

La Evaluación de Impacto de los Programas Sociales

Fundamentos Teóricos y Metodológicos y
Aplicación al caso peruano



Dr. Arístides Alfredo Vara Horna

Vara-Horna, Arístides (2007). *La evaluación de impacto de los programas sociales. Fundamentos teóricos y metodológicos y aplicación al caso peruano.* Lima: Fondo Editorial de la Universidad de San Martín de Porres.

AGRADECIMIENTOS

Esta obra es consecuencia del Concurso de Investigación Docente 2005 que organizó el Instituto de Investigación de la Facultad de Administración y Recursos Humanos de la Universidad de San Martín de Porres. La investigación titulada *“Nuevas estrategias metodológicas para evaluar el impacto de los programas sociales en el Perú”* fue una de las ganadoras, obteniendo una subvención para su desarrollo durante el año 2006. Por ello, debo mi especial agradecimiento al Instituto de Investigación, por su acogida institucional y por la promoción de la investigación científica; así como al Dr. Daniel Valera Loza, Decano de la Facultad, por su constante estímulo y ejemplo profesional.

Aristides A. Vara Horna.
Febrero de 2007

CONTENIDO

LISTA DE TABLAS	7
LISTA DE FIGURAS	11
RESUMEN.....	13
ABSTRACT.....	14
INTRODUCCIÓN	15
I. MARCO METODOLÓGICO.....	18
1.1. PLANTEAMIENTO DEL PROBLEMA	18
1.1.1. Descripción del Problema	18
1.1.2. Formulación del problema	20
1.2. JUSTIFICACIÓN E IMPORTANCIA	21
1.2.1. Justificación teórica	21
1.2.2. Justificación metodológica.....	21
1.2.3. Justificación práctica.....	22
1.3. OBJETIVOS	22
1.3.1. Objetivo general.....	22
1.3.2. Objetivos específicos	22
1.4. HIPÓTESIS Y VARIABLES	22
1.4.1. Hipótesis	22
1.4.2. Variables e indicadores	23
1.4.3. Matriz de consistencia.....	25
1.5. DISEÑO METODOLÓGICO.....	26
1.5.1. Población y muestra.....	26
1.5.2. Tipo de investigación.....	26
1.5.3. Diseño de la investigación	27
1.5.4. Instrumentos, procedimientos de recolección y análisis de datos.....	27
1.5.4.1. Sobre la simulación estadística	28
1.5.4.2. Proceso general de simulación estadística	30
II. BALANCE TEÓRICO	33
2.1. LOS PROGRAMAS SOCIALES	33
2.1.1. Definición y características.....	35

2.1.2. La evaluación y la gestión social	39
2.1.3. Los programas sociales en el Perú	42
2.2. LA EVALUACIÓN DE PROGRAMAS	46
2.2.1. Definición y características	48
2.2.2. Desarrollo histórico y modelos vigentes	53
2.2.3. Importancia social de la evaluación de programas	56
2.2.3.1. El rol de la Evaluación de Programas en la Política Social	59
2.2.4. La evaluación de programas sociales en el Perú	62
2.2.4.1. Aspectos diagnósticos de la situación actual	64
2.2.4.2. Sistemas de evaluación en la Administración Pública	66
2.3. LA EVALUACIÓN DE IMPACTO	70
2.3.1. Definición y características	71
2.3.2. Pertinencia de la evaluación de impacto	74
2.3.3. ¿Por qué siendo tan importante la evaluación de impacto se hace tan poco?	75
2.3.4. Técnicas de evaluación de impacto	81
2.3.4.1. Evaluaciones cuantitativas	82
2.3.4.2. Evaluaciones cualitativas	85
2.3.4.3. Evaluaciones mixtas cuali - cuantitativas	86
2.3.4.4. Otros enfoques de evaluación	87
2.3.4.5. ¿Cómo elegir el método de evaluación de impacto?	89
2.3.5. Principios metodológicos de la evaluación de impacto	94
2.3.6. Los indicadores de evaluación de impacto	98
2.3.7. Procedimiento general de la evaluación de impacto	103
2.3.7.1. Determinar la realización de la evaluación	104
2.3.7.2. Aclarar los objetivos de la evaluación	105
2.3.7.3. Examinar la disponibilidad de los datos	105
2.3.7.4. Diseñar la evaluación	107
2.3.7.5. Formar el equipo de evaluación	109
2.3.7.6. Evaluación de datos	110
2.3.7.7. Decisión sobre lo que se medirá	113
2.3.7.8. Muestreo	114
2.3.7.9. Instrumentos de medición	115
2.3.7.10. Administración de datos, análisis de la información y publicación	118
2.4. DEFINICIÓN DE CONCEPTOS	118

III. DESARROLLO METODOLÓGICO ACTUAL: UNA REVISIÓN SISTEMÁTICA.....121

3.1. DESARROLLOS NACIONALES	122
3.2. DESARROLLOS INTERNACIONALES	136
3.2.1. Experiencias institucionales	137
3.2.2. Desarrollo académico-institucional	140
3.2.2.1. Revisión sistemática de las tendencias de evaluación de impacto	246
3.2.2.1.1. Evaluación de impacto: Perspectiva global	246
3.2.2.1.2. Evaluación de impacto: Perú vs el mundo	253

3.2.2.2. Desarrollo metodológico: Nuevas estrategias.....	260
IV. ESTRATEGIAS METODOLÓGICAS DE EVALUACIÓN DE IMPACTO: MODELO ARYSIS	263
4.1. PRINCIPIOS.....	264
4.1.1. Principios del evaluador.....	265
4.1.1.1. Principio de sistematicidad en la evaluación	266
4.1.1.2. Principio de las competencias para desempeñarse como evaluador	267
4.1.1.3. Principio de la integridad y honestidad frente a los estudios.....	267
4.1.1.4. Principio de respeto a las personas	267
4.1.1.5. Principio de responsabilidad ante el bienestar general y público	268
4.1.2. Principios para el proceso de evaluación	268
4.1.2.1. Utilidad	268
4.1.2.2. Factibilidad	269
4.1.2.3. Rectitud	270
4.1.2.4. Precisión.....	270
4.2. PARÁMETROS TEÓRICOS	272
4.2.1. La inferencia causal	272
4.2.1.1. El problema de la validez de las evaluaciones.....	273
4.2.1.2. Modelos de la validez causal	275
4.2.1.2.1. Modelo original de Campbell y Stanley (1966).....	276
4.2.1.2.2. Modelo de Cook y Campbell (1979)	278
4.2.1.2.3. Modelos alternativos: UTOS de Cronbach.....	286
4.2.1.2.4. Evolución posterior de las tipologías de validez.....	288
4.2.1.3. La validez causal en ARYSIS.....	291
4.2.1.3.1. Estimación de los efectos de impacto	295
4.2.1.3.2. Procedimientos para disminuir el sesgo de selección	299
4.3. CRITERIOS METODOLÓGICOS	304
4.3.1. Definición de cuasi-experimento	305
4.3.2. Tipología cuasi-experimental.....	306
4.3.2.1. Los módulos básicos más empleados en la EvPro tradicionales.	309
4.3.2.1.1. Diseños pretest-postest sin grupo control	309
4.3.2.1.2. Diseños solamente postest con grupo control.....	310
4.3.2.2. Diseños con variable de asignación no conocida.....	311
4.3.2.2.1. Diseño con grupo control no equivalente	311
4.3.2.2.2. Diseño con doble pretest.....	315
4.3.2.3. Diseños de cohorte.....	316
4.3.2.3.1. Diseño básico de cohortes.....	317
4.3.2.3.2. Variaciones del diseño básico de cohortes.....	317
4.3.2.3.3. Diseño de cohortes con tratamiento dividido.....	319
4.3.2.4. Diseños de intercambio de tratamiento.....	320
4.3.2.5. Diseños con tratamiento invertido	321
4.3.2.6 Diseños sin grupo control	321
4.3.2.6.1. Diseños de grupo único.....	321
4.3.2.6.2. Diseños de línea-base no causal.....	323

4.3.2.7. Diseños con variables dependientes no equivalentes.....	324
4.3.2.8. Diseños con variable de asignación conocida.....	324
4.3.2.9. Diseños de series temporales interrumpidas	326
4.3.2.9.1. Diseño simple.....	328
4.3.2.9.2. Diseño con grupo control no equivalente	330
4.3.2.9.3. Diseño con variables dependientes no equivalentes	331
4.3.2.9.4. Diseño con replicaciones múltiples	331
4.3.2.9.5. Diseño con replicaciones intercambiables	332
4.4. SIMULACIONES EXPERIMENTALES	333
4.4.1. El modelo ideal	333
4.4.2. El modelo común: Grupos no equivalentes	338
4.4.3. Discontinuidad de la regresión.....	344
4.4.4. Series temporales interrumpidas	349
4.5. CONTRASTACIÓN DE HIPÓTESIS	357
CONCLUSIONES.....	361
RECOMENDACIONES	365
REFERENCIAS	367

LISTA DE TABLAS

<i>N°</i>	<i>Título</i>	<i>Pág.</i>
1.1	Matriz de variables e indicadores	24
1.2	Matriz de consistencia	25
1.3	Variables sistematizadas en el estudio biblio-integrativo	28
1.4	Características de la función de densidad normal	31
1.5	Significado de los comandos de simulación empleados con MINITAB 14 para generar variables	31
2.1	Definiciones sobre programas sociales	35
2.2	Tipos de evaluación y estados del proyecto	38
2.3	El proceso de gestión social, características y problemas asociados	40
2.4	Principales acontecimientos referidos a las políticas de gestión de los programas sociales en el Perú: 1990-2002	44
2.5	Definiciones sobre evaluación de programas sociales	49
2.6	Definición sucinta de la evaluación de programas sociales	52
2.7	Características comparativas de la evaluación tradicional y la evaluación en el marco de la gestión integral orientada a resultados.	52
2.8	Modelos metodológicos en la evaluación de programas	55
2.9	Definiciones sobre evaluación de impacto	71
2.10	Resumen de costos estimados de diversas evaluaciones de impacto del Banco Mundial	77
2.11	Ingreso mensual promedio de las familias que participan en el proyecto de vivienda de El Salvador y para un grupo control, 1976-1980.	80
2.12	Enfoques de evaluación de impacto	82
2.13	Opciones de diseño para reducir los costos de la obtención de datos	92
2.14	Procedimiento general de la evaluación de impacto	104
2.15	Puntos clave para identificar las fuentes de datos disponibles para la evaluación de impacto.	106
2.16	Repercusiones de las limitaciones de presupuesto, tiempo e información en la calidad de una evaluación de impacto.	108
2.17	Métodos de evaluación y necesidades de datos.	110
2.18	Algunas técnicas de recolección de datos para la evaluación de	115

	impacto.	
3.1	Revistas especializadas sobre evaluación de programas.	121
3.2	Evaluaciones de impacto más robustas de los programas sociales en el Perú.	126
3.3	Desarrollo académico de la evaluación de impacto de programas sociales en el mundo según investigaciones publicadas en Journals internacionales o instituciones de desarrollo.	141
3.4	Países donde se realizan las evaluaciones de impacto	247
3.5	Fuentes de publicación de los artículos o informes de evaluación de impacto	247
3.6	Área donde se realizó la evaluación de impacto	248
3.7	Población beneficiaria de los programas considerados en los artículos o informes de evaluación de impacto	249
3.8	Unidad de análisis considerada por los artículos de informes de evaluación de impacto	249
3.9	Origen de los datos utilizados para el análisis de la evaluación de impacto	250
3.10	Instrumento utilizado para la recolección de datos en los artículos o informes de evaluación de impacto	250
3.11	Metodología de investigación empleada en los artículos o informes de evaluación de impacto	250
3.12	Metodología de investigación empleada según el uso de grupo control en los artículos o informes de evaluación de impacto	251
3.13	Frecuencia de artículos o informes de evaluación de impacto según la metodología de investigación y el tipo de muestreo empleado	251
3.14	Tamaño de muestras reportadas en los informes de evaluación de impacto	251
3.15	Diseños de evaluación reportados en los informes de evaluación de impacto	252
3.16	Uso de variables instrumentales en los informes de evaluación de impacto	252
3.17	Uso de emparejamiento de grupos en los informes de evaluación de impacto	252
3.18	Análisis estadísticos empleados en los informes de evaluación de impacto	253
3.19	Impacto alcanzado por el programa o proyecto social evaluado	253
3.20	País donde fueron realizados los programas sociales contenidos en los informes de evaluación de impacto	254
3.21	Fuente de publicación de los artículos o informes de evaluación de impacto en el Perú y otros países	254
3.22	Áreas de evaluación de impacto en el Perú y otros países	254
3.23	Población beneficiaria de los programas evaluados en el Perú y otros países	256
3.24	Unidad de análisis considerada en los informes de evaluación de impacto en el Perú y otros países	256
3.25	Origen de los datos utilizados para el análisis según se indica en los	257

	informes de evaluación de impacto en el Perú y otros países	
3.26	Instrumentos utilizados para la recolección de datos según los informes de evaluación de impacto en el Perú y otros países	257
3.27	Metodología de investigación empleada en la evaluación de impacto en el Perú y otros países	257
3.28	Diseños de evaluación utilizados en la evaluación de impacto en Perú y otros países	258
3.29	Tipo de muestreo empleado en la evaluación de impacto en Perú y otros países	258
3.30	Técnicas de control de variables extrañas utilizadas en la evaluación del impacto de los programas sociales del Perú y otros países.	259
3.31	Análisis estadísticos empleados en la evaluación de impacto en Perú y otros países	259
3.32	Impacto alcanzado por el programa social evaluados en Perú y otros países	259
4.1	Tipos de validez y sus amenazas según Campbell y Stanley (1966).	277
4.2	Amenazas a la validez en Cook y Campbell, 1979.	279
4.3	Amenazas a la validez de la conclusión estadística según Cook & Campbell, 1979.	280
4.4	Amenazas a la validez interna según Cook & Campbell, 1979.	281
4.5	Amenazas a la validez de constructo según Cook & Campbell, 1979.	283
4.6	Amenazas a la validez externa según Cook & Campbell, 1979.	284
4.7	Evolución de las clasificaciones de amenazas a la validez en los diversos modelos de Campbell & Cols.	290
4.8	Amenazas a la validez y tratamiento en ARYSIS.	292
4.9	Ventajas y desventajas del Propensity Store Matching.	302
4.10	Diseño pretest-postest sin grupo control	309
4.11	Representación de resultados del diseño pretest-postest sin grupo control	309
4.12	Diseño solamente postest	310
4.13	Representación de resultados en diseño solamente postest	311
4.14	Diseño con grupo de control no equivalente.	312
4.15	Presentación de resultados del diseño con grupo de control no equivalente.	312
4.16	Diseño con doble pretest.	315
4.17	Presentación de resultados del diseño con doble pretest.	315
4.18	Diseño de cohortes básico.	317
4.19	Diseño de cohortes con pretest.	318
4.20	Presentación de resultados del diseño de cohortes con pretest.	318
4.21	Diseño con tres cohortes.	319
4.22	Extensión del diseño de tres cohortes.	319
4.23	Diseño de cohortes con tratamiento dividido.	320
4.24	Diseño de intercambio de tratamiento.	320
4.25	Diseño con tratamiento invertido	321
4.26	Diseño de grupo con tratamiento y control más típico.	322
4.27	Diseño de tratamiento repetido.	322

4.28	Diseño con grupo control no equivalente.	324
4.29	Diseño con variable de asignación conocida.	325
4.30	Presentación de resultados del diseño de discontinuidad de regresión.	326
4.31	Serie de tiempo interrumpida de diseño simple	329
4.32	Serie de tiempo interrumpida con grupo control	330
4.33	Serie de tiempo interrumpida con variables dependientes no equivalentes	331
4.34	Serie de tiempo interrumpida con replicaciones múltiples	331
4.35	Serie de tiempo interrumpida con replicaciones intercambiables	332

LISTA DE FIGURAS

<i>N°</i>	<i>Título</i>	<i>Pág.</i>
1.1	Relación entre variables e indicadores	23
2.1	Distribución legal entre políticas, programas y proyectos sociales.	37
2.2	Ciclo del proyecto de inversión según el SNIP.	68
2.3	La conversión de insumos en resultados.	95
2.4	Relación entre principios metodológicos de EI.	97
3.1	Distribución porcentual de la evaluación de impacto de los programas sociales según continente	246
3.2	Año de publicación de los artículos o informes de evaluación de impacto	248
3.3	Año de publicación de los artículos o informes de evaluación de impacto en el Perú y otros países	255
4.1	Modelo de Evaluación de Impacto ARYSIS.	264
4.2	Principios de Evaluación de Impacto ARYSIS.	265
4.3	Principios básicos del evaluador de impacto en ARYSIS.	266
4.4	Evolución desde Campbell & Stanley (1966) a Cook & Campbell (1979).	278
4.5	Estados posibles para beneficiarios y controles.	296
4.6	Tipología del diseño cuasi-experimental.	308
4.7	Relación entre postest y pretest cuando no existe impacto alguno del programa	334
4.8	Relación entre postest y pretest cuando existe impacto del programa en 10 puntos.	335
4.9	Relación entre postest y pretest cuando existe impacto del programa en 10 puntos y con una reducción del error de medida.	338
4.10	Relación entre postest y pretest cuando existe impacto del programa en 10 puntos y cuando los grupos no son equivalentes.	340
4.11	Relación entre postest y pretest cuando existe impacto del programa en 10 puntos y cuando los grupos no son equivalentes en perjuicio del GE.	343
4.12	Relación entre postest y pretest cuando existe impacto del programa en 10 puntos y cuando los grupos no tienen asignación	346

	conocida.	
4.13	Relación entre postest y pretest cuando no existe impacto del programa y cuando los grupos son de asignación conocida.	347
4.14	Serie simple con cambio de nivel posprograma, en supuesto significativo y no significativo.	350
4.15	Serie simple sin cambio de nivel y con cambio de tendencia del posprograma, en supuesto significativo y no significativo.	352
4.16	Serie simple con cambio de nivel y con cambio de tendencia del posprograma, en supuesto significativo.	354
4.17	Serie temporal con con cambio de nivel y con cambio de tendencia del posprograma en GE, y sin cambio de nivel y con tendencia en GC.	356

RESUMEN

La evaluación de impacto (EI) es especialmente importante en los países en desarrollo, donde los recursos son escasos y cada dólar gastado debe maximizar su efecto en la reducción de la pobreza. Sin embargo, a pesar de los miles de millones de dólares que se emplean en inversión social cada año, aún se conoce muy poco sobre el efecto real de tales proyectos en el Perú. La razón de ello es que los diseñadores de programas desconocen las metodologías modernas de evaluación de impacto, y los estudios adolecen de deficiencias y limitaciones en su aplicación.

Mediante un diseño bibliointegrativo se revisa sistemáticamente a) 408 investigaciones de impacto (1976-2006) publicadas en revistas especializadas o por organismos internacionales, así como b) 29 investigaciones de impacto de programas sociales con la mayor rigurosidad metodológica realizados en el Perú.

Entre los principales resultados se encuentra: a) la principal fuente de difusión y producción de EI en el mundo son los Organismos Internacionales, quienes han aportado más del 75% de la producción de informes, centrándose en los países en desarrollo y evaluando aspectos educativos, de salud, pobreza y empleo; b) Las EI han aumentado geoméricamente durante los últimos 30 años, concentrándose más del 77% en el periodo 2000-2006; c) Un importante porcentaje de EI (37.5%) utiliza datos secundarios, y en el caso de los primarios, el instrumento más usado es el cuestionario, d) El diseño más usado es el cuasi-experimental (41%), principalmente “diferencias en diferencias”; e) El tamaño de la muestra es muy variable (33.4% supera las mil unidades) y se centra principalmente en personas y familias; f) En cuanto al uso de técnicas de control especiales, su uso es muy limitado para el caso de las variables instrumentales (7.6%) y limitado para el matching (19.7%); g) La técnica estadística más empleada es la regresión (38%); h) Por lo menos 2 de cada 10 programas evaluados no muestran impacto positivo alguno.

Sobre la base de lo anterior, se determina los fundamentos teórico-metodológicos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú. Mediante un diseño de simulación computarizada se ejemplifica la aplicación del modelo ARYSIS, demostrando la importancia de sus supuestos teóricos.

Se concluye que los principios internacionales de evaluación, la teoría de la causación, el diseño cuasi-experimental avanzado y la estadística de simulación vía Montecarlo son los fundamentos básicos para la constitución del Modelo ARYSIS.

Palabras clave: Evaluación, Impacto, Programas, Gestión, Sociales, Perú, Cuasi-experimental, Causación.

ABSTRACT

The impact evaluation (IE) is especially important in developing countries, where resources are scarce and each expended dollar must increase its effect on poverty reduction. Nevertheless, in spite of thousands of million dollars employed in social investment each year, not much is known about the real effects of social programmes in Perú. The reason is that programme designers unknow modern methodologies of impact evaluation, and studies have shortcomings and limitations in their application.

By means a bibliointegrative desing it was done a sistematic review of a) 408 papers of impact evaluation (1976-2006) published both in specialized journals and by international organizations, as well as b) 29 papers of social programmes impact evaluation with the greatest methodological regurocity developed in Perú.

Between the main results it was found that: a) The main source of IE spreading and production around the world are International Organizations, wich have contributed with 75% of papers production and are concentrated in developing countries, taking account of issues such as education, health, poverty and emplorment; b) The IEs have had a geometric increased during the last 30 years, more than 77% is settled in 2000-2006 period; c) A significant percent of IE (37.5%) use secondary dates; and in case of primary information, the questionnaire was the tool most used, d) The researh desing most frequently used is the quasi-experimental (41%), specially “difference in the difference”; e) sample size is varied (33.4% exceed the thousand units) and are specially focused to people and families; f) Respect the use of special control techniques, it´s limited to intrumental variables (7.6%) and are limited to he matching (19.7%); g) Regression is the statistical technique most employed (38%); h) At least two of each ten evaluated programmes don´t show some positive impact.

Based on the previus descption, theoretical-methodological basis are determined for Sistema de Evaluación de Impacto ARYSIS of social programmes in Perú. By means of a computer simulation desing the ARYSIS model application is ilustrated, showing the importance of it´s theoretical assumptions.

It´s concluded that international principles of evaluation, causal theory, advance cuasi-experimental desing, and Via Montecarlo simulation statistics are esencial basis to the Modelo ARYSIS constitution.

Key words: Evaluation, Impact, Programmes, Management, Socials, Perú, Cuasi-experimental, causal theory.

INTRODUCCIÓN

El expresidente de los Estados Unidos, Bill Clinton, al firmar el *Government Performance and Results Act* (agosto de 1993) afirmó: *“Puede parecer asombroso, pero como otras grandes entidades, la nuestra está dominada por consideraciones de input: ¿Cuánto gastamos, cuánta gente tenemos, qué reglas nos gobiernan? y muchísimo menos por consideraciones de output: ¿funciona esto, y está cambiando y mejorando la vida de la gente?”* Esta afirmación, tan sencilla y aparentemente inofensiva, resume el concepto de un quiebre paradigmático que la administración pública y la gerencia social han venido experimentando en las últimas décadas.

En efecto, las tendencias teóricas recientes sobre la Gerencia Social y la Administración Pública enfatizan en las ventajas de promover una gestión orientada a resultados, en contraste con una orientación tradicional que haga hincapié en procedimientos y procesos de gestión (Barzelay, 2001; Caiden, 1998; Cunill & Ospina, 2003; CLAD, 1999; Moctezuma & Roemer, 1999; Olías de Lima, 2001; Osborne & Gaebler, 1994; Ramio, 2001). Estas nuevas tendencias (iniciadas en el Reino Unido, Australia, Nueva Zelanda, y luego en los Estados Unidos) enfatizan en la necesidad de evaluar las políticas públicas, proponiendo un cambio de paradigma en la administración pública, conocido como “neo-gerencialismo”.

En este nuevo contexto, la evaluación de la gestión pública pasa de ser un ejercicio técnico a ser una herramienta política, en la medida que: i) abarca los tres niveles de la acción social, el macro, meso y micro y ii) genera capacidad gerencial para establecer conexiones entre ellos, creándose así coherencia en los distintos niveles de la administración. De lo dicho, hoy se entiende a la evaluación de la gestión pública como una herramienta de modernización del Estado, más que una simple técnica administrativa.

Este nuevo paradigma tiene algunos aspectos característicos que son importantes resaltar. En primer lugar, otorga mayor énfasis a los resultados producidos por las organizaciones más que a los insumos o al proceso para conseguirlos (énfasis en la efectividad e impacto de la gestión). En segundo lugar, otorga mayor atención al punto de vista de los “clientes/consumidores” de los

servicios producidos, más que a los empleados o productores del servicio (énfasis en la calidad de los servicios). En tercer lugar, enfatiza en el desempeño de las organizaciones. En cuarto lugar, da mayor flexibilidad al gerente social, pero también pide más responsabilidad y un claro rendimiento de cuentas frente al desempeño de su organización (*accountability*). Y en quinto lugar, reduce la participación directa del Estado Central en la producción de servicios, buscando la descentralización y la coordinación inter-institucional.

En este contexto, la evaluación de los programas sociales, entendidos como manifestaciones materiales de las políticas públicas, se ha constituido en una necesidad obligatoria para el gerente social; entendiéndola asociada con el interés público, el mejoramiento de la sociedad y la contribución a la democracia.

Dentro de las modalidades de evaluación (ex ante, procesos, ex post, etc.), la evaluación de impacto es una de las más importantes y, paradójicamente, una de las menos conocidas. Si bien es cierto que todas las modalidades de evaluación son significativamente importantes, sólo la evaluación de impacto puede identificar los efectos provocados por un conjunto de acciones programadas y emitir un juicio de valor que califique, en conjunto, si las actividades realizadas o los servicios brindados, han producido los resultados esperados. Adicionalmente, la evaluación de impacto es la única modalidad de evaluación que permite sugerir las modificaciones necesarias para enfrentar los problemas detectados y aprovechar las fortalezas del programa, contribuyendo con un mayor logro de sus objetivos.

En términos generales, la evaluación de impacto contribuye a un empleo eficiente de los recursos y a una alta efectividad, al ofrecer información para mejorar los procesos de toma de decisiones. Así, evaluar el impacto de los programas sociales constituye un elemento fundamental para tomar decisiones y para optimizar las políticas, estrategias y asignación de recursos sociales. Alcanzar dicho objetivo implica realizar una evaluación de impacto integral, válida y confiable, acorde al desarrollo científico y tecnológico moderno; pues tal como afirman Cook, Leviton & Shadish (1985, 1991), Rossi & Freeman (1985) o Glass & Ellett (1980), la evaluación de programas es la aplicación sistemática de los procedimientos de la investigación social para probar nuevas ideas de intervención, evaluar la concepción, diseño, instauración y utilidad de los programas de intervención social, proporcionando, por tanto, una retroalimentación eficaz.

La evaluación de impacto es, pues, una herramienta técnica y política necesaria e indispensable para la gerencia social moderna, sin embargo, su complejidad y escaso acervo bibliográfico en habla hispana impide su posicionamiento en el imaginario académico e institucional. A estas limitaciones se suman muchas creencias erróneas sobre este método, desvirtuando su naturaleza y ahondando más la carencia comprensiva del tema y su escasa utilización.

En efecto, el gerente social moderno entiende la necesidad de evaluar sus procesos de gestión, sin embargo, aún carece de herramientas metodológicas y teóricas que le permitan valorar la calidad y rigurosidad de las evaluaciones de impacto. Por ello, la presente investigación busca integrar la hemerografía moderna sobre la metodología de evaluación de impacto para proporcionar los fundamentos teóricos y metodológicos necesarios que permitan entender esta compleja metodología y, además, aporta algunos lineamientos base para la constitución de

un nuevo modelo de evaluación de impacto de los programas sociales en el Perú y ejemplificar su aplicación con el uso de los diseños de simulación experimental computarizada.

Así, por tanto, la presente investigación se estructura en cuatro capítulos. El primer capítulo se refiere al marco metodológico de la pesquisa, el cual contiene el planteamiento del problema, la justificación e importancia de la investigación, los objetivos e hipótesis del estudio, las variables e indicadores así como el diseño metodológico propiamente dicho.

El capítulo II está dedicado a la elaboración del Balance Teórico. En una primera parte, se revisan los aspectos históricos, definición y características y ciclo de vida de los programas sociales y el contexto de la evaluación dentro del proceso de gestión social. Seguidamente, se hace una revisión bibliográfica sobre la definición y características de la evaluación de programas, su desarrollo histórico, importancia social, su papel en la política y la sociedad, así como la situación de la evaluación de programas sociales en el Perú y un análisis crítico de los sistemas de evaluación existentes en la administración pública hasta la fecha. Finalmente, se realiza un análisis teórico de la evaluación de impacto, realizándose un análisis minucioso de sus conceptos, su importancia, sus principios fundamentales, y las principales técnicas empleadas en su ejecución.

En el capítulo III se revisa el estado actual del desarrollo metodológico de la evaluación de impacto de los programas sociales. En una primera parte, se analizan los antecedentes nacionales inmediatos a la presente investigación. Posteriormente, en una segunda parte, se revisan, utilizando el método bibliointegrativo, los avances teóricos-metodológicos de la evaluación de impacto en el ámbito internacional, los cuales servirán de fundamento para el diseño del Sistema de Evaluación de Impacto ARYSIS. Se hace también una revisión sistemática con una síntesis estadística de las principales tendencias en la aplicación del método. Se realiza también un balance comparativo entre las evaluaciones realizadas en el Perú y en el resto del mundo.

En el capítulo IV se presenta y analiza los aspectos teórico-metodológicos fundamentales del Modelo de Evaluación de Impacto ARYSIS. En este capítulo se presentan los principios, parámetros teóricos, principios metodológicos y procedimientos estandarizados que fundamentan el modelo propuesto. Asimismo, se desarrollan algunos casos de simulación mediante el Método de Montecarlo para ejemplificar la aplicación del Modelo ARYSIS.

Finalmente, las conclusiones y recomendaciones son presentadas. Se espera que con la presente investigación aumente el nivel de comprensión de una herramienta metodológica necesaria en una realidad tan urgente de evaluación. Y, además, que la presente investigación contribuya con el desarrollo académico de los gerentes sociales, quienes, hoy más que nunca, tienen un papel protagónico en el desarrollo del país.

I. MARCO METODOLÓGICO

1.1. PLANTEAMIENTO DEL PROBLEMA

1.1.1. Descripción del Problema

La planificación es un requisito fundamental e imprescindible en el éxito de cualquier empresa o actividad social. En términos macro, el desarrollo de una nación depende, en gran medida, de la rigurosidad y compromiso en la elaboración de los planes de desarrollo. Pero planificar no es un proceso unívoco; se requiere, inevitablemente, contar con estrategias para evaluar la ejecución de los planes. Así, planificar y evaluar son dos aspectos del mismo proceso, recíprocos y mutuamente influyentes (BID, 1997; Tilley, 2000). En efecto, una planificación excelente considera, dentro de cada etapa, diversos mecanismos de evaluación. Es inconcebible planificar sin considerar las estrategias para evaluar, pues ¿cómo saber si los planes ejecutados están resultando o produciendo los efectos esperados?, ¿cómo seguir planificando, mejorando o cambiando su estructura si no se conocen los efectos iniciales de los planes originales? Sólo la evaluación puede responder tales preguntas.

Conocer los procesos, resultados e impactos de las políticas de desarrollo social constituye un elemento fundamental para mejorar la toma de decisiones y obtener mayor eficacia en las políticas diseñadas; sirve también para mejorar la implantación de sus estrategias y aumentar la eficiencia en la asignación de recursos (Abdala, 2004; Vedung, 1997). Sin embargo, todos estos beneficios se lograrán **solamente** si se realiza una evaluación integral, válida y confiable de las políticas aplicadas y, si disponen de recursos financieros suficientes y de recursos humanos capacitados para realizarla.

Debido a que la evaluación es un proceso amplio y continuo, existen diversos tipos de evaluación; sin embargo, la evaluación de impacto es quizá la más importante, pues nos informa sobre el efecto real de los programas sociales (Waissbluth, 2002). Si no se evalúa el impacto, no se sabrá si la intervención social está produciendo los beneficios previstos en la población; no se sabrá si se podría mejorar el diseño del programa para lograr los resultados; no se sabrá si se están

empleando los recursos en forma eficaz. Por eso, su desuso es negligencia, pues sólo el gerente social bien preparado entiende que estos son los tipos de preguntas que solamente se responden mediante una evaluación de impacto.

En efecto, la evaluación de impacto es una necesidad gubernamental, pues es especialmente importante en los países en desarrollo, donde los recursos son escasos y cada dólar gastado debe maximizar su efecto (Banco Mundial, 2000). Si los programas están mal diseñados, no llegan a los beneficiarios previstos y despilfarran los recursos, solo con la evaluación de impacto es posible rediseñarlos, mejorarlos o eliminarlos si se estima necesario (Blomquist, 2003; CONPES, 2002). El conocimiento obtenido de estos estudios también proporciona información decisiva para el diseño adecuado de programas y proyectos futuros (Cook, Leviton & Shadish, 1985, 1991; Rossi & Freeman, 1985; Glass & Ellett, 1980).

Sin embargo, a pesar de la irrefutabilidad de lo dicho, aún muchos gobiernos e instituciones, así como gerentes de proyectos, rechazan las evaluaciones de impacto por diversas razones (Medina Giopp, 2002). Son pocos los casos en América Latina (y especialmente en nuestro país) en los que se puede hablar de una evaluación sistemática del impacto de los programas sociales. De la revisión pormenorizada de sus estrategias se puede encontrar que los modelos de evaluación utilizados son –en su mayoría– demasiado simples, inadecuados, incompletos, inseguros y en muchos casos, inválidos. Además, de acuerdo a las revisiones de Sulbrandt (1993), la mayoría de las políticas y los programas públicos de América Latina no se evalúan y, salvo excepciones, las pocas evaluaciones que se han hecho no han provocado ningún cambio en el manejo de los programas ni en el aprendizaje de los ciudadanos.

Diversas son las causas de estas deficiencias, pero la gran mayoría se circunscribe al factor humano (Mokate, 2001; Baker, 2000). En efecto, existen muchas creencias erróneas y actitudes negativas hacia la evaluación de impacto. La gran mayoría de administradores de programas la consideran costosa, prolongada, técnicamente compleja y políticamente delicada (cuando sus resultados demuestran ineficiencia e ineficacia de la gestión realizada). Se critica también que los resultados se obtienen demasiado tarde, que no responden las preguntas adecuadas o que no se realizaron con suficiente rigor analítico (Medina Giopp, 2002). En fin, la evaluación de impacto recibe muchas críticas y descréditos, pero lo cierto es que ningún método de evaluación alternativo se ha mostrado superior para reemplazarla.

Según mí entender, todas estas creencias y actitudes negativas son producto de tres factores: a) la ignorancia generalizada sobre el tema, b) el desconocimiento de los modelos modernos de evaluación de impacto de los programas sociales y c) la escasez de especialistas.

Es verdad que la parte más dificultosa de la evaluación de impacto es su dimensión metodológica, pues evaluar el impacto de un programa requiere conocer mucho de metodología y estadística avanzada; es un terreno para especialistas. Sin embargo, el nivel de especialización de la evaluación de impacto no puede ser excusa para desdeñar sus aportes o para desentenderse de su importancia. Todo lo contrario, su especialización debe exigir una aproximación comprensiva para los

gerentes sociales y autoridades de la administración pública, pues sólo comprendiendo su real alcance y limitaciones puntuales se podrá demandar su ejecución y valorar la calidad de sus resultados.

Al respecto, existe una limitación seria: la bibliografía especializada de habla hispana es casi inexistente y las fuentes bibliográficas están dispersas en diversas revistas o informes institucionales, principalmente, de habla inglesa. En efecto, desde hace cuatro décadas diversos autores y organismos internacionales han propuesto una serie de modelos para planificar y realizar evaluaciones de impacto (Ej. el Modelo cuasi-experimental, el Modelo UTOS, el del Banco Mundial, el de la *Evaluation Research Society*, entre otros), sin embargo, en las estrategias y bibliografía de evaluación utilizadas en América Latina y en el Perú, se encuentran muy pocas aplicaciones de las mencionadas.

Urge así, pues, un material bibliográfico comprensivo y didáctico que, basándose en las investigaciones modernas, condense las nociones teóricas y metodológicas relacionadas a la evaluación de impacto de los programas sociales en el Perú. Además, sobre la base de lo anterior, se requiere la propuesta de un modelo ajustado a la realidad nacional.

De lo dicho, tal material proporcionaría a los gerentes de proyectos y a los analistas de políticas nacionales, las herramientas teóricas-metodológicas necesarias para evaluar el impacto de sus proyectos. Por ello, esta investigación, enfocada desde una perspectiva estrictamente metodológica, pretende llenar el vacío teórico-metodológico nacional con interesantes aportes en el ejercicio de la investigación evaluativa. Pretende, además, proponer un modelo integral de evaluación de impacto, acorde a los avances mundiales en el tema y a luz de la realidad nacional.

Por tanto, con la presente investigación se propone los fundamentos de un nuevo modelo de evaluación de impacto (cuyo nombre es ARYSIS), comprensivo y didáctico, acorde a los modernos avances científicos y metodológicos de la evaluación de programas y de la estadística experimental. Para ello, en primer lugar, se revisa biblio-integrativamente el estado de la cuestión sobre la evaluación de programas y la evaluación de impacto, para luego, en segundo lugar, sistematizar sus bases primarias, fundamentos teóricos y metodológicos, los cuales provienen del desarrollo de la estadística multivariada y del diseño cuasi-experimental, asociado a los criterios y conceptos de validez e inferencia causal.

1.1.2. Formulación del problema

Así, ante lo anteriormente expuesto, se busca respuesta al siguiente problema general: *¿Cuáles son los fundamentos teórico-metodológicos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú?*, el cual, para efectos de la investigación, se sistematiza mediante las siguientes preguntas específicas:

1. ¿Cuál es el estado actual del desarrollo teórico de los programas sociales, la evaluación de programas y la evaluación de impacto de los programas?

2. ¿Cuál es el estado actual del desarrollo académico-metodológico de la evaluación de impacto en los ámbitos internacional y nacional?
3. ¿Cuáles son los fundamentos principistas, teóricos y metodológicos del sistema de Evaluación de Impacto ARYSIS?

1.2. JUSTIFICACIÓN E IMPORTANCIA

En un contexto donde los recursos son escasos, las entidades que financian y ejecutan los programas sociales requieren cada vez más precisión en los resultados de la evaluación. En este sentido, la evaluación de impacto juega un papel fundamental porque mide los cambios en las condiciones de vida de los beneficiarios y explica en qué medida estos cambios pueden ser atribuidos al programa (Cook, Leviton & Shadish, 1985, 1991; Rossi & Freeman, 1985; Glass & Ellett, 1980). Sin embargo, a pesar de la importancia que tiene la evaluación de impacto en el desarrollo de los programas y de su falta de utilización, año tras año se invierten grandes sumas de dinero para el financiamiento de programas sociales que nunca llegan a determinar su impacto en la calidad de vida de las personas a las que van dirigidos. La escasez de evaluaciones de impacto obedece en gran medida, a la falta de capacitación idónea del personal encargado de diseñar, ejecutar y evaluar los programas. Para adquirir capacitación y aprender técnicas cuantitativas en evaluación de impacto de programas de salud, población, educación, entre otras, la presente investigación pretende compilar, analizar y proponer nuevos aportes teóricos y metodológicos de la evaluación de impacto aplicado a programas sociales en el Perú. Por tanto, la realización de la presente investigación se justifica por tres razones complementarias: teórica, metodológica y práctica.

1.2.1. Justificación teórica

Se recorren de manera exhaustiva las posibilidades metodológicas y técnicas que resultan aplicables a los programas sociales en una realidad como la nuestra. Esta exploración de alternativas planteadas en la bibliografía internacional permitirá poner al alcance de ejecutores y evaluadores locales, sofisticadas técnicas métricas y estadísticas, además de identificarse las bondades y limitaciones de cada una de ellas para la medición del impacto. Además, se identifican los parámetros teóricos de un modelo integral de evaluación de impacto, lo cual es parte de una aproximación biblio-integrativa.

1.2.2. Justificación metodológica

Se propone elaborar un documento de referencia que sirva a los diseñadores, gerentes sociales y evaluadores de programas sociales, como base para la construcción de estrategias propias de operaciones en evaluación de impacto de los programas sociales. Esta herramienta intenta ser de significativa importancia para la evaluación de programas sociales y, debido a la escasa

bibliografía de habla hispana, pretende convertirse en un manual de consulta, dentro de la bibliografía local especializada sobre la materia. De esta manera, para los diseñadores de programas y gerentes sociales, existirá una clara descripción de cómo incorporar la evaluación de impacto de los programas como un elemento paralelo al programa en sí mismo y cómo evaluar la calidad de los mismos.

1.2.3. Justificación práctica

Los gerentes sociales encontrarán las herramientas necesarias para implementar algunos procesos de la evaluación y poder dialogar y exigir calidad al evaluador externo, cuando este sea el caso. Además, los evaluadores encontrarán una discusión sobre las ventajas y limitaciones de las distintas metodologías de evaluación de impacto, la manera cómo se deben computar los diversos estimadores del mismo y algunas estrategias para controlar las amenazas contra la validez de los resultados.

1.3. OBJETIVOS

1.3.1. Objetivo general

Determinar los fundamentos teórico-metodológicos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú.

1.3.2. Objetivos específicos

1. Determinar el estado actual del desarrollo teórico de los programas sociales, la evaluación de programas y la evaluación de impacto de los programas.
2. Determinar el estado actual del desarrollo académico-metodológico de la evaluación de impacto en los ámbitos internacional y nacional.
3. Identificar y analizar los fundamentos principistas, teóricos y metodológicos del sistema de Evaluación de Impacto ARYSIS.

1.4. HIPÓTESIS Y VARIABLES

1.4.1. Hipótesis

Como la presente es una investigación biblio-integrativa y aplicada, entonces la hipótesis se centra en la identificación y sistematización de los fundamentos teórico-metodológicos del Modelo de Evaluación de Impacto ARYSIS. Por la naturaleza del estudio, esta hipótesis no se puede someter propiamente a una contrastación o verificación, sino que sólo representa una guía de cumplimentación de objetivos, basada en la revisión bibliográfica y experiencia previa. Hecho la salvedad, la formulación de la hipótesis general es como sigue:

“Los fundamentos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú son: los principios internacionales de evaluación, la teoría de la causación, el diseño cuasi-experimental avanzado y la estadística de simulación vía montecarlo”.

Las hipótesis específicas, también con las características de la hipótesis general, son las siguientes:

1. *“Aunque existe un desarrollo teórico amplio sobre los programas sociales y la evaluación de programas; no existe un nivel teórico comprehensivo de la evaluación de impacto de los programas sociales”.*
2. *“Existe un desfase significativo entre el estado actual del desarrollo académico-metodológico de la evaluación de impacto en el ámbito internacional y el desarrollo metodológico nacional”.*
3. *“Los fundamentos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú son: los principios internacionales de evaluación, la teoría de la causación, el diseño cuasi-experimental avanzado y la estadística de simulación vía montecarlo”.*

1.4.2. Variables e indicadores

De la hipótesis general formulada se pueden derivar las variables de investigación. Tal como se mencionó previamente, éste es un estudio bibliointegrativo, por tanto, se centra en la identificación y sistematización de fundamentos para un nuevo modelo de evaluación de impacto. En este caso, la **variable dependiente** es el modelo de evaluación de programas ARYSIS, mientras que las **variables independientes** son los fundamentos teórico-metodológicos: a) principios internacionales, b) la teoría de la causación, c) el diseño cuasi-experimental y, d) la simulación estadística de montecarlo.

La relación es causal por cuanto la identificación y desarrollo de cada uno de los fundamentos teórico-metodológicos (variables independientes) contribuyen a la generación y sustentación del modelo de evaluación de impacto ARYSIS (variable dependiente). La relación causal entre estas variables se esquematiza en la siguiente figura:

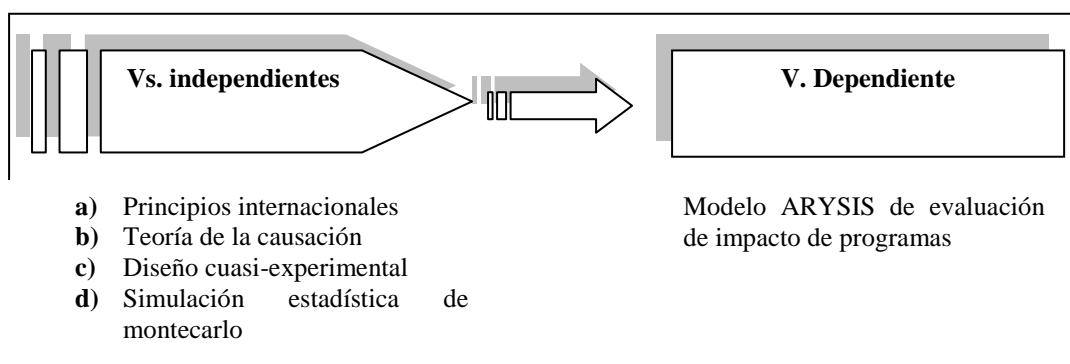


Figura N° 1.1. Relación entre variables e indicadores (Fuente: Elaboración propia).

En cuanto a los indicadores de cada una de las variables mencionadas, estas se simplifican en la siguiente tabla:

Tabla N° 1.1. Matriz de variables e indicadores (Fuente: Elaboración propia).

Variables	Definición operacional o constitutiva	Indicadores
Fundamentos teóricos metodológicos	Principios internacionales	<ul style="list-style-type: none"> • Principios del evaluador • Principios procesales
	Teoría de la causación	<ul style="list-style-type: none"> • Modelos de validez causal
	Diseño cuasi-experimental	<ul style="list-style-type: none"> • Diseños de asignación no conocida • Diseños de asignación conocida
	Simulación estadística de montecarlo	<ul style="list-style-type: none"> • Generación de datos • Contraste de parámetros • Estimación de sesgos • Corrección de sesgos
Modelo de Evaluación de Impacto ARYSIS	Sistema dinámico de principios, parámetros teóricos, criterios metodológicos y procedimientos estandarizados, organizados y regulados en subsistemas, que se dirigen hacia la identificación y obtención de datos pertinentes y válidos para evaluar el impacto de los programas sociales.	Idénticos a los indicadores de la variable independiente.

1.4.3. Matriz de consistencia

En la siguiente tabla se resume la relación de coherencia entre el problema de investigación, los objetivos, hipótesis y variables. La relación se presenta tanto para el aspecto general como específico.

Tabla N° 1.2. Matriz de consistencia (Fuente: Elaboración propia).

<i>Título</i>	<i>Problema</i>	<i>Objetivo</i>	<i>Hipótesis</i>	<i>Variables</i>
Fundamentos para un modelo integral de evaluación de impacto de los programas sociales en el Perú	<p>GENERAL: ¿Cuáles son los fundamentos teórico-metodológicos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú?</p> <p>ESPECÍFICOS: ¿Cuál es el estado actual del desarrollo teórico de los programas sociales, la evaluación de programas y la evaluación de impacto de los programas?</p> <p>¿Cuál es el estado actual del desarrollo metodológico de la evaluación de impacto en los ámbitos internacional y nacional?</p> <p>¿Cuáles son los fundamentos principistas, teóricos y metodológicos del sistema de Evaluación de Impacto ARYSIS?</p>	<p>Determinar los fundamentos teórico-metodológicos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú.</p> <p>Determinar el estado actual del desarrollo teórico de los programas sociales, la evaluación de programas y la evaluación de impacto de los programas.</p> <p>Determinar el estado actual del desarrollo metodológico de la evaluación de impacto en los ámbitos internacional y nacional.</p> <p>Identificar y analizar los fundamentos principistas, teóricos y metodológicos del sistema de Evaluación de Impacto ARYSIS.</p>	<p>Los fundamentos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú son: los principios internacionales de evaluación, la teoría de la causación, el diseño cuasi-experimental avanzado y la estadística de simulación vía montecarlo.</p> <p>Aunque existe un desarrollo teórico amplio sobre los programas sociales y la evaluación de programas; no existe un nivel teórico comprehensivo de la evaluación de impacto de los programas sociales.</p> <p>Existe un desfase significativo entre el estado actual del desarrollo metodológico de la evaluación de impacto en el ámbito internacional y el desarrollo metodológico nacional.</p> <p>Los fundamentos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú son: los principios internacionales de evaluación, la teoría de la causación, el diseño cuasi-experimental avanzado y la estadística de simulación vía montecarlo.</p>	<p>Variable dependiente: Modelo de evaluación de programas ARYSIS</p> <p>Variables independientes: Fundamentos teórico-metodológicos:</p> <p>a) Principios internacionales. b) Teoría de la causación. c) Diseño cuasi-experimental. d) Simulación estadística de montecarlo.</p>

1.5. DISEÑO METODOLÓGICO

1.5.1. Población y muestra

En la presente investigación, la población corresponde a tres tipos distintos:

1. Al ser un estudio biblio-integrativo, el primer tipo de población consiste en la bibliografía y hemerografía internacional especializada en temas de evaluación de programas, evaluación de impacto y diseño cuasi-experimental. En este caso la muestra está definida por la búsqueda a través de diversas bibliotecas especializadas, adquisición de textos y el uso de bases de datos internacionales como EBSCO HOST, PROQUEST, ABI INFORM, ERIC, SSRN, las cuales contienen artículos completos de más de 20,000 revistas a nivel mundial. Se ha utilizado también las bases de datos institucionales del Banco Mundial, el Banco Interamericano de Desarrollo, la Agencia Norteamericana de Desarrollo Internacional, entre otras. Para ser incluido en la muestra, el informe o reporte de investigación debió referirse exclusivamente a evaluaciones de impacto.
2. En segundo lugar, al ser una investigación aplicada a la realidad peruana, se identifica y revisa los informes de evaluación de impacto de los programas sociales evaluados en el Perú. La población incluye los diferentes tipos de programas considerados sociales, como son los programas educativos, de salud, laborales, asistenciales, alimentarios y de infraestructura. En este caso, la muestra es seleccionada bajo un criterio intencional no probabilístico, considerando los programas sociales con las evaluaciones más robustas y significativas. La ausencia de muestreo probabilístico se justifica en la medida que se pretende analizar las evaluaciones de impacto consideradas más robustas y completas en el ámbito local, considerando, además, que el número de evaluaciones de impacto en el Perú es muy pequeño.
3. Adicionalmente, se elabora una muestra significativa de datos de programas simulados o ficticios, utilizando los métodos de simulación de Montecarlo, esto con el fin de demostrar las posibilidades aplicativas del Sistema de Evaluación ARYSIS y una ejemplificación didáctica de su uso.

1.5.2. Tipo de investigación

La presente es una investigación de tipo aplicada, en la medida que pretende proponer y difundir, una vez sistematizada, un referente bibliográfico y un modelo teórico-metodológico de evaluación de impacto de programas sociales dentro del contexto nacional.

1.5.3. Diseño de la investigación

La presente investigación utiliza un diseño biblio-integrativo y de simulación experimental mediante casos ficticios.

Se utiliza un **diseño biblio-integrativo** (conocido por algunos autores como revisión sistemática) porque se busca identificar y sistematizar los fundamentos de un modelo comprensivo, didáctico, sencillo para evaluar el impacto de los programas sociales en el Perú. Los diseños biblio-integrativos son revisiones sistemáticas de la literatura científica, pormenorizados, selectivos y críticos que tratan de analizar e integrar la información esencial de los estudios primarios de investigación sobre un problema específico, en una perspectiva de síntesis unitaria de conjunto (Delgado-Rodríguez, 2002; Shea, Dube & Moher, 2001; Cook, Mulrow & Haynes, 1997; Morales, 1993; Slavin, 1987; Center for Reviews and Dissemination, 2001). En este caso, el diseño biblio-integrativo identifica y analiza las mejores fuentes internacionales y locales para, sobre ellas, proponer un esquema sintético de la mejor evidencia disponible sobre la evaluación de impacto de los programas sociales.

Por otro lado, se utiliza un **diseño de simulación experimental** mediante casos ficticios para ejemplificar la aplicación del Modelo ARYSIS y para precisar sus posibilidades de aplicación. En efecto, debido a que ARYSIS es un modelo nuevo, aún en estado experimental, éste debe probarse para precisar sus alcances y limitaciones. Sin embargo, debido a que la evaluación de impacto es costosa y prolongada, sería inadmisibles probarlo en situaciones reales. Por ello, resulta necesario utilizar simulaciones computarizadas, pues brindan información oportuna sobre los alcances de nuevos modelos propuestos y sirven también como medio didáctico (Guetzkow, 1962; Lehmann, 1980; Trochim & Davis, 1986).

1.5.4. Instrumentos, procedimientos de recolección y análisis de datos

Para la recolección de los datos se han utilizado las diversas bases de datos de artículos y disertaciones, entre las más reconocidas EBSCO HOST, PROQUEST, ABI INFORM, ERIC, SSRN; así como las bases de datos de instituciones internacionales como el Banco Mundial (BM) o el Banco Interamericano de Desarrollo (BID), entre otras; nacionales como el Instituto de Estadística e Informática (INEI), Ministerio de Economía y Finanzas, Consorcio de Investigaciones y Estudios Sociales (CIES), de APOYO, de GRADE y de las instituciones gubernamentales como Ministerio de Trabajo, DEVIDA, MIMDES, entre otros.

Las revisiones sistemáticas se han realizado considerando el análisis individual de informes de evaluación de impacto, sea tanto institucional como artículo publicado en *Journal*. En esta investigación se define al informe de evaluación de impacto como aquél que utiliza fuentes de datos directos (primarios o secundarios), instrumentos de recolección de datos cualitativos o cuantitativos y que analiza mediante algún procedimiento estadístico o categórico el impacto de

un programa. En este caso, se seleccionaron solamente los artículos e informes publicados durante los años 1976-2006. Las variables sistematizadas provenientes de cada informe de evaluación se presentan en la Tabla N°1.3:

Tabla N° 1.3. Variables sistematizadas en el estudio biblio-integrativo (Fuente: Elaboración propia).

VARIABLES SISTEMATIZADAS	DEFINICIÓN
Autores	Apellidos e iniciales de los autores
Año	Año de publicación del artículo
Título	Título del artículo de investigación
Muestra	Número y descripción de la muestra empleada. La unidad de análisis también es referida.
Metodología	Si es experimental, cuasi-experimental, pre-experimental o cualitativo.
Presencia de grupo control	Si el diseño cuenta con al menos un grupo control.
Diseño	Indicar el sub-diseño: diferencias de diferencias, pre-post, sólo posttest, series de tiempo, discontinuidad de la regresión, entre otros.
Instrumento/método de recolección de datos	Nombre del instrumento empleado. Modalidad: encuesta, escala, guía de entrevista, focus group, etc.
Análisis de datos	Métodos estadísticas empleados para el análisis de los datos: modelos de regresión, ecuaciones estructurales, ARIMAs, estadísticas descriptivas.
Uso de procedimientos de emparejamiento o variables instrumentales	Si uso el matching o el uso de variables instrumentales para controlar el sesgo de selección o asignación integral.
Conclusiones	Principales conclusiones del estudio. Determinar si el programa tuvo impacto significativo.
Lugar de ubicación de la muestra	País en dónde se ubica físicamente la muestra beneficiada por el programa evaluado.
Fuente de la investigación	Si es un artículo de Journal o es un informe institucional. Indicar la fuente institucional.
Aspectos estudiados	Dimensión de la evaluación de programas estudiados. Área de estudio (empleo, salud, educación, infraestructura, justicia, entre otros).

Para el análisis biblio-integrativo se elaboró una matriz de tabulación tipo R (variables x artículos) utilizando el Paquete Estadístico para Ciencias Sociales (SPSS versión 11). Se emplearon estadísticas univariadas y bivariadas de distribución de frecuencias y de tablas cruzadas para resumir los principales hallazgos. Previamente se han realizado tablas resumen que describen individualmente cada uno de los informes analizados.

En otro aspecto, utilizando el paquete estadístico MINITAB 14 se generan datos simulados para demostrar y ejemplificar las aplicaciones del Modelo ARYSIS. Estos datos han sido simulados y procesados en una computadora Pentium IV de 3.8 Ghz y de 1200 MbSDRAM; utilizando el Método de Montecarlo y las estadísticas de regresión lineal múltiple, análisis de covarianza y análisis de comparación de medias y de descripción univariada.

1.5.4.1. Sobre la simulación estadística

Los responsables de la toma de decisiones necesitan información cuantificable sobre diferentes hechos que puedan ocurrir. La simulación constituye

una técnica muy económica ya que permite ofrecer varios escenarios posibles de un modelo de evaluación, permite cometer errores sin provocar efectos sobre el mundo real y sin gastar dinero. Se puede afirmar, entonces, que la simulación es una rama experimental dentro de la investigación social evaluativa (Gilbert & Troitzsch, 1999; Trochim & Davis, 1986).

La simulación es el desarrollo de un modelo lógico-matemático de un sistema, de tal forma que se obtiene una imitación de la operación de un proceso real o de un sistema a través del tiempo (Azarang & García, 1996). En efecto, la simulación surge para evaluar numéricamente un modelo. Un modelo es una representación analítica de una realidad, es un cuerpo de información relativo a un sistema. Si las relaciones que integran ese modelo están definidas, entonces se puede tener una representación numérica de las mismas y calcular una solución exacta en términos deductivos. Así, entonces, la simulación es el proceso de diseñar y desarrollar un modelo computarizado de un sistema o proceso y conducir experimentos con este modelo con el propósito de entender el comportamiento del sistema o evaluar estrategias con las cuales se puede operar el sistema. De lo dicho, el modelo de simulación sería el conjunto de hipótesis acerca del funcionamiento del sistema expresado como relaciones matemáticas o lógicas entre los elementos del sistema.

Entre los diversos procesos de simulación, el más popular es el Método de Simulación de Montecarlo, el cual viene aplicándose desde 1940 gracias al desarrollo teórico inicial de Von Neumann & Ulam (1949) en su ya clásico *The Montecarlo Method*. Este método permite resolver problemas analíticos mediante la simulación y el muestreo sistemático de variables aleatorias (Rubinstein, 1981). El Método de Montecarlo da solución a una gran variedad de problemas matemáticos haciendo experimentos con muestreos estadísticos en una computadora. En la presente investigación, el Método de Montecarlo se utiliza con fines experimentales, es decir se elabora distintos modelos de tipo estocásticos¹ y se intercambian parámetros para estudiar cuáles son los posibles resultados. La elección de este diseño se basa, entre otras razones, en las significativas ventajas de su utilización (Azarang & García, 1996):

- Una vez construido, el modelo simulado puede ser modificado de manera rápida con el fin de analizar diferentes políticas o escenarios.
- Generalmente es más barato mejorar el sistema vía simulación, que hacerlo directamente en el sistema real.
- Es mucho más sencillo comprender y visualizar los métodos de simulación que los métodos puramente analíticos.
- Los métodos analíticos se desarrollan casi siempre, para sistemas relativamente sencillos donde suele haber un gran número de suposiciones o simplificaciones, mientras que con los modelos de simulación es posible analizar sistemas de mayor complejidad o con mayor detalle.

¹ En los modelos estocásticos, los valores de las variables dentro un modelo sufren modificaciones aleatorias con respecto a un valor promedio; dichas variaciones pueden ser manejadas mediante distribuciones de probabilidad (Azarang & García, 1996).

- En algunos casos la simulación es el único medio para lograr una solución.

1.5.4.2. Proceso general de simulación estadística

La simulación involucra la generación de una historia artificial de un sistema; la observación de esta historia mediante la manipulación experimental, ayuda a inferir las características operacionales de tal sistema. Así, existen dos pasos básicos de una simulación: a) desarrollo del modelo y b) experimentación. El desarrollo del modelo² incluye la construcción de ecuaciones lógicas representativas del sistema y la preparación de un programa computacional. Una vez que se ha validado el modelo del sistema, la segunda fase de un estudio de simulación entra en escena, experimentar con el modelo para determinar cómo responde el sistema a cambios en los niveles de algunas variables de entrada.

Actualmente se puede utilizar cualquier programa de análisis estadístico para realizar simulaciones. Debido a que estos programas funcionan con lenguajes generales de programación, si se aprenden tales lenguajes y se ingresa al panel de comandos, entonces se podrá obtener tales ventajas. En la presente investigación se utiliza el lenguaje de comandos del MINITAB 14, potente software estadístico que es muy atractivo por su versatilidad y fácil acceso³. La simulación se realiza en el programa gratuito de prueba (acceso libre por un mes) y puede ser obtenido de la página web institucional (<http://www.minitab.com/products/minitab/14/demo/>).

Debido a que existen algunos requisitos matemáticos para modelar un sistema simulado, se debe diferenciar entre dos tipos de datos: los parámetros, los cuales permanecen sin cambio a través del tiempo, y las variables, las cuales presentan cambios a través del tiempo (Azarang & García, 1996). En este caso, debido a que el modelado de un sistema mediante simulación es útil cuando la información del sistema tiene carácter dinámico y probabilístico (estocástico), la variabilidad de las variables debe modelarse siguiendo ciertas ecuaciones matemáticas que sean capaces de reproducirla, en este caso, se pueden utilizar las distribuciones de probabilidad (Ej. Uniforme, normal, Z, exponencial, weibull, triangular, lognormal, Bernoulli, uniforme discreta, binomial, Poisson, Geométrica,

² Un modelo es una representación de los objetos de un sistema y refleja de manera sencilla las actividades en las cuales esos objetos se encuentran involucrados. Un sistema en cambio, es una colección de variables que interactúan entre sí dentro de ciertos límites para lograr un objetivo.

³ El programa Minitab permite simular, de forma rápida y sencilla, la obtención de muestras de variables aleatorias definidas por una determinada distribución de probabilidad, así como calcular valores correspondientes a su función densidad de probabilidad o función de distribución. Otra importante posibilidad que permite este programa es la de crear archivos ejecutables en formato ASCII, formados por conjuntos de instrucciones que pueden ejecutarse todas seguidas de forma automática. Esta opción está especialmente indicada cuando se debe realizar una acción de forma repetitiva. Además, desde Minitab también se pueden ejecutar programas compuestos por instrucciones típicas de programación, tales como IF, ELSEIF, ELSE, DO, WHILE, NEXT, GOTO, además de los comandos propios de Minitab. Estos programas pueden ser utilizados en diversos tipos de aplicaciones tales como automatizar procesos de análisis de datos, resolver problemas por simulación, etc.

etc.). En la presente investigación se ha utilizado, principalmente, la distribución normal, al ser una de las distribuciones continuas más comunes en la realidad.

Tabla N° 1.4. Características de la función de densidad normal (Basado en Azarang & García, 1996).

Características	$N(\mu, \sigma)$
Función de densidad	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (-\infty \leq x \leq \infty)$
Distribución acumulada	No existe ecuación
Parámetros	Parámetro de localización: μ Parámetro de escala: σ
Rango	$(-\infty, \infty)$
Media	μ
Varianza	Σ

Cuando se usa el programa Minitab para hacer simulaciones, se requiere el ingreso previo de una serie de comandos. Para el caso específico de la investigación, se requiere primero elaborar las variables con distribución normal aleatoria, asignándole sus respectivos parámetros (media y desviación estándar), para una determinada población (número de observaciones) y sus respectivos errores de medida para cada variable. Estos comandos son, generalmente, cómo siguen:

Tabla N° 1.5. Significado de los comandos de simulación empleados con MINITAB 14 para generar variables (Fuente: Elaboración propia).

Comandos en Minitab	Significado
MTB > random 1000 c1; SUBC> normal 50 5.	Se crea una variable aleatoria de 1000 observaciones, la cual se ubicará en la columna c1. Esta variable es creada con una distribución de probabilidad normal, promedio de 50 y desviación estándar de 5.
MTB > random 1000 c2; SUBC> normal 0 5.	Se crea una variable aleatoria de 1000 observaciones, la cual se ubicará en la columna c2. Esta variable es creada con una distribución de probabilidad normal, promedio de 0 y desviación estándar de 5.
MTB > random 1000 c3; SUBC> normal 0 5.	Se crea una variable aleatoria de 1000 observaciones, la cual se ubicará en la columna c3. Esta variable es creada con una distribución de probabilidad normal, promedio de 0 y desviación estándar de 5.
MTB > add c1 c2 c4 MTB > add c1 c3 c5 MTB > name c1='puntaje verdadero' c2='error pretest' c3='error postest' c4='pretest' c5='postest'	Se suman las variables c1, c2 y el resultado se ubica en la columna c4. Se suman las variables c1, c3 y el resultado se ubica en la columna c5. Se asigna nombre a cada una de las variables creadas.
MTB > set c6 DATA> 1:1000 DATA> end MTB > code (1:500) 0 c6 c6 MTB > code (501:1000) 1 c6 c6	Se crea una variable en la columna c6, con valores de 1 a 1000. Se codifica las primeras 500 observaciones con el valor 0 (grupo control). Se codifica las restantes 500 observaciones con el valor 1 (grupo experimental).

Creadas las variables, el siguiente paso consiste en verificar los parámetros formulados y luego experimentar acorde a los objetivos teóricos del sistema, analizando los datos modificados mediante alguna técnica estadística, en este caso,

el análisis de regresión múltiple. El resultado típico del análisis de regresión emitido por el programa Minitab tiene la siguiente estructura:

```

MTB > regress c5 1 c3
The regression equation is
P. sin cambio de nivel = 48.8 + 4.72 grupos (antes-después)
Predictor          Coef  SE Coef      T      P
Constant          48.783  1.693  28.81  0.000
grupos (antes-después)  4.720  2.395   1.97  0.064

S = 5.35434    R-Sq = 17.8%    R-Sq(adj) = 13.2%

Analysis of Variance
Source          DF      SS      MS      F      P
Regression        1   111.39  111.39   3.89  0.064
Residual Error   18   516.04   28.67
Total             19   627.43

```

En primer lugar se encuentra el comando que solicita el análisis de regresión, indicando la variable dependiente (c5), la cantidad de variables independientes (1) y la variable independiente (c3). Posteriormente, el primer resultado emitido es la **ecuación de regresión**, que tiene una fórmula lineal (VD predicha = constante⁴ + coef.b⁵ x VI)⁶. Seguidamente se presenta la tabla de coeficientes de regresión, errores del coeficiente, sus respectivos valores “t” y la probabilidad de significancia para aceptar la hipótesis nula. Luego la desviación estándar del error del modelo (S) y el coeficiente de determinación (R²) que indica el porcentaje de variación en la respuesta que es explicado por el modelo. Finalmente, el resultado culmina con el análisis de varianza del modelo de regresión⁷.

⁴ Indica el valor promedio de la variable de respuesta Y cuando X es cero. Si se tiene certeza de que la variable predictor X no puede asumir el valor 0, entonces la interpretación no tiene sentido.

⁵ Indica el cambio promedio en la variable de respuesta Y cuando X se incrementa en una unidad.

⁶ Para un modelo con multiples predictors, la ecuación es: $[y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + e]$ donde Y = respuesta, X = predictores, b_k = Coeficientes de regresión poblacionales, y e = error con distribución normal, media de 0, y desviación estándar de s. La ecuación muestral es: $[\hat{Y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k]$ donde \hat{Y} = valor predico y b_k = coeficiente de regresión estimado. En la regresión lineal simple, donde se incluye un solo predictor, el modelo es: $[y = b_0 + b_1x_1 + e]$. Con los estimadores muestrales b_0 por b_0 , b_1 por b_1 , y s por s; la ecuación de regresión es entonces: $[y = b_0 + b_1x_1]$.

⁷ Donde DF = Grados de libertad; SS = Suma cuadrática, MS = Media cuadrática, F = Valores de distribución “F”; P.= Probabilidad de aceptar la hipótesis nula.

II. BALANCE TEÓRICO

Este capítulo está dedicado al desarrollo del Balance Teórico, consecuencia de la revisión y análisis de la bibliografía vigente sobre el tema de investigación. En una primera parte se revisan los aspectos históricos, definitorios, características, el ciclo de vida de los programas sociales y el contexto de la evaluación dentro del proceso de gestión social. Seguidamente, se hace una revisión bibliográfica sobre la definición y características de la evaluación de programas, su desarrollo histórico, importancia social, su papel en la política y la sociedad, así como la situación de la evaluación de los programas sociales en el Perú y un análisis crítico de los sistemas de evaluación existentes en la administración pública hasta la fecha. Finalmente, el punto siguiente está dedicado al análisis teórico de la evaluación de impacto, modalidad especializada dentro del contexto de la evaluación de programas. También, se realiza un análisis minucioso de sus conceptos, su importancia y de las principales técnicas empleadas en su ejecución; además de presentar los principios fundamentales de la evaluación de impacto.

2.1. LOS PROGRAMAS SOCIALES

El origen de los programas sociales guarda relación directa con el origen de las políticas públicas de desarrollo social, las llamadas “políticas sociales”. Las políticas sociales, tal como fueron impulsadas por los gobiernos europeos al finalizar la Segunda Guerra Mundial se orientaron a procurar condiciones de vida aceptables para el conjunto de la sociedad. Esto implicó intentar, mediante su intervención, resolver los más diversos problemas sociales: desempleo, bajos salarios, analfabetismo, escasa escolaridad, hacinamiento en viviendas, carencia de servicios básicos, enfermedades, muertes evitables, entre otros.

En las décadas de 1970-1980, este modelo entró en crisis por los insuficientes ingresos para solventar los gastos sociales. Tal situación llevó a los Estados a introducir las llamadas “reformas de primera generación” para dismantelar el intervencionismo estatal, aplicar políticas de descentralización y de privatización de empresas públicas, reducir el tamaño del Estado y emplear el mecanismo de mercado como nueva centralidad. En este contexto neoliberal, el

Estado debía limitarse a garantizar las condiciones para que las empresas privadas ofrecieran los bienes y servicios públicos, que antes habían constituido un componente fundamental de sus funciones (Cardozo, 2003).

En la década siguiente (los '90) los problemas sociales de América Latina continuaron creciendo: había aumentado el número de habitantes en condiciones de pobreza y de pobreza extrema (253 millones de personas, que constituían 50% de la población en 1990, frente a 38% que existían en 1980); éste se había concentrado en el medio rural y en grupos particularmente vulnerables, como niños (60%), mujeres, indígenas, ancianos y discapacitados. La pobreza se reforzó con situaciones de desigualdad, marginación y exclusión (Klikberg, 1997). Ante esta situación, los Estados comenzaron a aplicar las “reformas de segunda generación”, centradas en la eficiencia de la provisión de servicios, la efectividad para mejorar su cobertura y calidad, así como el reforzamiento de las condiciones para el desarrollo de la actividad privada (Nickson, 2002).

La preocupación por la innovación y la eficiencia en la gestión ha llevado a los gobiernos, entre otras estrategias, a la focalización de sus intervenciones sociales. Si bien la política social sigue planteándose el objetivo de mejorar el nivel de vida de las personas, el acento se ha puesto actualmente en el combate a la pobreza extrema, lo que provoca que otros grupos sociales (trabajadores, por ejemplo) vean disminuir los servicios públicos que antes satisfacían sus necesidades (principalmente de salud, educación y vivienda), tanto en cantidad como en calidad. Esta nueva política social, sea que asuma objetivos de corto plazo (vinculados a la transferencia de recursos para paliar la situación actual de los pobres en relación con la satisfacción de sus necesidades elementales más inmediatas) o que se plantee objetivos de largo plazo (por medio de la inversión en capital humano y social, en cuyo caso se habla de políticas de desarrollo social), exigen la inclusión de medios de evaluación sistemáticos.

En este contexto, la introducción de la evaluación de los programas sociales gubernamentales ha sido muy difícil en todos los países, especialmente en los menos desarrollados. Si bien en los países escandinavos y otros como Suiza, Estados Unidos o Canadá, la idiosincrasia, el nivel de desarrollo y el sistema político han permitido mayores avances (Aucoin, 2005), también existen otras naciones que aun gozando de altos grados de desarrollo como Alemania, Reino Unido o Francia, han enfrentado importantes resistencias para su aplicación. Generalmente las dificultades aumentan en países menos desarrollados y se agudizan en casos como el nuestro, cuya historia política ha estado signada por un agudo presidencialismo y, hasta hace poco tiempo, por el predominio de una dictadura donde no se rindieron cuentas a la ciudadanía y se desatendía a la cultura de la evaluación.

En el aspecto positivo, México, por ejemplo, ha hecho importantes avances en la materia a partir del año 2000, debido principalmente a la contratación de evaluaciones externas (Cardozo, 2003). Su mejor evaluación fue la realizada por expertos extranjeros al programa “Progresá” la cual se hizo gracias a una disposición de recursos, de la cual carecen las evaluaciones realizadas en nuestro país, impidiendo a estas últimas contar con la fiabilidad y validez deseadas.

En todo caso, de lo dicho, resulta necesario, además de mayores recursos financieros, generar mayor desarrollo de las capacidades para evaluar programas, tanto en lo que se refiere a especialistas técnicos como a los distintos actores sociales participantes en los mismos, lo que constituye un desafío para nuestras universidades en los niveles de postgrado, educación continua y servicios de asesoría.

2.1.1. Definición y características

Existen muchas definiciones de lo que son programas. En el término amplio de la palabra, en general, los diversos autores entienden que un programa es un plan o sistema bajo el cual una acción está dirigida hacia la consecución de una meta. Los autores coinciden en afirmar que los programas sociales son refuerzos cuidadosamente planificados, comprensivos y sistemáticos para lograr objetivos claramente articulados. En la siguiente tabla se presenta una serie de definiciones de diversos autores sobre los programas sociales.

Tabla N° 2.1. Definiciones sobre programas sociales (Fuente: Elaboración propia).

Autores	Año	Definición
Stame, N.	2003	“Los programas pueden ser entendidos como las acciones que tienen el propósito de obtener un cambio, y tienen que ser implementadas con los medios dados para obtener resultados esperados dentro de un plazo dado”. (Pág. 37) (Traducción propia).
Jariot, M.	2001	“...proceso de intervención que se inicia con una evaluación y priorización de necesidades educativas, de modo que a partir de esta evaluación inicial se marquen los objetivos educativos del programa destinado a potenciar aquellos aspectos más deficitarios o relevantes de un grupo concreto de personas. Intervención que se realiza dentro de un contexto determinado, llevado a cabo mediante el diseño, planificación e implantación de una serie de actividades que conforman el proceso de implementación del programa, apoyadas y fundamentadas en un marco teórico de referencia. Este proceso, además, deberá estar sometido a una evaluación constante con el fin de mejorar la intervención y valorar el logro de objetivos”. (Pp. 113-114).
Development Assistance Comité - DAC	2002	(Programa de desarrollo) “...es una intervención de duración limitada que comprende múltiples actividades que pueden abarcar varios sectores, temas y/o zonas geográficas” (Pág. 30).
United States General Accounting Office (GAO)	1998	“Un programa debería ser una actividad, proyecto, función o política que tiene un propósito determinado o un grupo de objetivos”. (Pág. 3) (Traducción propia).
De Miguel	2000b	Es el plan, proyecto o servicio mediante el cual se diseña, organiza y se pone en práctica un conjunto de acciones y recursos materiales dirigidos a la consecución de una meta.
Fernández	1999	El programa es “...el conjunto específico de acciones humanas y recursos materiales diseñados e implementados organizadamente en una determinada realidad social, con el propósito de resolver algún problema que atañe a un conjunto de personas” (Pág. 477).
Alfie, E.	2006	“Fundamentalmente son acciones, en algunos casos desorganizadas, tendientes a satisfacer las urgencias, resolver demandas y problemáticas presentadas en cada campo específico social; situación que impide una mejor utilización de los recursos y una adecuada referencia a los resultados” (Pág. 1).
Fernández-Ballesteros	1996	“Un programa se interpreta como una intervención o tratamiento, es decir, como un conjunto de manipulaciones que han sido programadas para ser implantadas en una determinada realidad social” (Pág. 28).

Autores	Año	Definición
Bartolomé, M.	1990	“Un conjunto de estrategias y elementos orientados a la consecución de determinadas metas y objetivos” (Pág. 39).
Pérez, R.	1992	“Plan sistemático de intervención, diseñado y elaborado intencionalmente para la consecución de unos objetivos de mejora”. (Pág. 46).
Fink	1993	“Los sistemáticos esfuerzos realizados para lograr objetivos pre-planificados con el fin de mejorar la salud, el conocimiento, las actitudes y la práctica” (Pág. 2).

En concordancia con las definiciones presentadas se dirá que un programa es un conjunto específico de acciones humanas y recursos materiales, diseñados e implantados organizadamente en una determinada realidad social, con el propósito de resolver algún problema que atañe a un conjunto de personas. Todo programa social persigue mejorar algunas condiciones de vida que pueden ser de salud, infraestructura, educación, asistencia, justicia, entre otras. En la literatura el término “programa social” se usa indistintamente de forma genérica para cualquiera de estas áreas.

En otro aspecto, algunos autores plantean que existen diferencias entre “proyectos” y “programas”. Aunque en términos políticos y legales programas y proyectos sociales tienen una diferencia de jerarquía, sin embargo, en términos metodológicos, no existe una diferencia sustancial entre ellos, son indistintos. En efecto, no existe en la literatura una diferenciación entre los elementos metodológicos asociados con la evaluación de políticas, de programas y de proyectos. En el terreno práctico, las diferencias entre políticas y programas son inexistentes; pues un programa nacional de gran envergadura puede ser considerado como la expresión real de una política e incluso puede tomar el nombre de la política que lo generó (Ej. programa de lucha contra las drogas) o también puede llamársele proyecto (Ej. proyecto de desarrollo rural). Al ser metodológicamente indistintos, entonces, es importante aceptar que la evaluación concierne directamente con cualquiera de estos niveles (políticas, programas y proyectos), y que las técnicas de la evaluación son aplicables a cualquiera de ellos (Weis, 1998). En cambio, en términos políticos y legales, los proyectos son definidos como la mínima unidad de asignación de recursos para el logro de uno o más objetivos específicos (Sanín, 1999). Tienen una población objetivo y localización espacial determinadas y un tiempo de inicio y finalización previstos (Naciones Unidas, 1998). Los programas, por su parte, se definen como conjuntos de proyectos que persiguen los mismos objetivos (Frechtling, Stevens, Lawrenz & Sharp, 1993), tal como puede apreciarse en el gráfico siguiente, donde se esquematiza esa relación jerárquica⁸.

⁸ Actualmente la normatividad nacional considera explícitamente esta relación jerárquica. Por ejemplo, el Sistema Nacional de Presupuesto Público considera una relación funcional programática donde los programas supeditan a los proyectos y actividades. Para mayor detalle véase la estructura funcional programática.

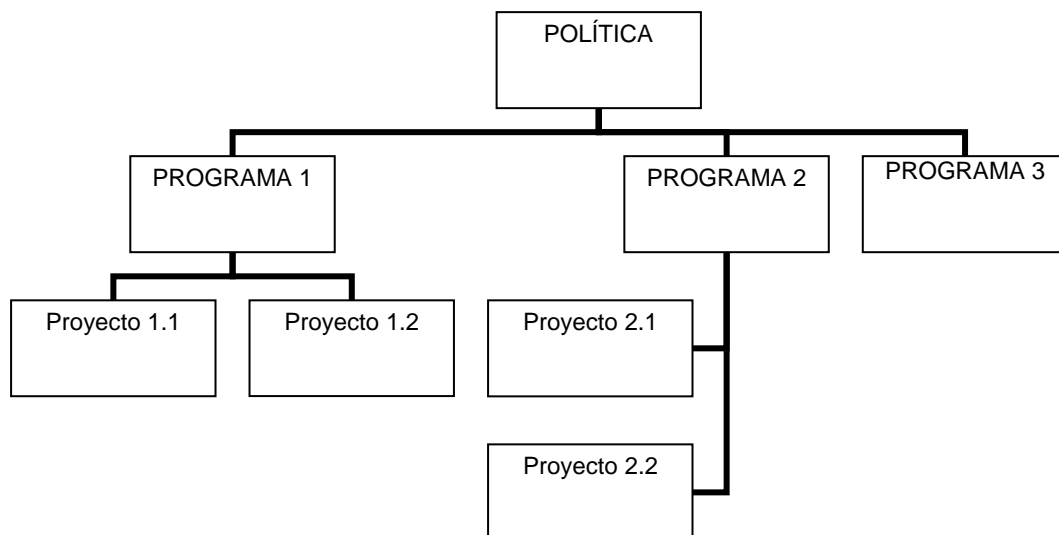


Figura Nº 2.1. Distinción legal entre políticas, programas y proyectos sociales (Basado en Naciones Unidas, 1998).

Los programas sociales son la materialización de la política social, elaborados para satisfacer las necesidades de la población. En el caso de los problemas sociales, se opera pretendiendo satisfacer necesidades productos de “demandas insatisfechas”. Las necesidades básicas insatisfechas de la población que presenta mayores carencias, pueden, así, ser concebidas como problemas y los proyectos sociales como soluciones a los mismos. De ahí que los extremos polares de un proyecto social puedan ser definidos como una “situación actual” en la que aparece un problema y una “situación deseada” en la que el mismo ha sido eliminado o ha disminuido como resultado de la intervención que el programa o proyecto representa.

Queda claro entonces que indistintamente cómo se le refiera (programa o proyecto), metodológicamente hablando el programa social es un conjunto organizado de acciones y recursos, diseñados e implantados sistemáticamente en una determinada realidad social, para resolver algún problema que atañe a una población y mejorar su calidad de vida en algún aspecto.

En otro aspecto, se acepta con unanimidad que los programas o proyectos sociales tienen un ciclo de vida universal (CTA, 2005; ONU, 1998; Abdala, 2001; BID, 1997, ILPES & CEPAL, 2003; Sanín, 1999). El ciclo de vida describe la secuencia de estados, etapas y acciones que se deben realizar en un programa o proyecto para materializarlo. De acuerdo a las Naciones Unidas (1998) es posible distinguir tres estados generales en el ciclo: Preinversión/Formulación; Inversión, y Operación. Dado que esta nomenclatura tiene su origen en los proyectos productivos, la etapa de inversión es considerada un elemento central. Sin embargo, en los proyectos sociales puede que no exista inversión en activos físicos, pero sí en activos intangibles. Los estados del ciclo de vida son los siguientes:

1. **Preinversión/Formulación.** Abarca la identificación del problema, el diagnóstico de la situación, la definición de áreas de intervención, la elaboración de alternativas de solución y el análisis comparativo de las

mismas para seleccionar la óptima. Este proceso se lleva a cabo con niveles de profundidad crecientes, reflejado en sus etapas: Idea del proyecto, Perfil, Prefactibilidad y Factibilidad. Avanzar en las etapas puede ser descrito como una adquisición progresiva de certidumbre. Esto es, la incertidumbre, disminuye, en la medida en que se afrontan los costos adicionales de profundizar los estudios que suponen el pasar de una a otra etapa.

2. **Inversión.** Es la instancia de adquisición de los bienes de capital requeridos para la operación del proyecto (compra de terrenos, construcción de edificios y equipamiento). En el campo de lo social, es necesario diferenciar dos situaciones típicas: Proyectos que requieren inversión (postas, hospitales, escuelas, etc.) y aquellos en los que ésta no existe o es de carácter marginal (proyectos de capacitación, asistencia técnica, campañas de salud, etc.). Cuando la inversión no es requerida, se pasa directamente del estado de la preinversión a la operación del proyecto.
3. **Operación.** Comienza cuando el proyecto inicia su proceso productivo (compra de insumos, articulación de recursos y generación de productos) para entregar los productos (bienes o servicios) a la población objetivo.

Se encuentra también que a cada estado del ciclo de vida se le asocia un tipo de evaluación. En la siguiente tabla se resume la relación:

Tabla N° 2.2. Tipos de evaluación y estados del proyecto (Fuente: Elaboración propia).

Estado del proyecto	Tipo de evaluación	Funciones	Cuando Utilizarla
Planificación del Programa (Pre-inversión / Formulación)	Evaluación Ex - ante	Evaluar la viabilidad y sustentabilidad del programa en términos financieros, políticos e institucionales, para priorizar y seleccionar las alternativas de componentes y actividad que maximicen el impacto de la inversión. Es la herramienta de evaluación que permite tomar la decisión de implementar el programa.	Se diseña y se realiza durante la Planificación.
Ejecución del Programa (Operación)	Monitoreo	Indagar y analizar el grado en que las actividades realizadas y los resultados obtenidos responden a lo planificado. Busca detectar oportunamente eventuales deficiencias, obstáculos y/o necesidades de ajuste en la planificación y ejecución	Se diseña durante la Planificación y se desarrolla permanentemente, durante toda su ejecución.
	Evaluación Diagnóstica	Evaluar en profundidad aspectos o problemas del programa relacionados con la planificación, ejecución y logros de resultado, con la perspectiva institucional.	Cuando existe la necesidad de profundizar el conocimiento de su desempeño.
	Evaluación desde la Perspectiva de los beneficiarios	Evaluar aspectos o problemas del programa relacionados con la planificación, ejecución y logro de resultados, desde la perspectivas de los beneficiarios, incorporando su participación como aspecto central.	Cuando existe la necesidad de profundizar el conocimiento de su implementación y sus efectos, según sus propios beneficiarios.

Estado del proyecto	Tipo de evaluación	Funciones	Cuando Utilizarla
Finalización del Programa	Evaluación de Resultados	Evaluación del grado de cumplimiento final de los objetivos específicos del programa, en términos de efectividad, eficiencia y eficacia.	Se diseña en la planificación del programa y se realiza inmediatamente después de finalizada la ejecución del mismo.
	Evaluación de Impacto	Identificar y explicar la modificación del valor inicial de los indicadores del problema que dio origen al programa (línea de base).	Se diseña durante la planificación del programa y se ejecuta un tiempo después de finalizada la ejecución del mismo.

Así, en la preinversión, los proyectos se formulan para que puedan ser evaluados *ex-ante*. Esta evaluación se utiliza como criterio de decisión, permitiendo determinar si el proyecto satisface o no los requisitos de rentabilidad o costo/impacto establecidos y, al mismo tiempo, cuál es la alternativa óptima para su implementación. En la etapa de inversión, corresponde realizar el seguimiento (**monitoreo**) en la misma. Ésta se lleva a cabo durante la ejecución midiendo el avance de las obras y los recursos empleados en función del cronograma establecido. Su función es detectar desviaciones con respecto a la programación inicial y, en el caso que existan, a qué y a quiénes son imputables. La **evaluación ex-post** se realiza durante la fase de operación del proyecto. No supone que éste haya concluido. Comienza al mismo tiempo que la entrega de los bienes o servicios a la población beneficiaria. Las principales variables que se tienen en cuenta en la evaluación ex-post de un programa se relacionan con el análisis de la población objetivo, la producción del programa y su capacidad de gestión (Sanín, 1999). El último eslabón de la evaluación ex-post lo constituye la evaluación de impacto.

Actualmente, el Sistema Nacional de Inversión Pública (SNIP) del Perú considera que el ciclo del proyecto de inversión tiene tres etapas: preinversión, inversión y post-inversión; y tan igual como en el modelo general, se les asocia diversos tipos de evaluación. (Para mayor detalle véase el punto 2.2.2.5.1).

2.1.2. La evaluación y la gestión social

Los programas son la traducción operacional de las políticas sociales. Las políticas sociales, o también conocidas como políticas públicas, son decisiones que se toman en el ámbito político gubernamental para atender una demanda social, ya sea de carencia o ya sea de necesidad de desarrollo. En general, la literatura científica (Arce, 2004) resalta algunas características propias de una política pública “óptima”, o propias de un “buen gobierno”: a) que sean representativas, b) que tengan demanda social prioritaria, c) que mejoren la eficiencia de procesos y gestión, d) que estén integradas con el sistema de políticas, e) que consideren en su diseño a la evaluación y f) que haya control en los resultados.

Esta visión es coherente con las tendencias teóricas recientes sobre la Gerencia Social y la Administración Pública, las cuales enfatizan en las ventajas de promover una gestión orientada a resultados, en contraste con una orientación

tradicional que hace hincapié en los procedimientos y los procesos de gestión (Barzelay, 2001; Caiden, 1998; Cunill & Ospina, 2003; CLAD, 1999; Moctezuma & Roemer, 1999; Olías de Lima, 2001; Osborne & Gaebler, 1994; Ramio, 2001).

Hoy en día se entiende que las políticas públicas constituyen los focos sobre los que se debe fijar la atención para introducir una mayor racionalidad en el gasto social a través de acciones que aseguren su impacto y eficiencia (Waissbluth, 2002; Osborne & Gaebler, 1994). Esta mayor racionalidad plantea dos ejes estratégicos: la evaluación y la gestión social. La evaluación se preocupa del cuánto, de contrastar el logro alcanzado o estimado de los objetivos de una política, programa o proyecto, con las metas propuestas y el costo de su implementación. La gestión aborda el cómo, la manera en que se estructuran los diversos componentes del programa y la dinámica de los actores, funciones y procesos que buscan dar cuenta de las metas planteadas de producción e impacto.

En el plano específico de los programas sociales, se afirma que el proceso de gestión de los programas sociales comprende las funciones de: Organización, Dirección, Programación, Ejecución, Monitoreo y Evaluación de Impacto, a partir de las que se construye el acrónimo difundido por las Naciones Unidas (1998) como “ODPEME”. Se espera que en cada una de estas etapas se realice una actuación eficiente, sin embargo, se han documentado algunos problemas que limitan la gestión eficiente de las políticas sociales en cada una de ellas (CEPAL, 1998). A continuación, en la siguiente tabla, se presenta una breve descripción de cada una de ellas:

Tabla N° 2.3. El proceso de gestión social, características y problemas asociados (Basado en CEPAL, 1998).

Gestión	Características	Problemas asociados
Organización	<ul style="list-style-type: none"> Estructura e integra los recursos y órganos responsables de la gestión, estableciendo sus relaciones y atribuciones. Permite una utilización equilibrada de los recursos. Relaciona el trabajo (incluyendo sus instrumentos y localización) con el personal que lo debe ejecutar. Conforma la estructura social y material para el proceso de producción y distribución. El ámbito de esta función depende del grado de exhaustividad de la formulación. Mientras más detallada sea la formulación, más adelantada estará la organización de las actividades. Introduce las modificaciones necesarias a la organización original, a la luz de la información emanada de la operación del programa, cualquiera sea el momento y la periodicidad de dichos cambios. 	<ul style="list-style-type: none"> Los programas y proyectos no siempre están bien definidos, no responden a enunciados explícitos de políticas, tienen horizontes de realización irreales, carecen de una orientación hacia el logro de resultados y de criterios para la medición de la eficiencia y del impacto, con escasa focalización y la persistencia de subsidios encubiertos de los sectores pobres a los de mayores ingresos. Competencias del sector social fragmentadas entre diversas dependencias u organismos, programas e instituciones, con capacidad insuficiente para dar solución integral a los problemas, exigiendo una continua coordinación y colaboración. Escasas experiencias de descentralización. Técnicas de gestión y formas de organización del trabajo inadecuadas e inconsistentes con la naturaleza de las tareas de los programas y proyectos.
Dirección	<ul style="list-style-type: none"> Orienta el comportamiento de los 	<ul style="list-style-type: none"> Excesiva burocracia, con resistencias

Gestión	Características	Problemas asociados
	<p>operadores del proyecto a la consecución de sus objetivos.</p> <ul style="list-style-type: none"> • Comprende la conducción, supervisión, sincronización de las acciones individuales y coordinación de las actividades que están definidas para ser desarrolladas por las distintas unidades de la estructura. • Se expresa mediante un particular estilo de liderazgo. Conduce personas y equipos (socialización de valores, comunicación interpersonal y participación). • Se encarga de la conducción estratégica del programa, de la prevención y de la resolución de problemas y conflictos (por áreas, grupales y personales). • Realiza un balance de las amenazas, oportunidades, fortalezas y debilidades. • Guía, supervisa y coordina las actividades definidas en la organización bajo criterios de desempeño (eficiencia interna e impacto externo). 	<p>al cambio y a la innovación, con fragmentación de tareas, numerosas capas jerárquicas, funciones departamentales aisladas e incomunicación entre los decisores y los operadores.</p> <ul style="list-style-type: none"> • Ausencia de incentivos para el mejoramiento de la gestión, con asignaciones presupuestarias a las instituciones que gestionan los programas y proyectos que no están basadas en el desempeño. • Funcionarios poco motivados, que perciben bajos salarios, con insuficiente actualización y capacitación.
Programación	<ul style="list-style-type: none"> • Establece la secuencia cronológica de las actividades a realizar, optimizando los recursos asignados e identificando a los responsables y las interacciones requeridas para su ejecución. • Asigna los recursos presupuestarios y realiza la gestión financiera. Se basa en la organización y formulación. 	<ul style="list-style-type: none"> • La programación no considera acciones contingenciales.
Ejecución	<ul style="list-style-type: none"> • La ejecución corresponde al desarrollo de las actividades enmarcadas en los procesos (sustantivos y de apoyo) para alcanzar los objetivos de producto e impacto buscados por el programa. 	<ul style="list-style-type: none"> • Dificultades en la participación social de los grupos prioritarios de la política social, los más pobres, que carecen de información, influencia y organización, están dispersos y no cuentan con mecanismos para el ejercicio de sus derechos o la movilización de redes sociales para influenciar las decisiones de la burocracia.
Monitoría	<ul style="list-style-type: none"> • Es la observación concomitante de la ejecución. Compara lo realizado con lo programado, con estándares (promedios u óptimos) y con otros programas. • La información resultante de la monitoría se canaliza a través de distintos circuitos de información (ejecución, programación, dirección y organización) con el objetivo de producir acciones correctivas en dichas funciones y procesos. • Su implementación requiere indicadores, periodicidad de recolección de datos, fuentes de información y su sistematización. 	<ul style="list-style-type: none"> • Ausencia de sistemas de información fiables y precisos para monitorear y evaluar la implementación e impacto de las políticas sociales, con limitado uso de tecnologías de información, produciéndose una baja calidad/producción de los servicios sociales.
Evaluación de impacto	<ul style="list-style-type: none"> • Tiene por finalidad determinar la magnitud de los cambios producidos 	<ul style="list-style-type: none"> • Escasez de expertos en políticas sociales, con capacidad técnica para

Gestión	Características	Problemas asociados
	<p>por los programas, tanto buscados como no buscados.</p> <ul style="list-style-type: none"> • A la luz de los resultados generados, es posible realizar acciones correctivas en los objetivos y procesos implementados. • Dicha información es un insumo para la función de organización, la formulación y evaluación ex-ante de otros programas y las políticas del sector. 	<p>el análisis de programas y proyectos.</p> <ul style="list-style-type: none"> • Escasez de recursos porque no se la considera importante.

Así, analizar la gestión social, etapa por etapa, y garantizar la disminución de los problemas asociados permitiría un mejor control de resultados y una mayor eficiencia funcional. De igual forma, al ser un proceso inserto dentro del sistema de gestión del programa social, la evaluación no puede ser pensada como una etapa desconexa o ajena a ella. En efecto, esta estrecha relación condiciona una gestión preocupada por la evaluación desde su planificación, durante su ejecución y después de su aplicación. Este proceso coincide con el ciclo de vida de los programas y se torna una constante obligatoria.

2.1.3. Los programas sociales en el Perú

Desde la década del noventa la política social peruana se ha orientado a erradicar la pobreza extrema⁹. Si bien se alcanzaron importantes avances en este periodo, la política social nunca logró transformarse en un estrategia integrada a largo plazo que permitiera una asignación ordenada y eficiente de los recursos hacia las principales prioridades sociales y tuviera un impacto importante y sostenible sobre los indicadores sociales (Shack, 2000).

Durante los primeros años del nuevo siglo, la situación de los programas sociales tampoco ha mejorado. A pesar de que se han elaborado diversos planes que buscan mejorar la gestión de los programas, estos simplemente no se han materializado. Por el contrario, se han introducido nuevos problemas referidos a la gestión de los programas como resultado de una creciente burocratización o “partidarización” de los mismos. De esta forma, a la ineficiencia de los programas se le han sumado nuevos problemas, los cuales se han visto incrementados durante el proceso de regionalización (Du Bois, 2004).

Du Bois (2004), al evaluar la situación de los programas sociales en el Perú, concluye que

“...estos todavía siguen constituyendo una serie de esfuerzos inconexos que, en la mayoría de los casos, no cumplen con los objetivos para los cuales fueron creados. Más aún, tampoco han tenido un impacto significativo sobre los niveles de ingreso de la

⁹ El gasto social per cápita alcanzó un crecimiento real de casi 100% entre 1991 y el 2000 (pasó de US\$ 90 a US\$ 180). Según CEPAL (2003) el Perú fue el tercer país con mayor crecimiento real del gasto social per cápita entre 18 países de América Latina y el Caribe, sólo por detrás de Paraguay y Colombia.

población que es beneficiaria de estos programas, con excepción de los programas de inversión social y productiva en zonas rurales, por lo que en realidad éstos simplemente no constituyen una inversión en capital humano. El ejemplo más dramático es el caso de los programas de alimentación y nutrición. Así, mientras entre 1997 y el año 2000, el Estado invirtió aproximadamente US\$ 1000 millones en este tipo de programas, la tasa de desnutrición se mantuvo estable” (P.1).

En este escenario de escasa efectividad del gasto en programas sociales, la inacción gubernamental del último lustro ha sido la norma. Al respecto, sólo se pueden resaltar dos aspectos positivos:

- El reordenamiento a nivel ministerial de los programas de carácter alimentario y nutricional (aunque sin reformular su gestión).
- La creación de los programas de empleo temporal (aunque con limitaciones de focalización, es decir, no llegar a las personas efectivamente desempleadas sin otras posibilidades de generar ingresos).

A pesar de lo dicho, la focalización de los programas sociales durante la década pasada mostró severas deficiencias, al no alcanzar efectivamente a los más pobres, permitiendo que una gran cantidad de no pobres se vea beneficiada, con un nivel de filtración (% de beneficiarios no pobres que accedieron a los programas) que superaba el 60% para los años 1996 y 1998 (Chacaltana, 2001). Esta ineficiencia era mayor en los programas de alimentación y nutrición (Vásquez et al, 2000).

Además, la intervención del Estado mediante los programas de ayuda alimentaria generó conductas “clientelistas” por parte de los grupos de interés ya formados (clubes de madres, comedores populares), los cuales se vieron acrecentados a partir de 1998, cuando el PRONAA fue utilizado con fines proselitistas electoreros (Saavedra y Pasco-Font, 2001).

Otros problemas de gestión importantes fueron la superposición de programas (generando duplicidad de gastos¹⁰) y la casi absoluta ausencia del control sobre el impacto de los programas sociales, hecho que impedía la reformulación o fortalecimiento, según el caso (Du Bois, 2004).

A partir del nuevo milenio, el gobierno ha optado por iniciar el proceso de descentralización de los principales programas sociales, sin embargo, lo ha hecho sin antes reformularlos, reestructurarlos o fusionarlos. Esta situación ha hecho que los problemas de gestión se mantengan.

Con fines descriptivos, a continuación se desarrolla una cronología breve y comentada, pero no exhaustiva, de la evolución de los principales programas sociales en el Perú, durante las últimas dos décadas.

¹⁰ Por ejemplo, las políticas de nutrición y alimentación fueron implementadas por el Ministerio de Agricultura, del Ministerio de Educación, Ministerio de Pesquería, Ministerio de Salud, Ministerio de Promoción de la Mujer y Desarrollo Humano, del PRES a través de FONCODES y del Ministerio de Economía y Finanzas a través de los municipios. Vásquez y Mendizábal (2002) señalan que la duplicidad de gasto en el año 2000 ascendió a US\$ 100 millones, un 40.6% del gasto total de programas alimentarios y 11.5% del total de programas de lucha contra la pobreza.

Tabla N° 2.4. Principales acontecimientos referidos a las políticas de gestión de los programas sociales en el Perú: 1990-2002. (Fuente: Elaboración propia).

Año	Programa o lineamiento	Objetivo	Comentarios
1990	Programa de Emergencia Social (PES)	Proteger a la población más vulnerable. Se centró en temas alimentarios y de salud.	El PES fracasó por importantes restricciones fiscales y por el reducido equipo de trabajo. Este fue reemplazado por el Sistema Nacional de Desarrollo y Compensación Social (SNDCS).
1991	Fondo de Compensación y Desarrollo Social (FONCODES)	Reemplazar al SNDCS y constituir un fondo de inversión basada en la demanda, para financiar proyectos de infraestructura, apoyo social y desarrollo productivo.	Ha funcionado muy bien, logrando cobertura y generando participación activa de la población (Alcázar y Wachtenheim, 2001; Francke y Espino, 2001).
1992	Programa Nacional de Asistencia Alimentaria (PRONAA)	Otorgar apoyo alimentario en zonas rurales y urbano marginales.	Se ha mostrado ineficaz, por su escasa focalización y gestión con las bases (Vásquez, et al, 2000) así como la generación de clientelismo (Saavedra y Pasco-Font, 2001).
1992	Ministerio de la Presidencia (PRES)	Centralizar el manejo de distintos programas destinados a atender a la población más pobre. Se centralizó la gestión de los programas como FONCODES, PRONAA, FONAVI, INFES, entre otros.	Aunque esta integración permitió mayor discrecionalidad gubernamental, la desarticulación y superposición de los programas sociales nunca fueron superadas (Vásquez et al, 2000). Además, la Comisión Intersectorial de Asuntos Sociales (CIAS) del PRES nunca logró funcionar de manera adecuada (Du Bois, 2004). Lo mismo ocurrió a nivel regional con las CTAR.
1993	Se elaboraron los Lineamientos de la Política Social, centrándose en la focalización y la participación activa de la población.	Intentar corregir la desarticulación y superposición de los programas.	Sirvieron de lineamientos de acción durante toda la década.
1994	Programa de Focalización del gasto social básico.	Mejorar la calidad del gasto social en cuatro sectores: nutrición, salud, educación y justicia.	Fue un paso importante, dando origen a estrategias certeras como el Mapa de Pobreza Distrital, beneficiando al área rural (Francke y Espino, 2000; Chacaltana, 2001; Valdivia y Dammert, 2001).
1994	Programa de Salud Básica para Todos	Asegurar la provisión de un conjunto de intervenciones básicas en los establecimientos de primer nivel de atención.	De la misma forma se implantó otros programas destinados a atender aspectos específicos como la salud materno-infantil y la planificación familiar. La política del sector también dio importancia a la infraestructura y equipamiento básico. Se aumentó así la cobertura en un 100% (Ricse, 2000).
1994	Programa de Administración Compartida (PAC) y los Comités Locales de Administración Compartida (CLAS).	Permitió al MINSA fomentar las redes de administración y gestión de los centros y puestos de salud básicos.	La introducción del CLAS produjo mejoras en términos de calidad y equidad (Du Bois, 2004).

Año	Programa o lineamiento	Objetivo	Comentarios
1994	Programa de Mejoramiento de la Calidad de la Educación Peruana (MECEP) en convenio con el Banco Mundial	Hacer frente a los problemas de calidad e la enseñanza, modernizar la administración educativa y sustituir y rehabilitar la infraestructura educativa.	Sobre ello, el MINEDU desarrolló varios programas, la mayoría ineficientes, salvo el programa de universalización de la educación primaria.
1996	Mapa de Pobreza Distrital	Se estableció como un instrumento de focalización geográfica para la inversión en infraestructura.	PRES y FONCODES usaron efectivamente esta herramienta.
1997	Programa de Administración de Acuerdos de Gestión (PAAG)	Tiene como objetivo que los propios hospitales realicen una mejor asignación de sus recursos, mejorando la calidad de los servicios debido a la mayor autonomía recibida.	Este plan piloto fracasó, ya que no se logró cambiar la estructura administrativa de los hospitales (Arroyo, 2001).
1999	Programas Sociales Protegidos (PSP)	Lista de programas que no podían estar sujetos a recortes presupuestarios y evitar que su ejecución se vea perjudicada.	No se estableció mecanismo alguno que pudiera asegurar esta protección presupuestaria. Así, finalmente, no se cumplió con los objetivos y muchos programas se vieron recortadas, en especial la inversión de carácter social y productivo (Saavedra y Pasco-Font, 2001).
2001	Mesas de Concertación para la Lucha contra la Pobreza, la cual se materializó en la Carta de Política Social.	Convertirse en un medio de concertación en torno de la forma más idónea de luchar contra la pobreza en cada región, provincia y distrito del país.	Acciones realizadas en el periodo de transición, como un medio para ordenar los programas sociales.
2001	Programa a Trabajar Rural (FONCODES). Programa a Trabajar Urbano (Ministerio de Trabajo y Promoción Social).	Creación de trabajos orientados a la rehabilitación, mantenimiento y puesta en valor de la infraestructura social de las comunidades.	Este programa ha generado 201,643 empleos temporales hasta el 2003. Sin embargo, Chacaltana (2001) demuestra que el impacto de programa ha sido deficiente.
2002	Bases para la Estrategia de Superación de la Pobreza y Oportunidades Económicas para los Pobres (Elaborado por CIAS)	Establecer una estrategia integral que defina un nuevo tipo de gestión social, sustentada en la transparencia y la participación ciudadana, vía descentralización.	Para lograrlo se requería una profunda racionalización de los programas, así como la reestructuración ministerial y la implementación de un sistema de seguimiento y evaluación de impacto de la inversión social. Esto generó la desaparición del PRES y la creación del MIMDES (FONCODES y PRONAA pasaron a ser parte del MIMDES), pero sin mayores cambios en gestión y evaluación.

Así, tal como se observa en la Tabla 2.4, desde fines de la década pasada existe una tendencia consensual sobre la necesidad de reformular los programas sociales, sobre todo si se considera la cantidad significativa de gasto sin mayor impacto en el beneficio de la población (Chacaltana, 2001, Du Bois, 2004). Sin embargo, este interés por una gestión eficiente y eficaz se ve truncado por el mismo Estado, quien, lamentablemente, tiene hasta ahora hábitos contrarios a una administración moderna. A decir de Du Bois (2004):

“...los programas sociales han sido burocratizados en los últimos años, incrementándose en 126% el gasto en personal y obligaciones de los programas alimentarios y nutricionales entre el 2001 y el 2003; mientras que, en caso de FONCODES, el aumento del gasto administrativo entre los mismo años ha sido de 310%. Sin embargo, no sólo la burocracia está aumentando, sino que los recursos destinados simplemente no llegan a los hogares beneficiados. En el caso del Vaso de Leche, casi un 30% de los recursos se pierde entre los Comités y los hogares” (P.144).

Considero que esta situación mejoraría mucho si se incluyesen sistemas de reforma basada en la administración moderna y si se utilizasen sistemas continuos de evaluación y monitoreo. Tema del que se discutirá a continuación.

2.2. LA EVALUACIÓN DE PROGRAMAS

Hace cuatro décadas, el investigador social Donald Campbell sostenía que la sociedad podía reformarse a sí misma mediante retroalimentación proveída por la experimentación social. Para Campbell, la evaluación es un ejercicio imprescindible para que los políticos y tecnócratas decidan sus cursos de acción. En efecto, existe unanimidad en afirmar que la evaluación social es una disciplina que favorece la decisión sobre políticas públicas, la rendición de cuentas y el conocimiento de la ciudadanía sobre las acciones de sus gobernantes. Por un lado, es un efectivo instrumento de aprendizaje para la ciudadanía, pues se constituye en un excelente medio para controlar al gobierno. Pedir cuentas a los gobernantes requiere conocimiento e información y éstos se consiguen con procesos evaluativos que develan ante la opinión pública y la ciudadanía los resultados de las políticas gubernamentales. Por otro lado, para el gobierno, la evaluación es un instrumento útil para diseñar políticas, particularmente aquellas orientadas al mejoramiento de la condición de vida de la población, evaluar su gestión, producir ajustes convenientes y optimizar el aparato institucional (Bovens, Hart & Kuipers, 2005).

Por otro lado, además de ser un mecanismo instrumental, la evaluación representa también una cultura de gobierno que no se limita a ser un proceso meramente técnico. Aunque es una herramienta práctica para tomar decisiones gubernamentales y generar control político-social por la ciudadanía, su utilidad más trascendental radica en ser un proceso deliberativo que permite analizar la acción específica del Estado, el contexto en el que se desarrolla esta acción, su relación con el sistema institucional y político vigente, y cuestionar la ideología dominante presente en el accionar político. De hecho, la aplicación de programas evaluativos desde diversos frentes permite cuestionar, de manera útil y constructiva, el marco ideológico que soporta la acción política así como su adecuación con el contexto social y político (Solarte, 2002).

Aunque la evaluación puede aplicarse sobre cualquier tipo de organización, su mayor campo de acción es y será el Estado. Cuando se trata de la acción gubernamental, la evaluación integra componentes del sistema político, social,

económico, e institucional, convirtiéndola en una valiosa herramienta de análisis y aprendizaje.

Durante las últimas cuatro décadas, se han ido conformando diversas corrientes teóricas que, tanto en Europa como en los Estados Unidos, orientan la evaluación hacia diversos usos tales como la decisión pública, el aprendizaje social, el fomento a la democracia, el mejoramiento de la gestión, entre otros¹¹.

En América Latina, la idea de la evaluación como una práctica gubernamental o social aún no está posicionada. Durante el período de reformas sociales de carácter populista (en los años 50 y 60) su aplicación ha sido mínima. Hacia los años 70's se incrementa su uso de manera puntual, orientado básicamente hacia la evaluación de algunos programas de desarrollo y promovida por organismos multilaterales y ONG's internacionales que desarrollaban programas asistenciales. En los años 90's ocurre un debilitamiento del debate sobre las políticas de desarrollo (Sánchez, 2000), al ejecutar procesos de ajuste en la economía y los gobiernos. Durante este período se intentó desarrollar sistemas nacionales de planificación y evaluación, algunos con mayor éxito y continuidad que otros en diversos países como Chile, Colombia (SINERGIA en el gobierno de Ernesto Samper), Bolivia (SISPLAN), México (modelo de la Secretaría de Desarrollo Social - SEDESOL), con diversas orientaciones, centrados en los programas sociales, algunos en el desempeño de las agencias estatales y otros en el presupuesto del gobierno.

La teoría y práctica evaluativa posee una particularidad interesante por cuanto en ella confluyen diversas disciplinas: la ciencia política, la ciencia de las políticas públicas, la investigación social, la sociología, la administración, entre otras, hecho que la convierte en un amplio campo de exploración temática. Sin embargo, la bibliografía aborda de manera independiente el tratamiento de la filosofía política, de la gestión pública, de las políticas públicas y de la evaluación social. En el caso de las políticas públicas, generalmente la evaluación se presenta como un paso más dentro del ciclo de políticas, con una fuerte orientación teórica (Bovens, Hart & Kuipers, 2005). Por el contrario, la literatura sobre evaluación presenta un fuerte carácter instrumental, desligado del contexto político y filosófico. La gestión pública por su parte se ocupa de la evaluación desde una perspectiva netamente administrativa o de desempeño. A pesar de lo anterior, es evidente la fuerte relación teórica que existe entre ellas y cómo la teoría política, así como lo hace con las políticas públicas, ofrece implícitamente en algunos casos y explícitamente en otros, la fundamentación principal para el desarrollo de la teoría evaluativa¹².

¹¹ En efecto, la evolución de la teoría y práctica evaluativa a partir de la post guerra ha ido de la mano con el avance del Estado de bienestar liberal, en Europa y los Estados Unidos, como herramienta para el diseño de las reformas sociales. En este período, los principales demandantes de procesos evaluativos fueron los gobiernos, siendo la sociedad el objeto de estudio a reformar. Una vez que se produce el giro mundial hacia la reducción del tamaño del Estado, la orientación hacia el mercado y el auge de las escuelas gerencialistas de la administración pública, la evaluación deja de concentrarse en la reforma de la sociedad, y se convierte en un mecanismo de reforma del Estado, buscando aumentar su eficacia.

¹² Si bien existen autores como E. House y F. Fisher que trabajan en esta dirección, su intención es más la de demostrar el carácter deliberativo de la evaluación.

2.2.1. Definición y características

Uno de los conceptos claves de esta investigación es el de “evaluación”. Se trata de un término que suscita variados tipos de asociaciones, en su mayoría con connotaciones más bien negativas, como por ejemplo, examen, fracaso, premio, corte de financiamiento, comparación; sin embargo, su connotación es más amplia.

El concepto de evaluación, particularmente en el campo de las intervenciones públicas y sociales, ha estado marcado por el debate sobre su alcance y utilidad (Tilley, 2000). Una primera tendencia define a la evaluación de manera instrumental, como un proceso sistemático que recoge información, la procesa y determina los resultados de una intervención con el objeto de tomar decisiones en el ámbito político o gubernamental, generalmente con un alcance determinado por los límites que impone la intervención evaluada. Una segunda tendencia plantea a la evaluación como algo inherente al proceso de construcción social, que sirve para conocer la realidad, percibir los problemas sociales, construir colectivamente propuestas y mejorar la sociedad. Esta última tendencia le da a la evaluación un alcance social más amplio y deliberativo que traspasa la simple intervención e incursiona en el cuestionamiento de la sociedad, el Estado y la ideología política que lo respalda.

La bibliografía reciente es unánime en afirmar que la evaluación debe ser parte integral del diseño y desarrollo de cualquier programa o intervención que pretenda aliviar o solucionar problemas asociados a la calidad de vida de las personas. La OECD u OCDE (1998) indica que *“es una valoración tan sistemática y objetiva como sea posible de un proyecto, programa o política que se está desarrollando o se ha completado, su diseño, implementación y resultados. La evaluación debería brindar información que es creíble y útil, posibilitando la incorporación de las lecciones aprendidas en el proceso de toma de decisiones”*. Por su parte, Rossi, Freeman & Lipsey (1999) entienden que la evaluación es *“el uso de procedimientos científicos para sistemáticamente investigar la efectividad de los programas... es un instrumento para estudiar, comprender y ayudar a mejorar los programas en todos sus aspectos importantes, incluyendo el diagnóstico de los problemas a los que se dirigen, su conceptualización y diseño, su implementación y administración, sus resultados y su eficiencia”*.

Weiss (1998) define la evaluación como el análisis sistemático de la operación y/o de los efectos de una política o programa, comparándolos con un grupo de estándares implícitos o explícitos, como medio para contribuir a mejorar el programa o la política. Encontramos aquí una visión sistémica que compara una intervención con parámetros determinados para concluir sobre su conveniencia y producir mejoras. De igual forma, Stufflebeam et al. (1971) definen a la evaluación como *“...el proceso de identificar, obtener y proporcionar información útil para juzgar alternativas de decisión”*. (Pág. 40).

Una visión similar ofrece Patton (1982) cuando afirma que la práctica evaluativa incluye la *“recolección sistemática de información sobre las actividades, características y efectos de programas, personal y productos de uso de determinadas personas para reducir incertidumbre, efectividad y tomar*

decisiones con respecto a lo que estos programas, personas o productos están logrando y afectando” (p.15). En el mismo sentido, Rossi & Freeman (1985), Stufflebeam et al (1987) y Robson (1993), la conciben como la sistemática aplicación de procedimientos de investigación social en evaluar la conceptualización y el diseño, la implementación y la utilidad de los programas de intervención social.

A partir de estas definiciones resulta claro que la evaluación es una herramienta fundamental para iluminar la toma de decisiones en las diferentes fases de ejecución de los programas e intervenciones, desde su concepción y diseño hasta su posterior valoración en términos de logros e impactos. Evaluar implica establecer un juicio sobre el valor o mérito de algo, una conclusión valorativa (Stufflebeam & Shinkfield, 1987). Para ello se requiere siempre de la comparación con algún estándar o punto de referencia, que ha sido establecido explícita o implícitamente. Determinar, por ejemplo, que un cierto programa de inmunizaciones tiene un 82% de cobertura no es una evaluación, sino solo una medición. La evaluación se realiza cuando se compara este resultado contra el nivel planeado, digamos 90% y se establece un juicio valorativo con base en esta comparación. El establecimiento de significados sustantivos para la interpretación a partir de la comparación entre resultados obtenidos y planeados es, en ocasiones, un proceso de cierta complejidad que requiere del concurso experto (Mohr, 1995). Ocurre con cierta frecuencia que ni los mismos ejecutores de programas han determinado con precisión cuáles son esos niveles esperados de desempeño del programa, es entonces tarea de los evaluadores determinar estos valores de manera más exacta.

Para garantizar la validez de sus interpretaciones la evaluación debe regirse por criterios de rigurosidad científica. De esta forma los evaluadores utilizan teorías, métodos y técnicas de las ciencias sociales. En evaluación, al igual que en una investigación científica, se recolecta, analiza e interpreta evidencia empírica para fundamentar las conclusiones. Entre más sólida sea la evidencia y la inferencia causal, más credibilidad tendrá el juicio evaluativo.

En términos concretos, en la bibliografía se encuentran múltiples definiciones sobre la evaluación de programas sociales (EvPro). A continuación, en la siguiente tabla, se mencionan algunas:

Tabla N° 2.5. Definiciones sobre evaluación de programas sociales (Fuente: Elaboración propia).

Autores	Año	Definición
Briones	1985	“...acto de juzgar o apreciar la importancia de un determinado objeto, situación o proceso, en relación con ciertas funciones que deberían cumplirse o con ciertos criterios o valores, explicitados o no... Se trata de un tipo de investigación que analiza la estructura, el funcionamiento y los resultados de un programa con el fin de proporcionar información de la cual se deriven criterios útiles para la toma de decisiones en relación con su administración y desarrollo” (p.4).
Riecken	1972	La evaluación de programas es la medición de las consecuencias deseables o indeseables de una acción psicosocial ejecutada para facilitar la obtención de metas perseguidas por el sujeto.
Perloff, Perloff y Sussana	1976	La determinación de la medida en que un programa logró uno o más de sus objetivos, las razones por las cuales no los alcanzó y la relación existente entre los efectos del programa y una amplia diversidad de variables y características del programa.

Autores	Año	Definición
Spaniol	1975	Es un proceso sistemático y continuo que proporciona información acerca del efecto o impacto de un programa, con el objeto de contribuir a adoptar una decisión política de trascendencia.
Ato, Quiñones, Romero y Rabadán	1989	La evaluación de programas es un proceso que consiste en valorar si un programa consiguió o no alcanzar los objetivos para cuya consecución fue ideado.
Fernández-Ballesteros	1996	Consiste en la investigación sistemática a través de métodos científicos de los efectos, resultados y objetivos de un programa, con el fin de tomar decisiones sobre él.
Freeman y Rossi	1993	La aplicación sistemática de procedimientos de investigación social para evaluar la conceptualización, diseño, implementación y utilidad de los programas de intervención social.
Rossi, Freeman y Lipsey	1999	Es el uso de procedimientos de investigación social para investigar sistemáticamente la efectividad de programas de intervención social. Requiere: a) recolección de datos y b) existencia de criterios de valoración.
Aguilar y Ander-Eggs	1993	“...una forma de investigación social aplicada, sistemática, planificada y dirigida; encaminada a identificar, obtener y proporcionar de manera válida y fiable, datos e información suficiente y relevante, en que apoyar un juicio acerca del mérito y el valor de los diferentes componentes de un programa (tanto en la fase de diagnóstico, programación o ejecución), o de un conjunto de actividades específicas que se realizan, han realizado o realizarán, con el propósito de producir efectos y resultados concretos; comprobando la extensión y el grado en que dichos logros se han dado, de forma tal, que sirva de base o guía para una toma de decisiones racional e inteligente entre cursos de acción, o para solucionar problemas y promover el conocimiento y la comprensión de los factores asociados al éxito o al fracaso de sus resultados” (Pág. 18).
Kröger et al.	1998	“...la evaluación de una intervención, un proyecto o un programa supone la recogida, análisis e interpretación sistemáticos de la información relativa a su funcionamiento y a sus posibles efectos. Los datos recogidos suelen utilizarse para decidir cómo mejorar la intervención y si debe ampliarse o abandonarse. En esencia, una evaluación debe responder a las siguientes preguntas básicas: ¿cuál es la naturaleza y el alcance del problema?; ¿qué intervenciones pueden incidir sobre el problema?; ¿a qué grupo objetivo debe dirigirse la intervención?; ¿afecta realmente la intervención al grupo objetivo?; ¿se aplica la intervención según lo previsto?; ¿es eficaz la intervención?” (Pág. 1).
Markiewicz, A.	2005	“...involucra a una gama de beneficiarios quienes representan una variedad de intereses políticos y organizacionales. Es importante reconocer el valor de la diversidad como reflejo de las diferentes perspectivas y posiciones sociales. Sin embargo, esto es un desafío para el evaluador en la gestión de las diferencias que se generan entre los beneficiarios y el evaluador, y entre los mismos beneficiarios en relación al enfoque y resultados de la evaluación” (Pág. 13). (Traducción propia)
Senate Standing Comittee on Social Welfare	1979a	“...el proceso de revisión minuciosa y crítica de la eficiencia, eficacia y adecuación de un programa o grupo de programas” (p.5). (Traducción propia).
Lipsey, M	1996	“Este es el proceso de determinar la estructura del programa y lógica por la cual podría ser razonable esperar la obtención de los resultados buscados” (Pág. 298) (Traducción propia).
Abdala, E.	2004	“...constituye un proceso sistemático, metódico y neutral que hace posible en conocimiento de los efectos de un programa, relacionándolos con las metas propuestas y los recursos movilizados.” (Pág. 26)
Jacinto, C. & Gallart, A. M.	1998	“...consiste en el examen riguroso y sistemático, a partir de criterios específicos, de los logros obtenidos por una política o programa social en relación a las metas que se han propuesto y a los efectos esperados en los grupos sociales beneficiarios” (Pág. 17).
Stufflebeam, D.	1971	Es un proceso que facilita la identificación, la recolección y la interpretación de informaciones útiles a los encargados de tomar decisiones y a los responsables de la ejecución y gestión de los programas.

Autores	Año	Definición
Repetto, E.	1987	“...un modo de investigación en orientación educativa que se propone indicar el valor de los programas orientadores de acuerdo con determinados criterios” (Pág. 250).
DAC	2002	(evaluación) “Apreciación sistemática y objetiva de un proyecto, programa o política en curso o concluido, de su diseño, su puesta en práctica y sus resultados. El objetivo es determinar la pertinencia y el logro de objetivos, así como la eficiencia, la eficacia, el impacto y la sostenibilidad para el desarrollo” (Pág. 21).
GAO	1998	“...son los estudios sistemáticos individuales conducidos periódicamente o sobre bases <i>ad hoc</i> para medir cuan bien un programa está trabajando. Estos son generalmente conducidos por expertos externos al programa, ni dentro o fuera de la agencias, también como por los gerentes del programa” (Pág. 3) (Traducción propia).
De Miguel	2000	Es el conjunto de principios, estrategias y procesos que fundamentan la evaluación de toda acción o conjunto de acciones desarrolladas de forma sistemática en un determinado contexto con el fin de tomar decisiones pertinentes que contribuyan a mejorar las estrategias de intervención social.
Levine, H.M.	1975	El examen de los efectos, resultados y output de un programa.
García, J.L.	1996	“Proceso de identificación, recolección y tratamiento de datos para obtener una información que justifique una determinada decisión. Debe servir no sólo para analizar un determinado programa, sino también para ayudarnos a comprender el proceso de aprendizaje” (Pág. 46).
Stufflebeam, D.L.	1989	Es el proceso de identificar, obtener y proporcionar información útil y descriptiva acerca del valor y el mérito de las metas, la planificación, la realización y el impacto de un objeto determinado, con el fin de servir de guía para la toma de decisiones, solucionar los problemas de responsabilidad y promover la comprensión de los fenómenos implicados.
Alkin, M C.	1990.	Un proceso de determinación de áreas de decisión sobre asuntos de interés mediante la selección de la información adecuada y analizada para realizar un informe con datos útiles para los que toman decisiones.
Pérez Juste, R.	1992	Proceso sistemático diseñado intencional y técnicamente, de recogida de información rigurosa –valiosa, válida y fiable- orientado a valorar la calidad de un programa, como base para la posterior toma de decisiones de mejora tanto del programa como del personal implicado y, de modo indirecto, del cuerpo social en el que se encuentra inmerso.
Northwerst	2002	“La evaluación de programas es una herramienta que ayuda al personal a evaluar objetivamente los efectos del programa a través de una recolección sistemática, análisis y reporte de información del programa” (Traducción propia).

Existen otros conceptos de importantes autores (Tyler, 1942; Beeby, 1977; Ruthman, 1977; Kaufman y English, 1979; Scriven, 1980; Joint Comité en el «*Standards for evaluations of educational programs*» 1981; Cronbach, 1983; Alvira, 1985; Stufflebeam et al, 1987; Caride, 1989; Pérez Juste, 1991; De la Orden, 1993; Aguilar y Ander Egg, 1992; Fernández Ballesteros, 1996; Martínez Mediano, 1996; De Miguel, 2000), pero todos ellos hacen mención a algunas características comunes:

- Se trata de una investigación y, por tanto, debe estar sometida a rigor científico. Los métodos, como guía del proceso, pueden modificarse en el tiempo. Y por ello, es obvio aceptar su flexibilidad estructural.
- El concepto lleva implícito la emisión de un juicio, término que reúne en sí determinados rasgos proactivos, perfectivos o de mejora, objetividad, etc. cuyo valor reside en orientar la acción.

- El significado final del proceso tiene cualidades comprensivas que deben explicar en gran medida la orientación de ese juicio. Por consiguiente, hablamos de modalidades de investigación cuyos resultados de estudio no tienen sentido fuera de su situación.

Así, la evaluación de programas es un proceso permanente y continuo de indagación y valoración de la planificación, la ejecución y la finalización del programa social. Su finalidad es generar información, conocimiento y aprendizaje dirigidos a alimentar una toma de decisiones oportuna y pertinente que garantice la eficiencia, eficacia y calidad de los procesos, los resultados y los impactos de los programas, todo ello en función del mejoramiento de las condiciones de vida de sus poblaciones beneficiarias. La siguiente tabla resume la definición sintetizada:

Tabla N° 2.6. Definición sucinta de la evaluación de programas sociales (Fuente: Elaboración propia).

<i>Evaluación de programas sociales, es un:</i>	
Proceso	No es una etapa ni una fase
Permanente	Durante toda la vida del programa
De valoración	Investigación evolutiva y científica
De indagación	Construcción de juicios valorativos y objetivos
De la gestión del programa	Planificación y ejecución del programa
Para generar conocimientos y aprendizajes	No es control ni fiscalización
Que alimente la toma de decisiones	Mejor gestión y mejor resultado

La evaluación es un proceso permanente que se inicia desde la identificación del problema y acompaña toda la vida del programa hasta finalizar su ejecución. La evaluación no es una acción de control y fiscalización, sino un proceso que permite a los distintos actores involucrados aprender y adquirir experiencias de lo planificado y actuado para tomar decisiones que optimicen la gestión del programa y garanticen mejores resultados e impactos.

Debe quedar claro que hoy en día ya se ha desterrado la “concepción tradicional” del concepto de evaluación, por cuanto el paradigma vigente es el de la “Gestión integral orientada a resultados” (Barzelay, 2001; Caiden, 1998; Cunill & Ospina, 2003; CLAD, 1999; Moctezuma & Roemer, 1999; Olías de Lima, 2001; Osborne & Gaebler, 1994; Ramio, 2001). En la siguiente tabla se compara las características de cada modelo de evaluación.

Tabla N° 2.7. Características comparativas de la evaluación tradicional y la evaluación en el marco de la Gestión Integral orientada a Resultados (Fuente: Elaboración propia).

	Evaluación tradicional	Gestión integral orientada a Resultados
¿Cómo participa la evaluación en los programas sociales?	No es parte constitutiva del programa.	Se consolida progresivamente una “cultura de evaluación” de los programas sociales.
¿Cuál es la finalidad predominante?	Control y fiscalización	Conocimiento y aprendizaje para la toma de decisiones y para el mejoramiento de la gestión y los resultados. Otorgar transparencia a la política social.
¿Cuál es el significado de la evaluación?	Es una fase o etapa del ciclo de los programas.	Es un proceso permanente de gestión Integral de Programas Sociales.
¿Cuándo se realiza la evaluación?	Se realiza fundamentalmente al inicio (evaluación ex antes) y al final del programa social	En los procesos que acompañan en los programas social en todo su vida. Se evalúa la planificación, la ejecución y la

	Evaluación tradicional (evaluación ex post)	Gestión integral orientada a Resultados finalización del programa social (evaluación integral)
¿Cuales son los aspectos prioritarios que se evalúan?	Se centra en la evaluación físico - financiera de los productos de los programas sociales.	Se centra en resultados e impactos, pero para ello, se requiere la evaluación del proceso
¿Qué se evalúa en la evaluación?	El seguimiento se limita una evaluación de los productos.	La evaluación de la ejecución de los programas implica el análisis del desarrollo de procesos, el cumplimiento de las actividades y el logro de los resultados durante el desarrollo del programa.
¿Se considera la evaluación en la planificación?	La planificación y la formulación tradicional no prevén el desarrollo de parámetros e indicadores de evaluación.	Los parámetros e indicadores de evaluación y monitoreo se construyen a lo largo del proceso de planificación del programa social.
¿Cuál es el papel de las evaluaciones externas?	Se prioriza solo la evaluación externa, sin considerar procesos de evaluación que desarrolla la misma institución ejecutora del programa social.	El diseño y puesta en marcha de una evaluación continua por parte del equipo de gestión de programa contribuye y viabilidad la realización de evaluaciones externas.
¿Cuál es la concepción metodológica de las evaluaciones?	Se centra en criterios metodológicos derivados de la evaluación económica - financiera de proyectos.	Se utilizan distintas herramientas de evaluación que tienen características específicas y complementarias. Se articulan métodos y técnicas cuantitativas y cualitativas de acuerdo con los aspectos que se evalúan.

2.2.2. Desarrollo histórico y modelos vigentes

La evaluación de programas sociales ha tenido un enorme desarrollo en los últimos cuarenta años (Mark, 2001). Se pueden distinguir tres etapas:

Primera etapa:

En la década de los años cincuenta-sesenta se inicia una gran revolución iniciada por los trabajos de Cronbach (1963), Scriven (1967) y Stufflebeam (1966). Es una época en la que la asignación presupuestaria en servicios sociales (sobre todo en Estados Unidos) propicia el desarrollo de estrategias evaluadoras para dar cuenta de su efectividad¹³.

Los programas de acción social que se establecieron fueron muy diversos: programas contra la pobreza, programas de salud, programas compensatorios en zonas marginales, programas de intervención educativa a largo plazo¹⁴, etc. Como

¹³ La evaluación de programas nació en Estados Unidos con el *Planning-Programming-Budgeting-System* (PPBS), el cual fue importado a los países europeos en la década de los 70'. Estas instituciones realizaban evaluaciones ex ante y evaluación ex post. Tuvieron un objetivo interministerial e introdujeron racionalidad científica en el proceso gubernamental (Toulemonde, 2000).

¹⁴ En la década de los cincuenta y comienzos del sesenta se buscó determinar los efectos de la nueva currícula en las llamadas nuevas ciencias y matemáticas; posteriormente a insistencia del senador Robert Kennedy se instituyó la Elementary and Secondary Education Act of 1965 que obligó realizar evaluaciones de los programas educacionales para estudiantes en desventaja económica; este mandato luego se expandió rápidamente a todos los programas sociales federales (House, 1993).

en todos ellos se trataba de estudios de campo, los científicos sociales tomaron la iniciativa en diseñar y llevar a cabo evaluaciones de esos programas a la luz de sus métodos tradicionales de investigación (Anderson & Ball, 1983; McLaughlin, 1976; Kellogg Foundation, 1998).

Sin embargo, estos modelos pronto vieron su insuficiencia y se hizo imprescindible una labor de consulta interdisciplinar (Cook & Shadish, 1986). Por eso, se diseñaron programas de intervención longitudinales que exigieron la participación de especialistas en evaluación desde los campos de la educación, psicología, sociología, ciencias políticas, economía y antropología, con lo que la evaluación tomó un carácter ecléctico, sin seguir un modelo específico o concreto (Talmage, 1982). Estas disciplinas se enriquecieron unas a otras, aportando o reconsiderando distintos métodos con los que afrontar los diseños de evaluación.

Segunda etapa:

Una década después, Lee Cronbach y colaboradores (1980) identificaron tres aspectos centrales en la evaluación de programas sociales: a) Comprensión por parte de los analistas políticos que la investigación de campo mejoraría la planificación del programa; b) reconocimiento de la política y la ciencia como parte integral en la evaluación de programas; c) Promoción de la metodología experimental con el fin de evitar las críticas de las evaluaciones de intervención longitudinales de la etapa anterior. Los dos primeros aspectos reconocidos por Cronbach (1980) son, en estos momentos, aspectos ampliamente aceptados, en líneas generales, por los evaluadores de distintas disciplinas. No así el tercero. El lugar de la metodología experimental entre las técnicas de evaluación está todavía sujeto a disputa entre los evaluadores, sobre todo en los países en vías de desarrollo¹⁵. En efecto, se puede afirmar que la evaluación de los programas sociales en América Latina está todavía en su infancia, tanto en su desarrollo profesional como en la literatura generada.

Tercera etapa: Auge de la metodología cuasi-experimental:

Campbell y colaboradores desarrollan la metodología cuasi-experimental y sus principales diseños los aplican en diversos programas sociales en países desarrollados. Desde allí, han aparecidos escuelas y teorías sobre la evaluación de programas basado en el modelo original de Campbell.

En esta tercera etapa, sin embargo, en vez de una convergencia evaluativa, se ha generado una divergencia de modelos de evaluación (Cronbach, 1983; Guba, 1978; Guba & Lincoln, 1985; House, 1980, 1993; Popham, 1975; Ross y Cronbach, 1976; Rossi & Freeman, 1985; Sandefur, Freeman & Rossi, 1986; Stufflebeam & Webster, 1980; Stufflebeam & Shinkfield, 1987). A partir de allí, Talmage (1982) divide a los investigadores de la evaluación de programas en cuatro grupos según la línea metodológica adoptada: 1) experimentalistas, 2) eclécticos, 3) descriptivos y 4) analistas costo/beneficio.

- En el primer grupo se ubican los investigadores interesados en diseños de evaluación robustos y de base experimental, centrados en establecer

¹⁵ Según Perloff, Perloff y Sussna (1976), el término *program evaluation* no apareció en el Psychological Abstract hasta recién en enero de 1973.

relaciones causales entre el programa y los resultados. Entre los máximos representantes están Cook & Campbell (1979), Riecken & Boruca (1974) y Rivlin & Timpane (1975).

- El segundo grupo adopta una posición ecléctica, inspirándose en diseños experimentales que hacen afirmaciones causales posibles, junto con métodos que describen el proceso de implementación del programa así como de variables contextuales que producen el programa. De esta forma el diseño de evaluación buscará una causalidad múltiple o generará explicaciones plausibles que se aproximan a la realidad. Sus máximos representantes son Bryk (1978), Cronbach et al (1980) y Weiss & Rein (1972).
- El tercer grupo rechaza los diseños experimentales, por considerarlos un método inapropiado para comprender los programas sociales, así como para determinar sus efectos. Para autores como Parlett & Hamilton (1989), Patton (1980) y Stake (1975a), House (1980), Colás & Rebollo (1993), García Ramos (1991), los datos significativos son sólo posible a través de descripciones en profundidad del programa en el contexto en que se desarrollan y a través del testimonio personal de los que participan en él.
- El cuarto grupo, representado por Haller (1974), Levin (1975) y Thompson (1980), trae a la evaluación de programas un enfoque metodológico prestado de la economía: el análisis costo-beneficio.

En la tabla siguiente se una comparación dimensional de cada modelo mencionado:

Tabla N° 2.8. Modelos metodológicos en la evaluación de Programas (Basado en Talmage, 1982).

Dimensiones	Experimentalista	Eclécticos	Descriptivos	Análisis Costo/Beneficio
Base filosófica	Positivista	Pragmático	Fenomenológico	Lógico-Analítico
Base disciplinaria	Psicología	Psicología, sociología, ciencia política	Sociología, antropología	Economía
Centro de la metodología	Determinar relaciones causales	Aumentar la búsqueda de relaciones causales con datos de proceso y contextuales	Describir el programa globalmente y desde la perspectiva de los participantes	Juzgar el valor del programa en términos de costo y beneficios
Metodología	Diseños experimentales y cuasiexperimentales	Diseños cuasiexperimentales, estudios de casos, descripciones	Etnografía, estudios de casos, observación participante	Análisis costo-beneficio
VARIABLES	Predeterminadas como input-output	Predeterminadas más las que surjan	Las que surjan en la evaluación	Predeterminadas
Grupo de control	Sí, es necesario	Sí es posible	No es necesario	Sí
Rol de los participantes en la evaluación	Ninguno	De ninguno a interactivo	Varía (puede estar en función de las observaciones de campo).	Ninguno
Rol del evaluador	Independiente del programa	Cooperativo	Interactivo	Independiente del programa.

Dimensiones	Experimentalista	Eclécticos	Descriptivos	Análisis Costo/Beneficio
Presiones políticas (internas-externas).	Controladas en el diseño o ignoradas	Integradas	A describir	A ignorar
Objeto del informe de evaluación	Tomar/no tomar una decisión	Interpretar y recomendar el mejoramiento del programa	Presentar una descripción global del programa en curso	Emitir un juicio

Es importante señalar que aunque a simple vista parece que las distinciones entre los modelos mencionados son fuertes, en la práctica hay un considerable solapamiento. En efecto, las diferencias se desvanecen a la hora de llevar a cabo evaluaciones concretas. Además, Mines, Gressard & Daniels (1982) realizaron una revisión de la literatura sobre diversos modelos de evaluación y llegaron a la conclusión de que había poca información disponible acerca de los méritos relativos, limitaciones y utilidad de dichos modelos. Naturalmente, esto no quiere decir que la efectividad de todas las técnicas y modelos de evaluación no pueda ser demostrada, sino que el proceso de determinar si una técnica o estrategia particular sirve para una finalidad dada es probable que lleve implícito algún tipo de juicio y error. No obstante, cualesquiera que sea el modelo de evaluación, siempre que se use más de uno de forma combinada, los beneficios que se obtengan al evaluar los programas serán mucho mayores que los riesgos (Bishop & Trembley, 1987).

2.2.3. Importancia social de la evaluación de programas

Evaluar constituye una forma particular de hacer investigación, no sólo para conocer y explicar una realidad social, sino también aportar al proceso de toma de decisiones y la optimización de la actividad evaluada. Diversas corrientes teóricas asocian a la evaluación de programas con el interés público, el mejoramiento de la sociedad y su contribución a la democracia (Scriven, 1967; Weiss, 1982). Se trata, por tanto, de un proceso que implica:

- a) Identificar los efectos provocados por la acción o el conjunto de acciones por evaluar y los costos en que se haya incurrido.
- b) Comparar la medida lograda con otra que sirva de parámetro para la evaluación, fundamentalmente, en el caso de los programas sociales, para constatar si hubo mejoramiento de las condiciones de bienestar de la población.
- c) Explicar los resultados comparativos encontrados, en función del diseño y las condiciones de aplicación del programa.
- d) Emitir un juicio de valor que califique, en conjunto, las actividades realizadas, los servicios brindados, sus efectos y su repercusión global.
- e) Sugerir las modificaciones necesarias para enfrentar los problemas detectados y aprovechar las fortalezas del programa, de manera que se contribuya a un mayor logro de sus objetivos.

En términos generales, se espera que la evaluación contribuya a un empleo eficiente de los recursos y a una alta efectividad de su impacto, al ofrecer información para mejorar los procesos de toma de decisiones. En particular,

Ospina (2001) señala que son pocos los casos en América Latina (cita como ejemplos a Chile, Colombia y Costa Rica) en los que se puede hablar de una evaluación sistemática en la administración pública, en sus tres o cuatro niveles: políticas públicas, programas públicos, organizaciones públicas y desempeño de los empleados públicos. Considera que la evaluación de la gestión pública está dejando de ser un ejercicio técnico para transformarse en una herramienta política útil para resolver problemas de gobernabilidad y legitimidad, al contribuir al fortalecimiento de la democracia y la rendición de cuentas a la sociedad. Así, una mejor gestión del gasto público debería contribuir a mejorar la imagen del Estado y aumentar su legitimidad, liberar recursos que pueden destinarse a satisfacer necesidades de grupos sociales de menores ingresos y aumentar la equidad. Sin embargo, aún no se reportan casos en América Latina en que estos resultados deseables se presenten empíricamente (Ospina, 2001).

Cohen (2001) resalta la importancia de proceder en la evaluación de programas sociales en América Latina considerando los siguientes aspectos técnicos: a) La magnitud del gasto social, b) el porcentaje del gasto social efectivamente redistributivo, c) la eficiencia en la aplicación del gasto anterior, y d) su impacto en la solución del problema. Las razones de esta secuencia son obvias: no pueden haber resultados relevantes en programas que carecen de recursos suficientes; no basta con gastar, pues muchos recursos pueden insumirse en los procesos sin llegar a la población objetivo¹⁶; y, suponiendo que éstos lleguen, por lo general se desconoce su efectividad para resolver o aliviar el problema que dio pie al programa, el grado en que justifican su costo y quiénes son sus verdaderos beneficiarios y perjudicados.

En otro aspecto, la evaluación de programas de desarrollo social es también importante porque facilita el aprendizaje de los ciudadanos interesados en el programa para afianzar los procesos de democratización (Blanet et al, 2000). En efecto, Monnier (1991) resalta la importancia de la participación de los actores sociales participantes y propone realizar evaluaciones con una actitud pluralista fundamentada en el reconocimiento de la diversidad de sistemas de valores que coexisten en el seno de la sociedad y que se construyen a partir de los conflictos sociales emergentes; lo que lleva al evaluador a actuar como conciliador entre las partes en desacuerdo, en busca de una solución satisfactoria, no óptima, mediante un proceso continuo de aprendizaje colectivo¹⁷.

¹⁶ El autor refiere un trabajo que demostraría que en Chile sólo 13% del gasto social total llega al 30% más pobre de la población; situación que, según considera, se repite en otros contextos nacionales de América Latina. Otro trabajo realizado en seis países latinoamericanos registra que 90% del gasto público social corresponde a gasto corriente.

¹⁷ Hengel & Carlson (2002) plantean la existencia de tres diferentes perspectivas para impulsar el aprendizaje desde las evaluaciones, con diferentes niveles de la complejidad social: A) El enfoque tradicional de tomadores de decisiones y organismos financieros, interesados exclusivamente en un aprendizaje que permita la retroalimentación sobre el propio desempeño de políticas y programas. B) El nivel que impulsa el aprendizaje organizacional de tomadores de decisiones, donantes, pero también de todo otro actor social relevante, cuya participación se torne fundamental. La evaluación y el aprendizaje incluyen acciones dentro de políticas existentes; el sistema mismo de acción gubernamental y los principios, normas y valores de la conducta colectiva. C) El impulso al aprendizaje de la sociedad, cuyo conocimiento, en términos del Banco Mundial, se entiende como un bien público global, que requiere empezar por hacer un uso público de las evaluaciones,

En el caso latinoamericano, Sulbrandt (citado en Kliksberg, 1997) considera que la mayoría de las políticas y los programas públicos de América Latina no se evalúan y esto se debe a que la evaluación tradicional no ha satisfecho las expectativas en lo que respecta a los programas sociales, de características muy diferentes a los dirigidos a la inversión económica. Salvo excepciones, la evaluación no ha conducido a provocar ningún cambio en el manejo de los programas ni en el aprendizaje de los ciudadanos.

Otra posición crítica de la evaluación en América Latina la plantea Solarte (2002), quien también considera que la evaluación es imprescindible en la región para aumentar la transparencia y responsabilidad de las acciones de gobierno y buscar políticas exitosas que promuevan la eliminación de la pobreza y la inequidad. Considera necesaria la evaluación, con énfasis en

“...la generación de aprendizaje social e institucional que permita aprender de las experiencias —negativas y positivas— que analice la conveniencia y resultados de las acciones políticas, sociales y económicas y que facilite el establecimiento de acuerdos sociales y políticos sobre lo que se debe emprender” (P.1).

Solarte (2002) estima que la evaluación es útil al gobierno para un mejor diseño de políticas y también a los ciudadanos para controlar al gobierno. La evaluación se vuelve herramienta de formación individual y de participación social, de aprendizaje social, de rendición de cuentas y control parlamentario, de mejoramiento de la gestión pública, de la política del gobierno (imagen pública, legitimidad, gobernabilidad), de validación del sistema social e ideológico (modelo de Estado y sociedad, democracia, interés público, bien común, ideologías dominantes, etcétera).

La participación en los procesos evaluativos tiene ya un fuerte arraigo en Estados Unidos y en los países europeos, especialmente los nórdicos, en forma de paneles, talleres, diálogos, investigación participativa, investigación-acción, comunidades epistemológicas, modelos multiactor, descentralización de decisiones, dinámicas de empoderamiento, redes, o incluso sociodramas, dibujos o metáforas para facilitar la comunicación, que hacen hincapié en los procesos de evaluación con aprendizaje (Isaac, 1999; DeBono, 1999). En el contexto japonés se ha desarrollado a partir del concepto *ba*¹⁸. También ha sido recientemente impulsada por organismos multilaterales y agencias del desarrollo como el Banco Interamericano de Desarrollo (BID, 2000), la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2001) y el Banco Mundial (2001). Este último ha publicado en uno de sus reportes el trabajo denominado *“Voices of the poor”*,

entenderlas, para luego estar en condiciones de criticarlas y, posteriormente, de realizarlas. Su desarrollo implica un mayor compromiso con los actores sociales. Obviamente, esta posibilidad no está exenta de riesgos, como algunos casos en que la participación ha provocado resultados no deseados, como una mayor inequidad; lo que lleva a la existencia de autores que se oponen terminantemente a ella, como Cooke & Kothary (2001).

¹⁸ El concepto *ba* describe un lugar de aprendizaje, de carácter físico, virtual o mental, con el objeto de relacionar individuos u organizaciones, para facilitar el aprendizaje y desarrollo creativo, mediante procesos de socialización, externalización, combinación e internalización. Se trata de una forma de colaboración que, en Occidente, podría considerarse formación de “capital humano” y “capital social”.

que recupera la participación, en discusiones en pequeños grupos y entrevistas grupales e individuales, de 60,000 personas pobres en 60 países.

En síntesis, la evaluación de programa es importante porque tiene una fuerte relación con los procesos democráticos de relación Estado-ciudadano, por cuanto se configura en instrumentos de control, de mejora y de participación en las políticas públicas y en la optimización del Estado.

2.2.3.1. El rol de la Evaluación de Programas en la Política Social

La evaluación de programas se mueve en un campo sociopolítico muy ambicioso que le da sentido y funcionalidad en diversos ámbitos, tales como el proceso de gobierno, la conformación de ciudadanía y democracia, la validación de ideologías y el mejoramiento del Estado, entre otros. A este respecto, Wholey (1992) plantea que la evaluación constituye “*el centro del debate político en una sociedad, revelador de los procesos de acción, de las iniciativas y de las estructuras de poder*” (P.201). En efecto, la evaluación de programas cumple diversos roles socio-políticos. A continuación se mencionan algunos de ellos:

La evaluación como herramienta de formación ciudadana y participación social.

Esta perspectiva recoge la idea de que la evaluación permite al ciudadano informarse sobre el desempeño del gobierno con respecto a sus intereses particulares, mejorando su capacidad de opinión, participación y decisión con respecto a su futuro (Frechtling, Stevens, Lawrenz & Sharp, 1993).

La evaluación como herramienta de aprendizaje social.

Se considera que la sociedad puede aprender a partir de un análisis formal de las intervenciones sociales y sus resultados, mejorando su capacidad para generar consenso con respecto a lo que es conveniente o no para sus intereses colectivos. Así, la evaluación como práctica social, es el instrumento para facilitar este proceso adquiriendo un rol de mediación entre los diversos actores. Adicionalmente, la evaluación contribuye a que una sociedad se reconozca a sí misma a partir del conocimiento y debate de los intereses y opiniones de los diferentes grupos que la conforman (Lane, 2000; Furukawa & Hoshino, 2001), facilitando su *empowerment* (Fetterman, 2000) y su capacidad para construir su futuro (Sangra, 2000).

La evaluación como herramienta de responsabilización.

La tendencia actual de los gobiernos es hacia la reducción del gasto público y la búsqueda de mayor eficiencia en su gestión. El ciudadano, a su vez, ha pasado a ser considerado como un cliente que “adquiere” bienes y servicios del Estado. De lo dicho, profundas reformas al Estado se han emprendido en la mayoría de los países y se han promovido procesos descentralizadores en búsqueda de mayor autonomía y equilibrio fiscal, así como la aparición de nuevas formas de prestación de servicios. Este escenario ha generado mayores demandas de

“responsabilización”¹⁹ y transparencia al interior del gobierno, desde los subordinados hacia sus superiores y desde el gobernante hacia la ciudadanía, con el objeto de presionar el uso apropiado de los recursos y la consecución de resultados en la gestión.

El concepto de “responsabilización” ha evolucionado desde el control de gestión clásico -a través de herramientas de fiscalización tradicionales- hasta el control por parte de la ciudadanía²⁰. Para este fin, la evaluación es una herramienta que apoya efectivamente diversas formas de responsabilización, tales como el control parlamentario, la lógica de resultados en la administración pública y el control social. En el primer caso (el control parlamentario) la evaluación puede ser encomendada o desarrollada por el Parlamento para obtener información objetiva que facilite su labor de control del ejecutivo. En el segundo caso, la introducción de lógicas de control y evaluación de resultados en la administración y la realización de evaluaciones independientes de las entidades ejecutoras proveen elementos para determinar el éxito o fracaso de una gestión en función de la consecución de cambios objetivos y valorizables en las situaciones sociales que se desean afectar. En cuanto al control social como mecanismo de responsabilización, se ha mencionado anteriormente que la evaluación puede servir como instrumento para mejorar el conocimiento de la comunidad sobre la gestión de sus gobernantes, permitiéndole ejercer de manera responsable e informada su derecho a exigir resultados a los gobernantes elegidos democráticamente.

La evaluación como herramienta de mejoramiento de la gestión pública.

Aunque la teoría administrativa y la teoría política incluyen a la evaluación como parte de la gestión de las organizaciones y del proceso de las políticas públicas, la evaluación ha adquirido gran relevancia particularmente con la denominada “nueva administración pública”, en la década de 1990. La “nueva administración pública” presenta una clara orientación hacia la gestión por resultados (Waissbluth, 2002; Ospina, 2001), y la calidad de los servicios. Este enfoque presenta a la evaluación como una herramienta de gestión de gobierno, orientada a determinar el funcionamiento del aparato institucional, sus programas de acción, la consecución de resultados, medir los niveles de eficiencia y la generación de impacto. Este compromiso requiere diversos niveles evaluativos: 1) el primer nivel macro es la evaluación del desempeño del gobierno entendido como el compromiso democrático con la ciudadanía a partir de determinadas políticas públicas y sus logros; 2) el nivel dos tiene que ver con el desempeño de los programas que cruzan varias organizaciones; c) el nivel tres corresponde al

¹⁹ El término responsabilización intenta traducir el concepto inglés de accountability, aunque – a diferencia del original incluye el control social por la ciudadanía. CLAD – BID – EUDEBA. (2000). La Responsabilización en la nueva gestión pública latinoamericana. Buenos Aires: Ed. Universitaria de Buenos Aires.

²⁰ La “responsabilización” incluye dar cuenta sobre el cumplimiento de mandatos constitucionales y legales del gobernante, la cantidad y calidad de los bienes o servicios que el gobierno provee, el cumplimiento de los objetivos y planes programados, el uso de los fondos públicos, el cumplimiento de los objetivos de largo plazo, los niveles de eficiencia, el cumplimiento de procedimientos, la utilización de recursos etc., y debe ser ejercida en los distintos niveles del gobierno.

desempeño de cada organización que ejecuta algún aspecto de las políticas públicas; y, finalmente, d) el nivel del desempeño de los empleados en una organización (Ospina, 2001).

Cada nivel evaluativo alimenta procesos de mejoramiento institucional o de política general (en el diseño y formulación de políticas). Adicionalmente, la evaluación de aspectos constitutivos de la estructura del Estado, así como el marco institucional, los mecanismos de regulación, las políticas de descentralización, la política fiscal, entre otros, es promovida como una herramienta no solo para mejorar el gobierno, sino para orientar los procesos de reforma del Estado²¹ con el objeto de mejorar el bienestar de los ciudadanos.

La evaluación como herramienta política de gobierno.

Con relación al gobierno y su proceso, la evaluación es una especie de lente mediante el cual el gobernante percibe los asuntos claves de la problemática social y de la opinión ciudadana, con el objeto de incluirlos en su programa o agenda de gobierno. Le permite valorar constantemente su imagen pública a partir de la evaluación de los resultados de las acciones emprendidas por su gobierno y la investigación acerca de la percepción ciudadana sobre su labor. Asimismo, la evaluación le puede ofrecer una visión más comprensiva de la gobernabilidad con que cuenta su administración y de las demandas e intereses de sus competidores políticos, o de los demás poderes del Estado.

La práctica evaluativa desde diversos ámbitos tales como el gobierno, el congreso, las organizaciones civiles, los partidos políticos etc., aún con objetivos y usos distintos, es una actividad que influye poderosamente en la conformación y funcionamiento de las fuerzas políticas en virtud del impacto potencial que tiene la difusión y aplicación de sus hallazgos en la imagen pública de los gobernantes, los partidos políticos y funcionarios en general. Evaluaciones positivas se convierten en armas políticas relevantes para los partidos y el gobierno en la medida que sus resultados sean conocidos por la opinión pública, contribuyendo a generar confianza (Chelimsky & Shasish, 1997). Asimismo, el Parlamento puede utilizar estrategias evaluativas para controlar las actuaciones del gobierno, ampliar su conocimiento sobre los problemas sociales y la percepción ciudadana, desarrollar nueva legislación, aprobar presupuestos (Valadez & Bamberger, 1994), generar debates públicos y reflejar ante la ciudadanía su preocupación por los logros sociales.

Las evaluaciones debatidas en público o en foros académicos pueden tener marcada incidencia en la validación de las acciones gubernamentales, contribuyendo a la conformación de nuevas iniciativas partidistas de rechazo o apoyo a la gestión pública. El evaluador, por lo tanto, es un “referente fáctico” en el debate político (Meny & Thoening, 1992) y su legitimidad dependerá de las características de los clientes de la evaluación.

²¹ Esta última visión es impulsada fuertemente por la corriente neoinstitucional, aunque autores como Campbell trabajaron desde los años 60's el concepto de la evaluación como elemento para la reforma social.

La evaluación como herramienta de validación del sistema social e ideológico.

Una perspectiva mucho más formativa, desde el punto de vista de la ciencia política y la democracia, es la evaluación como herramienta de validación del sistema social y político, que contribuye a la construcción de conocimiento sobre su funcionamiento y la correlación con las teorías políticas y sociales que lo soportan bajo un modelo de Estado o de sociedad. Efectivamente, la investigación evaluativa provee evidencia empírica que puede facilitar la construcción de visiones más realistas en un sistema democrático. Este punto de vista de conocimiento y de investigación, generalmente relegada a los académicos, actúa en diversos niveles y se construye a partir de la conjunción de cada uno de los ejercicios evaluativos que se realizan (Fisher, 1997).

La evaluación de la política concierne aquí con las concepciones sobre el alcance del gobierno (nivel de intervención), las estrategias que se utilizan para ejercer las funciones, las condiciones políticas vigentes, la problemática social que se privilegia y su conexión con postulados políticos que el gobierno representa, la forma cómo se realizan las distribuciones en términos de equidad o de enfoque hacia determinada población, la relación entre gobierno y mercado y el rol de cada uno en el mejoramiento social, las características políticas y su relación con las políticas públicas desarrolladas, la conformación de la estructura social, elites y grupos de opinión y la forma en que la acción pública incide en su conformación o modificación, el análisis de la estructura institucional del Estado y la influencia de las políticas en su ajuste o cambio²².

2.2.4. La evaluación de programas sociales en el Perú

En nuestro país, durante las últimas décadas del siglo XX, más que evaluaciones de las políticas y programas sociales implementados y seguimientos de avance de metas, se han encontrado mecanismos de control jurídico y financiero del gasto (Ej. actividades de Contraloría). Peor aún, cuando se han registrado recortes presupuestales, las primeras actividades recortadas han sido las de evaluación, siempre consideradas “menos útiles”.

Esta situación se ha visto reforzada por la actitud de las organizaciones sociales nacionales quienes han considerado a la evaluación una forma de “desviar” sus recursos financieros y humanos en actividades no sustantivas. A esto se suma el rechazo que tradicionalmente han generado las evaluaciones provenientes de las fundaciones y organismos que aportan su financiamiento, lo que provoca que se las conciba más como un “control externo indeseado” que como un proceso propio que puede ayudar al aprendizaje de los participantes y de la comunidad en su conjunto. Todas estas razones han contribuido para que, hasta fechas muy recientes, se preocupen muy poco en capacitarse en temas de

²² Frank Fisher (1997) plantea la necesidad de cuatro componentes evaluativos para las políticas públicas, desde una óptica de argumentación política: evaluación de logros de la política, validación de la pertinencia de la política con respecto a la situación problema, el análisis de la contribución de la política a la construcción de valor en la sociedad, y la evaluación ideológica. Esta última corresponde al análisis del sistema de valores e ideologías que soportan la política.

evaluación y, en todo caso, prefirieran las evaluaciones cualitativas, que si bien permiten profundizar en casos concretos, no ofrecen una visión agregada de conjunto, ni miden el impacto con precisión.

En el ámbito universitario la situación tampoco es óptima, pues se observa una enorme carencia de material didáctico propio y de cursos especializados en la materia. Debido a que la mayoría de libros y revistas que exponen métodos de evaluación presentan experiencias de aplicación rigurosos provienen de Estados Unidos y de Europa o Canadá, resulta imprescindible y urgente desarrollar materiales bibliográficos adaptados que contribuyan a un proceso de aprendizaje que, a su vez, incremente la eficacia de las evaluaciones de programas, la transparencia en el uso de los recursos y la participación ciudadana en los procesos de elaboración de políticas públicas.

Noche (1982) identificaba, en Francia, tres tipos fundamentales de problemas ligados al atraso relativo en torno de una cultura de la evaluación en su país y que pueden considerarse válidos para la realidad peruana: los sociopolíticos, los administrativos y los metodológicos:

- Los sociopolíticos se consideran más importantes y, al mismo tiempo, más difíciles de cambiar porque están enraizados en las culturas y tradiciones de los pueblos y sus gobiernos. La tradición presidencialista y dictatorial ha propiciado en el Perú una suerte de resignación ciudadana frente a las inapelables decisiones del poder ejecutivo.
- Los obstáculos administrativos incluyen el diseño de planes, programas y presupuestos con objetivos muy vagos y ausencia de metas cuantificadas que faciliten su evaluación; una organización de trabajo que no define con claridad qué entidad administrativa se encargará de realizar la evaluación integral de los programas públicos; la falta de recursos destinada a esta actividad poco legitimada por su discutida aportación al bienestar social; el rechazo a engorrosos controles que distraen de la actividad sustantiva; etc. Todas estas razones están presentes en el contexto peruano, donde la actividad controladora se ha enfocado más a la realización de auditorías financieras para evitar malos manejos de fondos y al cumplimiento de la legalidad en los procedimientos de aplicación de recursos, que a la medición de la eficacia o la repercusión de las acciones realizadas.
- Los problemas metodológicos tampoco pueden soslayarse. La mayoría de técnicas empleadas en evaluación (análisis costo-eficacia, costo-beneficio, métodos multi-criterios, encuestas, paneles, estudios de caso, historias de vida, etc.) enfrentan serios cuestionamientos que disminuyen su credibilidad. Si bien ninguna técnica es perfecta, una buena combinación de las mismas, con aportación de las más diversas disciplinas como la sociología, la política, las matemáticas, la informática, etc., puede disminuir considerablemente los problemas no resueltos, y contribuir a desarrollar un proceso de aprendizaje a partir de la experiencia realizada.

En el ámbito gubernamental agregaría, a los problemas enunciados por Nioche (1982), otro de naturaleza más psicológica, vinculado al conjunto de

temores que la evaluación desata. Anteriormente, cuando se aplicaba la evaluación ésta respondía a un concepto restringido y punitivo (vinculado a la aplicación de sanciones, despidos, etc.). Así, toda evaluación se movería en un campo minado de intereses en el que, para los evaluados, se pone en juego el mantenimiento de su empleo, su nivel de ingreso, reconocimientos, prestigio, status, autoestima, etc. Por eso, muchos candidatos a ser evaluados se oponen a implantar sistemas de evaluación como un mecanismo de defensa frente al riesgo, y lo aplican casi en forma automática, pero en caso de ponerse en práctica la evaluación, plantean mecanismos de evasión, de justificación o de cumplimiento meramente formal de la meta, sin comprometerse realmente con el impacto de su actuación. Entonces, resulta imprescindible avanzar hacia una concepción más positiva de la evaluación, entendida como un proceso de aprendizaje individual y social a partir de los aciertos y los errores, de la detección de necesidades de capacitación y desarrollo, de la premiación de esfuerzos, etc., que sólo por excepción tenga que aplicar medidas represivas.

En el caso particular de la administración pública nacional, lo anterior se combina con obstáculos de índole cultural y tradicional: históricamente, los equipos de trabajo se han conformado más por relaciones de camaradería y lealtad que por selecciones basadas en los conocimientos, habilidades y actitudes de los candidatos a integrarlos. Esto ha provocado que quien ocupaba el puesto de nivel superior tendiera más a proteger a su equipo que a evaluarlo técnica y objetivamente.

Por otro lado, se afirma que la calidad de los servicios que brinda el Estado en el campo social a las personas de menores ingresos es en promedio deficiente debido a que se asigna muy pocos recursos para el gasto social. Aunque esto es una gran verdad, lo cierto es que otros países logran mejores resultados con recursos igualmente magros para el sector o con problemas de pobreza (Bolivia, por ejemplo). Evidentemente, el Perú debiera aumentar en lo posible los recursos que se destinan al sector, pero el problema de gestión y mejora de la calidad resulta crítico para que tales recursos rindan el fruto debido. Y, para ello, se considera la necesidad del fortalecimiento de los sistemas de medición y difusión de resultados e impactos.

2.2.4.1. Aspectos diagnósticos de la situación actual

A pesar que desde la década de 1990 se hizo explícita en la gestión pública la importancia de los sistemas de monitoreo y evaluación de proyectos sociales, estos son aún limitados y en muchos casos deficientes (Ñopo & Robles, 2004). A modo de ejemplo, de un total de 20 proyectos, que implicaban un monto de US\$ 700 millones en el año 2000, apenas la mitad tenía cierto avance en este rubro. Increíblemente, sólo 3 de los veinte proyectos analizados que sólo implicaban US\$ 17 millones de los US\$ 700 millones, contaban con manuales para sus sistemas de evaluación y monitoreo (Apoyo, 2000)²³.

²³ Si bien 9 de los 20 proyectos de la muestra de proyectos analizados por Apoyo (2000) realizaron alguna evaluación intermedia o ex-post, ésta habría sido de cobertura limitada, o sólo de tipo

La escasa y deficiente evaluación de impacto de los programas sociales en el Perú plantea la necesidad urgente de establecer, a la brevedad posible, requisitos legales referidos a los sistemas de monitoreo y evaluación y obtener compromisos sostenibles para su puesta en vigencia efectiva. El Perú no puede seguir gastando dinero en programas sociales sin que la sociedad en su conjunto cuente con mecanismos para saber cuál es el efecto del dinero invertido. Ni el sistema político ni la sociedad muestran indicios de aprender sobre estos temas, identificando mejor qué funciona y qué no. La discusión permanece relegada a un grupo de técnicos, cuyo impacto en las decisiones de política es limitado.

El Perú ya tiene una proporción bastante considerable de su presupuesto asignado al área social y no tiene los mecanismos institucionales para saber si lo que gasta sirve de algo o no, qué parte sirve y cuál no, y por qué sirvió y por qué no (Mostajo, 2002). Las decisiones de política se toman con información incompleta y sin suficiente análisis, lo que favorece que los actores políticos tomen posiciones sobre la base de prejuicios e ideologías. Los programas sociales tienen, en el Perú como en cualquier otro país, un gran nivel de inercia y son difíciles de gerenciar. Lograr implantar acciones que sean de verdad en ayuda efectiva para mejorar la calidad de vida de los peruanos, de una manera sostenible, es difícil.

En términos presupuestarios, no hay evaluación de resultados ni de calidad. Tampoco hay políticas financieras. En general, los estudios revelan que las evaluaciones del presupuesto ex post son desfasadas, solamente de carácter procedimental que focalizan en auditorías e informes, y poco preventiva. Además, se financian actividades y proyectos de baja calidad; se miden y evalúan procesos, procedimientos y actividades en lugar de resultados y, existe una carencia de procesos de rendición de resultados y corresponsabilidad entre niveles y entre unidades públicas. Todo ello hace que exista inadecuada información estratégica para la toma de decisiones (Mostajo, 2002).

Por todo ello, los autores nacionales recomiendan que la evaluación independiente de los programas y proyectos sociales debiera ser obligatoria, así como la publicación de los resultados, alimentando el proceso de asignación

cuantitativo, o sólo de una parte del proyecto. Asimismo, en pocas oportunidades se habrían realizado evaluaciones de impacto y sostenibilidad, al punto que en ninguno de los programas existe una línea de base y sólo tres, que representan cerca del 19% de la muestra en términos de los desembolsos efectuados en el 2000, utilizan grupos de control o comparación. Por otro lado, en los cuatro estudios de caso analizados por Apoyo (2000), que incluyen el Programa de Capacitación Laboral Juvenil – PROJoven, el Proyecto de Salud y Nutrición Básica (PSNB), el Programa Nacional de Wawa Wasi (PNWW), y el Programa de Mejoramiento de la Calidad de la Educación Primaria (MECEP), se encuentra que el origen de los sistemas de monitoreo y evaluación está en los convenios firmados con los organismos internacionales. En segundo lugar, que no existe una marcada cultura de monitoreo y evaluación en las autoridades políticas. Se constata también que sólo en el caso de PROJoven hubo interés por llevar adelante una evaluación de resultados por parte de la alta dirección cuando inició sus operaciones; en el PSNB y el MECEP la alta dirección no manifestó decidido interés pero tampoco obstruyó; y en PNWW el interés de la alta dirección se dio recientemente. En tercer lugar, que los organismos financieros cumplen un doble rol al exigir evaluaciones, vinculadas al desembolso o la operación del proyecto, y al brindar apoyo técnico y flexibilidad para la contratación de consultores. Por último, que los proyectos de mayor envergadura, como el MECEP, tienen un menor desarrollo de sistemas de monitoreo y evaluación en comparación a los proyectos de menor envergadura, como PROJoven y el PSNB.

presupuestaria. Esa tarea tiene que ser impulsada de manera sostenible por el Gobierno Central. Los autores recomiendan establecer legalmente la obligatoriedad de estudios de impacto independientes, anual o bianualmente, como parte del proceso presupuestario. Los informes deben ser abiertos al público y a la prensa, y servir para la discusión en la Comisión de Presupuesto. Se requiere promover su difusión activamente, para fortalecer el proceso de toma de decisiones y mejorar el gasto social. Sólo así se empezará un camino más o menos sostenible para mejorar la eficacia y eficiencia del gasto e intervención social, sobre la base de criterios técnicos y pragmáticos (Ortiz de Zevallos, 2001, 2003).

2.2.4.2. Sistemas de evaluación en la Administración Pública

A partir de la segunda mitad de la década pasada, se inicia un movimiento en la gestión pública peruana que intentaba introducir y fortalecer los Sistemas de Monitoreo y Evaluación (M&E) de los programas públicos con énfasis en lo social (Ortiz de Zevallos, 2001). Ello se lleva a cabo desde un conjunto de iniciativas vinculadas a la ejecución del gasto y al proceso presupuestario, desarrolladas desde el Ministerio de Economía y Finanzas, así como con el desarrollo de esquemas de M&E en diversos proyectos sectoriales, principalmente aquellos que reciben financiamiento de agencias de cooperación.

Como parte de estos esfuerzos, entre 1995 y 2000, el Ministerio de Economía y Finanzas impulsó tres iniciativas de M&E en la administración pública: la Oficina de Inversiones (ODI), el Sistema de Información Administrativa y Financiera (SIAF) y el Planeamiento Estratégico Sectorial Multianual (PESEM).

La Oficina de Inversiones (ODI)

La ODI funciona como una unidad rectora de las inversiones públicas: analiza las propuestas de inversiones presentadas por los diferentes sectores y procura que estas se formulen adecuadamente. A fines de 2000, promulgó la Nueva Ley de Inversiones estipulando que a partir de 2001 todos los proyectos de inversión deben cumplir con normas que definen cada una de las etapas del ciclo de proyectos, las cuales deben ser aprobadas para asegurar su viabilidad. La etapa de preinversión requiere de la elaboración del perfil del proyecto; la etapa de inversión consiste en la elaboración del expediente técnico y la ejecución del proyecto; y la etapa de post-inversión se refiere a la evaluación ex post de los proyectos.

Desde el punto de vista del M&E de proyectos en general, la principal limitación del nuevo marco legal es que solo abarca las intervenciones públicas de carácter eventual. De otro lado, la Ley establece que las unidades ejecutoras de los proyectos son las encargadas de hacer las evaluaciones ex post, sin embargo, lo más probable es que estas se encuentren desactivadas cuando culminen los proyectos. Resulta necesario que este esquema de M&E permita evaluar la “sostenibilidad” de los programas, luego de varios años de haber finalizado.

Por otro lado, tanto el SIAF como el PESEM no son propiamente esquemas de M&E. Veamos a continuación al SIAF.

El Sistema de Información Administrativa y Financiera (SIAF)

El SIAF es un sistema de ejecución, más no de formulación de los gastos de las unidades del gobierno central. Fue creado para que funcione como una herramienta de gestión del Tesoro, con el fin de que este pueda supervisar los gastos de las unidades ejecutoras. Como sistema de M&E, el SIAF posee limitaciones porque no permite analizar las causas por las cuales se avanzó o no con lo programado, por lo cual la información procesada por el sistema no retroalimenta a los organismos ejecutores.

Los Planeamientos Estratégicos Sectorial Multianual (PESEM)

El PESEM, por su parte, son un esfuerzo por planificar estratégicamente las operaciones que desarrollarán los sectores del gobierno central en el futuro. Cada uno de ellos establece claramente su visión, misión y objetivos. Luego, en caso de ser aprobados, pasan a integrar el Plan Estratégico Multianual (PEM), el cual se somete a la aprobación del Consejo de Ministros para la formulación del presupuesto del año siguiente. Desde el punto de vista del M&E, los PESEM no funcionan como un sistema de planificación estratégico, sino más bien como un sistema de programación. Además, en muchos casos, no existe consistencia entre los objetivos planteados y los indicadores definidos para el cumplimiento de las metas.

El SNIP

Actualmente está vigente el Sistema Nacional de Inversión Pública (SNIP). Durante muchos años fue común que las Entidades Públicas pasen directamente de la idea de un Proyecto a la elaboración del Expediente Técnico y de ahí a la ejecución de la obra, obteniendo como resultado (en muchos casos) proyectos que no resolvían problemas y que no contaban con recursos para su Operación y Mantenimiento. De esa manera se usaban ineficientemente los escasos recursos públicos destinados a inversión.

Ante esta situación, y sabiendo que los recursos disponibles para la inversión pública son limitados, se creó el Sistema Nacional de Inversión Pública (SNIP) por Ley N°27293, que establece que las entidades públicas encargadas de ejecutar proyectos de inversión pública deben aplicar una serie de principios, procesos, metodologías y normas técnicas que permitan optimizar el uso de los recursos públicos.

El SNIP es uno de los sistemas administrativos del Estado, que como el de Presupuesto, Tesorería, Contaduría, Control, Contrataciones y Adquisiciones, etc., es de observancia y cumplimiento obligatorio para todos los niveles de gobierno.

El SNIP busca optimizar el uso de los recursos públicos destinados a la inversión, con el fin de que su uso tenga un mayor impacto en el desarrollo económico y social del país. La principal herramienta que usa el SNIP es el análisis de pre-inversión, el cual permite una evaluación técnica, económica, financiera y de sostenibilidad de los proyectos previa a su ejecución.

En el SNIP, el proyecto de inversión²⁴ (PIP) atraviesa por el ciclo clásico de vida del proyecto. Este ciclo de proyecto tiene tres etapas: pre-inversión, inversión y post-inversión.

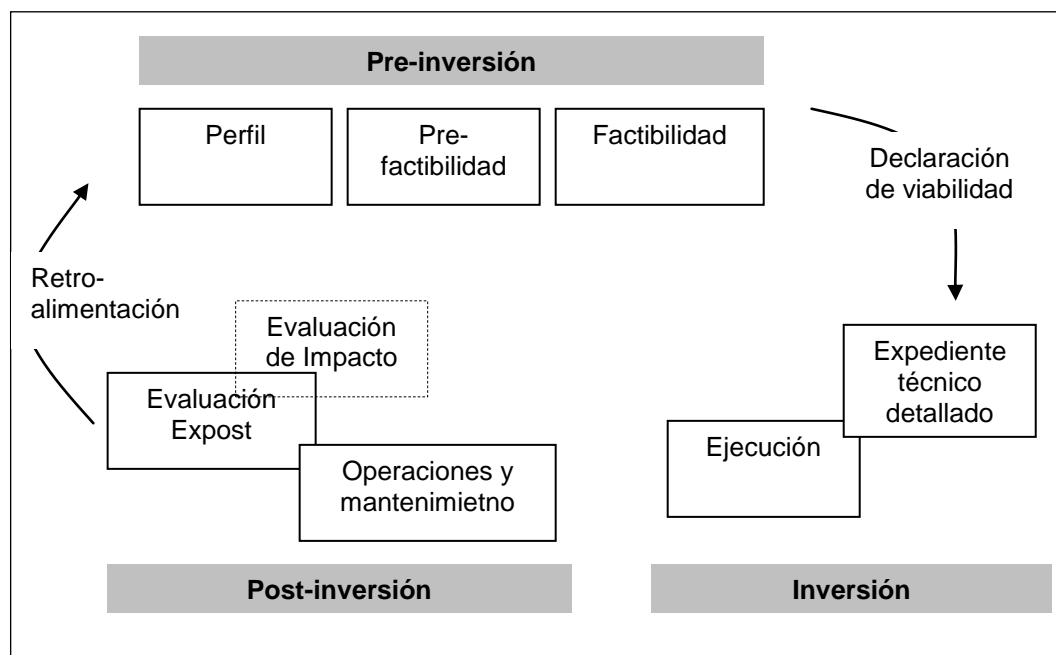


Figura N° 2.2. Ciclo del proyecto de inversión según el SNIP (Fuente: Elaboración propia).

Tal como se observa en la Figura 2.2, la evaluación de impacto está ubicada en la tercera etapa, en la etapa de post-inversión.

En la primera etapa, los estudios de Pre-inversión permiten reducir progresivamente la incertidumbre propia del riesgo a invertir. La elaboración del **Perfil** es importante y obligatoria para todos los proyectos, pues sirve para identificar el problema que ocasiona la necesidad de elaborar un PIP. Los estudios de **pre-factibilidad** son el segundo nivel de análisis de la fase de preinversión y tiene como objetivo acotar las alternativas identificadas en el nivel de perfil, sobre la base de un mayor detalle de la información. Incluye la selección de tecnologías, localización, tamaño y momento de inversión, que permitan una mejor definición del proyecto y de sus componentes. En esta etapa, la mejor calidad de la información permitirá descartar las alternativas menos eficientes. Finalmente, los estudios de **factibilidad** tienen por objetivo establecer definitivamente los aspectos técnicos fundamentales: la localización, el tamaño, la tecnología, el calendario de ejecución, puesta en marcha y lanzamiento, organización, gestión y análisis financieros, considerando un menor rango de variación en los costos y beneficios de la alternativa seleccionada en el estudio de prefactibilidad.

Los estudios de Pre-factibilidad y Factibilidad pueden no ser requeridos dependiendo de las dimensiones y características del PIP. Una vez concluidos los estudios de Pre-inversión, se solicita la Declaración de Viabilidad del PIP. La

²⁴ Según el SNIP, un proyecto de inversión pública tiene por finalidad crear, ampliar, mejorar, modernizar o recuperar la capacidad productora de bienes y servicios.

Declaración de Viabilidad no implica que inmediatamente se asigne al PIP una partida presupuestal. Declarado la viabilidad se inicia la etapa de la inversión, en donde se elabora el Expediente Técnico y luego se ejecuta.

Finalmente, en la fase de post-inversión se inserta la evaluación ex post. Esta evaluación es el estudio por el cual se busca determinar la eficiencia, eficacia e impacto de las actividades desarrolladas para alcanzar los objetivos del PIP.

La evaluación ex post es un proceso que analiza los efectos y los impactos de los proyectos en los beneficiarios. Es decir, mide el grado de cumplimiento de los objetivos y metas generales del proyecto. La evaluación ex post es el proceso que busca determinar los efectos y el impacto del proyecto (esperados e inesperados) con relación a las metas definidas a nivel de propósito y resultados, tomando en consideración los supuestos señalados en la matriz de marco lógico planteada en la evaluación ex ante. Constituye la última etapa del análisis del proyecto, por un lado analiza si las actividades desarrolladas realmente permitieron obtener los resultados, y si éstos realmente permitieron alcanzar el propósito. Del mismo modo busca determinar si el proyecto realmente contribuyó a resolver el o los problemas detectados (MEF, 2002).

Esta evaluación debe ser considerada y prevista desde el momento del diseño del proyecto e incorporada a la planificación operativa, definiendo su frecuencia, duración, responsables y recursos. Las fuentes sobre las que se basan un proceso de evaluación ex post son los informes de monitoreo, el seguimiento de los indicadores definidos para los niveles de propósito y resultado, e información externa. La evaluación ex post puede realizarse en dos momentos: inmediatamente después de finalizado el proyecto y/o un tiempo después de terminado el proyecto (MEF, 2002).

Desde ya, es importante mencionar que la evaluación ex post tiene dos etapas:

1. La primera de ellas, desde el punto de vista cuantitativo, trata de medir la rentabilidad real del proyecto con relación a su rentabilidad *ex ante*. Para esto se realiza una nueva evaluación costo – beneficio o **evaluación de resultados** en la que se reemplazan los valores del análisis ex ante por los resultados efectivos del proyecto (costos, beneficios, tasas de interés, etc.). Esta evaluación permite por una parte analizar los desvíos existentes entre el análisis *ex ante* y los efectivamente ocurridos, así como la rentabilidad real frente a la rentabilidad estimada. Esta etapa, desde el punto de vista cualitativo, está constituida por los análisis donde se identifican y consolidan las evaluaciones del proyecto, debería determinar tanto los problemas y virtudes encontrados en el análisis ex ante, las características de la implementación del proyecto incluyendo los desvíos ocurridos, los efectos positivos y negativos del proyecto y sobre todo aquellos inesperados.
2. La segunda etapa, la cual se recomienda llevarla a cabo tiempo después de concluido el proyecto, esta constituida por una **evaluación de impacto**, la cual se especializa en descubrir los cambios permanentes en población destinataria y en el ambiente. Igualmente busca determinar si se ha contribuido a mejorar la calidad de vida de la población objetivo en

los aspectos o dimensiones que el proyecto abordó. El fin principal de esta segunda etapa es determinar la efectividad de los resultados obtenidos para producir el impacto social deseado.

A pesar que existe una normatividad cada vez más sólida en el Perú, no debe olvidarse que los programas sociales nacionales son, intrínsecamente, muy políticos. En efecto, cada programa social crea beneficiarios que luego presionan para que dicho programa permanezca inalterado. En ese sentido, los malos diseños iniciales arrastran cola. En un país con pocos recursos como el Perú, con una gran proporción de pobres, darse el lujo de no tener debate político y presupuestal más informado sobre qué programas sociales deben o no crecer, cómo se deben mejorar, cuáles deben dejarse a un lado, etc., es un absurdo. Las decisiones para mejorar la efectividad de las políticas sociales son, casi siempre, políticamente difíciles. Solo ocurrirán (y se mantendrán) si hay suficiente respaldo político a lo largo del tiempo, lo que requiere que una mayor cantidad de actores esté consciente de las opciones disponibles y de las consecuencias previsibles de cada opción. Sin embargo, lo cierto es que los pobres del Perú no merecen que se tomen decisiones respecto de los programas sociales que le deben dar oportunidades de salida de su pobreza de manera poco informada, sin evaluaciones sistemáticas e independientes (Ortiz de Zevallos, 2003).

2.3. LA EVALUACIÓN DE IMPACTO

Cuando se habla de evaluación de programas es común referirse al tema con una concepción global y comprehensiva. En general, la palabra evaluación contiene aspectos como la evaluación *ex ante*, la supervisión, la evaluación de los procesos, la evaluación de costos-beneficios, la evaluación de viabilidad y la evaluación de impacto. A pesar del uso generalizado del término, lo cierto es que cada uno de estos aspectos son completamente diferentes. A modo de ejemplo:

- La supervisión ayuda a evaluar si un programa se está ejecutando de acuerdo a lo planificado. Un sistema de supervisión de programas permite una retroalimentación constante sobre el estado en que se encuentra la implementación del programa e identifica los problemas específicos a medida que surgen.
- La evaluación de los procesos se relaciona con la forma en que funciona el programa y se centra en los problemas de la entrega de servicios. En las evaluaciones de costos-beneficios o eficacia en función de los costos se estiman los costos de los programas (monetarios o no monetarios), en particular su relación con respecto a usos alternativos de los mismos recursos y a los beneficios que produce el programa.
- Por último, la evaluación del impacto tiene el objeto de determinar en forma más general si el programa produjo los efectos deseados en las personas, hogares e instituciones y si esos efectos son atribuibles a la intervención del programa (Blomquist, 2003). Las evaluaciones de impacto también permiten examinar consecuencias no previstas en los beneficiarios, ya sean positivas o negativas (CONPES, 2002).

De lo dicho, y en consonancia con el objetivo de la investigación, el análisis posterior se ceñirá a la evaluación de impacto, excluyendo la evaluación ex ante y la evaluación de procesos. A continuación se definirá y describirán sus características.

2.3.1. Definición y características

La evaluación de impacto mide los cambios en el bienestar de los individuos, cambios que pueden ser atribuidos a un programa o a una política específica. Los objetivos de la evaluación de impacto son proveer información y ayudar a mejorar la eficacia de los programas. En este sentido, es una herramienta que utilizan los responsables de la formulación de políticas y que posibilita que el público pueda exigir cuentas sobre los resultados de los programas (Baker, 2000; Prennushi, Rubio & Subbarao, 2000, Heckman, LaLonde y Smith, 1998).

Existen otros tipos de evaluación de programas, como las revisiones organizacionales y el monitoreo de procesos, pero éstos no miden la magnitud de los efectos ni atribuyen la causalidad que corresponde a tales efectos. La evaluación de impacto, en cambio, tiene un nexo ineludible con el análisis causal de la intervención de programas y sus efectos (directos, indirectos, positivos, negativos o neutros) en la mejora de la calidad de vida (Banco Mundial, 2000). En la siguiente tabla se presenta las definiciones más representativas de la evaluación de impacto.

Tabla N° 2.9. Definiciones sobre evaluación de impacto (Fuente: Elaboración propia)

Autores	Año	Definición
Mateu, P. & Vilca, J.	2004	Toda evaluación de impacto de un programa social tiene como objetivo determinar cuál es la mejora que perciben los individuos que participan de dicho programa. Esta mejora puede ser medida a través de variables de interés como pueden ser ingresos, empleo, reducción de la desnutrición, etc. Para ello, se requiere contar con un grupo de beneficiarios y controles que permitan comparar la situación de haber participado del programa con la de no haberlo hecho. (p. 53).
Abdala, E.	2004	Bajo la denominación de evaluación de impacto se entiende el proceso evaluatorio orientado a medir los resultados de las intervenciones, en cantidad, calidad y extensión según las reglas pre-establecidas. La evaluación de impacto abarca todos los efectos secundarios a la planeación y a la ejecución: específicos y globales; buscados (según los objetivos) o no; positivos, negativos o neutros; directos o indirectos (la puesta en marcha del programa puede generar por sí misma efectos sobre los directamente involucrados, hasta la sociedad toda). (Pág. 28-29).
Boothroyd, P.	1998	"...la medición de impacto se orienta a los efectos no esperados y los beneficios indirectos, o lo que los economistas denominan externalidades". (Traducción propia)
Apocada, P.	1999	"... la evaluación de los efectos producidos por un programa o intervención. Es decir, aquellos cambios habidos debidos a la citada intervención". (Pág. 363).
Anguera, T.	1989	"... la actividad encaminada a identificar, predecir, interpretar, comunicar y prevenir el impacto de un programa, proyecto, plan o acción y desemboca en la valoración de los efectos finales de una cadena de ellos que se inicia como una causa, que es precisamente la intervención o implementación de un programa". (Pág. 26).
De Miguel, M.	1997	"Este tipo de comparación se conoce como <i>effectiveness</i> -traducido unas veces por efectividad y otras por eficacia-, consideramos más oportuno la

Autores	Año	Definición
		utilización de “desarrollo” ya que este concepto define mejor los logros o resultados a largo plazo (<i>outcomes</i>), especialmente cuando estos resultados pretenden evaluar cambios en los valores sociales”. (Pág. 163).
Baker, J.	2000	“...la evaluación de impacto tiene el objeto de determinar en forma más general si el programa produjo los efectos deseados en las personas, hogares e instituciones y si estos son atribuibles a la intervención del programa. Las evaluaciones de impacto también permiten examinar consecuencias no previstas en los beneficiarios, ya sean positivas o negativas”. (Pág. 1).
GAO	1998	“La evaluación de impacto es una forma de evaluación de resultados que mide el efecto neto de un programa al comparar los resultados del programa con una estimación de qué habría pasado en caso de la ausencia del programa. Esta forma de evaluación es empleada cuando los factores externos son considerados como influyentes en los resultados del programa, de esta manera aislar la contribución del programa al logro de objetivos”. (Pág. 5) (Traducción propia)
CONPES	2002	“...identifica de manera sistemática los efectos (positivos o negativos, esperados o no) sobre los hogares e instituciones, generados por un programa de desarrollo social. Se distingue de las evaluaciones de procesos, gestión y resultados, al determinar la causalidad entre la intervención y sus efectos observados. Así, permite entender la magnitud de los cambios generados por dicha acción sobre el bienestar y en especial sobre la reducción de la pobreza. Concretamente, apuntan a retroalimentar y a mejorar la efectividad de los programas sociales, comparándolos con intervenciones alternativas.” (Pág. 4)
Mohr, L.	1995	“Es un medio para determinar la extensión hasta la cual un conjunto de actividades afecta el estado de algunos objetivos y examina por qué los efectos fueron mínimos o grandes” (Traducción propia).
Ezemenari, K.; Rudqvist, A.; Subbarao	1999	Una evaluación de impacto mide la extensión hasta la cual un programa ha causado cambios deseados en el público esperado: se refiere al impacto neto de una intervención sobre las familias e instituciones, atribuible únicamente a la intervención. Por tanto, la evaluación de impacto consiste en medir los resultados y el cambio de desarrollo de corto y largo plazo resultantes de una intervención. (Pág. 65) (Traducción propia).
Sandoval, J.M. & Richard, M.P	2003	La evaluación de impacto mide los cambios en el bienestar de los individuos que pueden ser atribuidos a un programa o a una política específica y sus objetivos son proveer información y ayudar a mejorar su eficacia. En este sentido, es una herramienta que utilizan los encargados de tomar decisiones en la formulación de políticas, y que hacen posible que el público pueda exigir cuentas sobre los resultados de los programas. (Pág. 7)
Regalia, F.	1999	“La evaluación de impacto es una herramienta indispensable para medir si un programa está logrando su objetivos, cuánto ha cambiado la situación de los beneficiarios como resultado del programa y cómo hubiera sido la situación si el programa no se hubiese desarrollado”. (Traducción propia).
Blomquist, J.	2003	La evaluación de impacto es la identificación sistemática de los efectos sobre los individuos, hogares e instituciones atribuidos a un programa o proyecto.
Verstraete, L.	1993	La Evaluación de Impacto tiene por objeto determinar si un proyecto ha producido los efectos deseados en las personas, hogares e instituciones y si estos efectos son atribuibles a la intervención. Las evaluaciones de impacto, por lo general, también permiten examinar consecuencias no previstas en los beneficiarios, ya sean positivas o negativas.
Castro, G. & Cháves, P.	1994	La evaluación de impacto de un proyecto social es un proceso de identificación, análisis y explicación de los cambios o modificaciones que, en función de un problema social, se hayan producido en las condiciones sociales de la población-objetivo y en su contexto, como consecuencia de la aplicación del proyecto que se evalúa.
Vela, R.	2003	Esta propuesta de evaluación, aunque mantiene la lógica de comparación entre situación inicial y situación final, se caracteriza por su intención en superar el enfoque unicausal de explicación de los cambios producidos y distinguir entre lo que el proyecto se propuso hacer y lo que realmente ocurrió.

En años recientes se ha desarrollado una gran literatura sobre evaluación de impacto de programas sociales (Heckman, Lalonde & Smith, 1999; ILPES & CEPAL, 2003, Baker, 2000) y, en general, es valorada como un proceso amplio y global, donde al abordaje cuantitativo se le agregan técnicas cualitativas (Abdala, 2001).

En concordancia con la tabla anterior, la evaluación de impacto puede entenderse como parte del proceso de evaluación donde se identifican, comprenden y explican cambios en variables y factores producidos por un programa. La evaluación de impacto analiza todos los efectos vinculados a la planeación, ejecución y operación de un proyecto, sean éstos específicos o globales. Los efectos que se analizan pueden estar vinculados a los objetivos del proyecto (explícitos o primarios) o aquellos que por el sólo hecho de la intervención se produzcan y no formen parte de los objetivos fijados en la planeación (implícitos o secundarios).

La identificación del efecto o impacto de un programa es un ejercicio bastante complejo, pues requiere responder la pregunta: ¿qué hubiera pasado si el proyecto no hubiera existido?²⁵. Es decir, el problema consiste en identificar los cambios en ciertos indicadores de la población objetivo, como resultado exclusivo de la intervención del programa. Lo que se busca en estas evaluaciones no es sólo identificar los cambios, sino que además es necesario saber si esos cambios pueden atribuirse a la intervención o no.

Según Baker (2000), para asegurar el rigor metodológico adecuado, la evaluación de impacto debe estimar el escenario contrafactual o simulado alternativo, es decir, lo que habría ocurrido si el programa nunca se hubiera realizado. En efecto, el escenario contrafactual nos dirá qué es lo que realmente cambia y en qué medida. Así, el impacto es la medida de cambio de una variable o factor en un horizonte de tiempo determinado, comparando un escenario con y sin intervención (Dar & Tzannatos, 1999).

Para identificar el estado contrafactual generalmente se usan grupos de control. La idea es simple: para cada individuo en el grupo de participantes, se trata de identificar un individuo similar en el grupo de control²⁶. Así, el efecto medio de la participación en el programa puede ser calculado como la diferencia media en los resultados obtenidos entre los participantes y controles²⁷ (Dehejia & Wahba, 1999).

²⁵ Una pregunta más compleja todavía es ¿qué hubiera sucedido si el proyecto se hubiera ejecutado de manera diferente?

²⁶ El término “grupo control”, propiamente utilizado en evaluaciones experimentales, es utilizado en este documento de manera indistinta con el término “grupo de comparación”.

²⁷ Existen dos supuestos que son cruciales para este ejercicio: a) el supuesto de Independencia Condicional (CIA). La idea es que si se controla por características observables, entonces el resultado que se obtendría sería el mismo en ambos grupos si el programa no se ejecutara. Esto permitiría atribuir cualquier diferencia entre participantes y controles, a la existencia del programa (Rosenbaum y Rubin, 1983). Este supuesto requiere disponer de una gran cantidad de información que explique la participación en el programa y también las variables resultado. Lamentablemente, este supuesto no puede ser testeado y por tanto, es necesario confiar en la teoría o estudios previos; y b) el supuesto de Participación Unitaria Estable (SUTVA). Implica asumir que el impacto de un programa en una persona no depende de otras personas o de cuantas personas están en el programa (Bryson, Dorsett & Purdon, 2002).

En años recientes se ha desarrollado un método alternativo para analizar la equivalencia de los grupos participantes y controles, basándose en probabilidades de participación (*propensity scores*) en vez de variables específicas. Este tipo de metodología ha sido utilizada intensivamente en los últimos años a nivel internacional y, por tanto, ha estado en el centro del debate metodológico²⁸. La idea general es que si hay un número grande de variables que influyen en el programa, se puede estimar la probabilidad de participación (condicionadas a estas variables) para participantes y controles (con una regresión binaria). De esta manera, se garantiza que el grupo de control tenga un perfil bastante similar al grupo de participantes, y se facilita el análisis, dado que el efecto condicionado a las variables será igual al efecto tratamiento condicionado en el *propensity score* (Rosenbaum & Rubin, 1983). Este procedimiento es desarrollado en extenso en el capítulo 4.

En otro aspecto, es importante tener una serie de criterios para evaluar la calidad de una investigación de impacto. Aunque existen algunos indicadores clave, aún no se ha desarrollado un sistema teórico-base que permita analizar su rigurosidad. Como intento primario, Fernández-Ballesteros (1996) ha propuesto un listado de cuestiones relevantes en evaluación de programas (LCREP) muy útil para evaluar “ex ante” la calidad del proceso de evaluación. En otras palabras, este instrumento trata de juzgar la evaluabilidad (factibilidad) del programa con base en: a) la calidad del proceso de planificación e implantación llevado a cabo y b) de las barreras que puede encontrar el evaluador al realizar el trabajo. Sin embargo, aún este instrumento carece de un marco teórico que le permita interpretar sus alcances

2.3.2. Pertinencia de la evaluación de impacto

La información generada por la evaluación de impacto ayuda a tomar decisiones sobre la necesidad de ampliar, modificar o eliminar cierta política o programa, y es posible utilizarla para asignarle prioridad a las acciones públicas (Fernández Ballesteros, 1996). Además, estas evaluaciones contribuyen a mejorar la eficacia de las políticas y programas (Baker, 2000) al abordar las siguientes preguntas: ¿Logra el programa las metas propuestas?, ¿Justifica el valor del programa su costo? ¿Son los cambios producidos resultados directo del programa, o son resultado de otros factores que ocurrieron simultáneamente? ¿Cambia el impacto del programa dependiendo del grupo al que se está tratando de beneficiar (hombres, mujeres, pueblos indígenas) o de la región o a través del tiempo? ¿Tuvo el programa efectos inesperados, ya sean positivos o negativos? ¿Qué tan eficiente es el programa en comparación con intervenciones alternativas?.

Hay cuatro preguntas que pueden ayudar a decidir cuándo realizar una evaluación de impacto, es decir cuándo es pertinente hacerlo:

²⁸ Desde las propuesta iniciales de Rosenbaum y Rubin (1983, 1984) se han desarrollado numerosas aplicaciones económicas por autores diversos Heckman, Ichimura y Todd (1997); Dehejia y Wahba (1999); Hotz, Imbens y Mortimer (1999); Lechner (1999); Heckman, LaLonde y Smith (1999); Sianesi (2001); y Bryson, Dorsett y Purdon (2002).

- 1) ¿Tiene el programa importancia estratégica en la reducción de la pobreza o mejora de la calidad de vida?²⁹. Se pueden evaluar las políticas y programas de los que se espera un mayor impacto en la calidad de vida, para asegurar que los esfuerzos en ese sentido siguen el camino correcto y permitir las correcciones necesarias.
- 2) ¿Contribuirá la evaluación de un determinado programa a llenar los vacíos en el conocimiento sobre lo que sirve y no sirve para mejorar la calidad de vida de la población? La decisión sobre qué evaluar también puede basarse en cuánto se sabe sobre la eficacia de intervenciones alternativas. Si hay lagunas en el conocimiento sobre qué es lo que mejor funciona para reducir la pobreza o aumentar la calidad de vida, entonces se justifica una evaluación de impacto.
- 3) Este programa, ¿pone a prueba enfoques innovadores para aumentar la calidad de vida? La evaluación debe fomentar el aprendizaje. Una evaluación de impacto puede ayudar a ensayar enfoques precursores y decidir si se deben ampliar y ejecutar a mayor escala. Por consiguiente, el carácter innovador de un programa o política puede ser una buena razón para evaluarlo³⁰.
- 4) La política o programa, ¿está destinado a grupos difíciles de alcanzar o se espera que su impacto dependa, por ejemplo, del género de los beneficiarios? Las políticas y programas destinados a grupos pobres difíciles de alcanzar se topan con una gran variedad de factores sociales, culturales, económicos y organizacionales, que pueden contribuir a su éxito o a su fracaso. Por estas razones, es de especial importancia que la evaluación esté bien diseñada y sea bien ejecutada.

2.3.3. ¿Por qué siendo tan importante la evaluación de impacto se hace tan poco?

Acorde con Medina Giopp (2002), existen diversas razones por las cuales se realizan muy pocas evaluaciones de impacto. A continuación mencionamos las más comunes:

- i) Muchos gobiernos e instituciones la consideran costosa, prolongada y de alta complejidad técnica.
- ii) Los resultados pueden ser “políticamente delicados”, controversiales, especialmente si son negativos, ya que asignan responsabilidad a los funcionarios políticos y

²⁹ El impacto final que persigue todo programa social es reducir la pobreza y mejorar la calidad de vida de la población. Sin embargo, por ser un fin último, los programas y proyectos sociales se diseñan y ejecutan para mejorar cualquier aspecto de la vida de las personas (educación, salud, infraestructura, etc.) o reducir situaciones precarias o riesgosas para su desarrollo. Por eso, cuando me refiero a aumentar la calidad de vida o reducir la pobreza, debe recordarse que son fines generales y últimos.

³⁰ Sin embargo, hay que hacer una advertencia importante: una evaluación provechosa requiere de un programa suficientemente maduro. Aunque un programa esté probando enfoques innovadores, necesita objetivos bien definidos y actividades bien delineadas, así como un marco institucional estable que se preste para la implementación.

administrativos de los programas. Es usual, entonces, que la información sobre el rendimiento se concentre en el volumen de recursos ejecutados y en cobertura, incurriendo en el mito de que gastar más y entregar más bienes o servicios es bueno *per se*. Lo cierto es que ningún programa o proyecto puede justificar su existencia por la mera entrega de bienes y/o servicios.

- iii) En ocasiones, las evaluaciones se critican por no contar con oportunidad, rigor académico o por no responder a ciertas preguntas.
- iv) Se consideran sinónimo de auditoría y control y por tanto innecesarias, lo cual es falso.
- v) Excepcionalmente las técnicas y métodos de evaluación se consideran como información pública, casi siempre han sido consideradas como informes “confidenciales” para funcionarios del más alto nivel.
- vi) Se consideran imposiciones externas, etc.

Todas estas razones han creado y fortalecido una cultura gerencial donde la evaluación de impacto es sinónimo de control y cumplimiento de condiciones o imposiciones de donantes, supervisores, o niveles jerárquicos superiores, más que una herramienta participativa y orientada al aprendizaje organizacional. Lo anterior ha derivado en mitos creados en torno de la evaluación que es necesario desterrar, pues produce, como señala Mokate (2000, 2001), la percepción de un “monstruo” y no la de un “aliado” para el desarrollo.

En efecto, la evaluación de impacto está plagada de una serie de mitos que son producto del desconocimiento de su real naturaleza. Con la intención de eliminarlos, se denuncia a continuación los principales y más nocivos.

Evaluar el impacto de los programas es demasiado costoso:

El costo promedio para realizar una evaluación de impacto –a nivel internacional- asciende a un promedio de 433 mil dólares. De este monto, el costo mayor corresponde a la recopilación de datos, que representa, en promedio, el 53.3% del total (Baker, 2000).

La inversión en evaluación, dadas estas cifras, parece enorme. De hecho, en varios países muchos funcionarios del gobierno y organizaciones de la sociedad civil afirman que con lo que se gasta en la evaluación de impacto podría lograrse una cobertura mayor de bienes y servicios a la población objetivo. Con cientos de miles de dólares o varios millones, podrían lograrse, argumentan, más beneficiarios. Sin embargo, lo que no se analiza es que sin evaluación de impacto nada puede garantizar que los bienes y servicios que se entregan producirán los beneficios esperados, es decir, la transformación deseada.

En realidad, de los proyectos que comprende un estudio del Banco Mundial (Baker, 2000), los costos de la evaluación de impacto representan, en promedio, solamente el 0.56% del costo total del proyecto, ni siquiera el 1%. Para el caso del programa Progreso de México, los más de tres millones de dólares invertidos en evaluación de impacto, representan solamente el 0.001% del costo total del

programa. A continuación se presenta una tabla resumen de los costos estimados en algunos proyectos del Banco Mundial.

Tabla N° 2.10. Resumen de costos estimados de diversas evaluaciones de impacto del Banco Mundial (Fuente: Baker, 2000)³¹

Proyectos	Costo estimado de la evaluación (US\$)	Costo % del costo total del proyecto	Desglose de los costos de evaluación (%)			
			Viaje	Personal del Banco Mundial	Asesores	Recopilación de datos
Administración escolar (Nicaragua)	495,000	1,26%	8,1%	18,1%	39,0%	34,8%
Administración escolar (Salvador)	443,000	0,60%	7,7%	7,4%	25,8%	59,2%
Programas de Vales (Colombia)	266,000	0,20%	9,4%	9,8%	21,8%	59,0%
Fondo Social (Honduras)	263,000	0,23%	3,0%	11,5%	53,2%	32,3%
Fondo Social (Nicaragua)	449,000	0,30%	4,9%	33,0%	7,8%	55,7%
Fondo Social (Bolivia)	878,000	0,50%	3,4%	14,6%	12,9%	69,1%
Trinidad y Tobago Capacitación Jóvenes	238,000	0,80%	7,6%	11,5%	17,9%	63,1%&
Promedio	433,000	0,56%	6,3%	15,1%	25,5%	53,3%

Pese a la imposibilidad de generalizar con tan pocos casos, la consulta realizada a diversos expertos de organismos internacionales como el Banco Interamericano de Desarrollo (BID) y diversas agencias de las Naciones Unidas (Ej. CEPAL, PNUD, FAO), revela que las evaluaciones de impacto en proyectos sociales raramente supera un 4% o 5% del costo total del programa.

Bajo esta lógica, resultará siempre, en todos los casos, mucho más costoso no saber si los proyectos son exitosos (o no) que invertir los recursos de la evaluación de impacto en la entrega de bienes y/o servicios, con la única expectativa de lograr más beneficiarios. Esto es así, ya que sí el proyecto posee deficiencias y fallas, el volumen de recursos perdidos derivado de éstas puede representar cientos (o miles) de veces lo que se gastaría en la evaluación. Es muy costosa y riesgosa la incertidumbre que puede existir sobre el rendimiento de un proyecto derivada de una decisión explícita de no evaluar (Valadez & Bamberger, 1994)³².

³¹ Este costo no incluye el costo del personal local de contrapartida no financiados con el préstamo o crédito. Las cifras se refieren al periodo dentro del cual se seleccionaron los proyectos de la muestra de las evaluaciones y no al financiamiento total proporcionado alguna vez por el Banco y otros a esas instituciones.

³² Por ejemplo, una evaluación de impacto de Progres (Programa de educación, salud y alimentación de México) reveló que uno de sus objetivos educativos “cerrar la brecha de género en la matrícula, especialmente en el nivel de secundaria” había producido un impacto nulo. El modelo estadístico que analizaba las variables del ausentismo y deserción asignaba una alta capacidad de explicación del comportamiento de éstas a la distancia de las escuelas de los hogares de estudiantes femeninos. En tal sentido, se asignaron becas más altas a jefes de hogares con mujeres en edad secundaria para estimular su asistencia y cubrir los costos derivados del transporte desde el hogar a la escuela. Así, la evaluación reflejó ausencia de impacto significativo en cerrar la brecha de género en la matrícula de nivel secundaria (Boltvinik, 2000). Este resultado, y otros derivados de

En el caso de los programas sociales del Perú, Yamada & Pérez (2005), encuentran que las evaluaciones de impacto pueden costar menos de 1% del monto total de proyectos grandes y entre 2% y 4% del monto total de proyectos pequeños o pilotos, lo que representa dinero bien invertido si se considera las grandes sumas de recursos públicos que podrían orientarse mejor luego de una sólida evaluación de impacto.

Más impacto se logra al gastar más:

Existe una tendencia en América Latina a destinar cada vez mayores recursos para combatir la pobreza, tanto en el área rural como urbana. La década de los años 90 refleja claramente esta tendencia, pues la proporción del gasto social ha aumentado significativamente en ella³³.

Es indudable que para enfrentar la pobreza se requiere contar con más recursos, sin embargo, esto es insuficiente ya que es posible gastar más pero lograr menos o nada. Así, es posible alcanzar una baja proporción de gasto social efectivo respecto del volumen del gasto social destinado. Los datos son contundentes. A pesar que el volumen de recursos destinados a lo social se ha incrementado, los niveles de pobreza, bienestar, desigualdad, exclusión y población en riesgo de padecer pobreza³⁴, así como la violencia e inseguridad no han logrado reducciones significativas en la región en los últimos años, más bien y en un buen número de países, algunas de estas condiciones se han agravado. Dado lo anterior, la evaluación de impacto, favorece los procesos de asignación de recursos, al identificar las áreas de mayor “rentabilidad social” de la inversión. Por otra parte, la evaluación, acompañada del monitoreo, favorece la eficiencia en la implementación de programas.

El carácter reservado de las evaluaciones:

Aún existen funcionarios que conciben a los resultados de las evaluaciones como “propiedad” de la alta dirección (Medina Giopp, 2002; Medina Giopp &

evaluaciones de impacto, no son anecdóticos o excepcionales, sino recurrentes. Es el caso de microempresarios que reciben capacitación en temas que no son relevantes, o bien que la escala de crédito a la que tienen acceso es insuficiente, o bien que los equipos recibidos por pequeños agricultores a través de proyectos de mecanización, por ejemplo tractores, resulta inviable ya que los repuestos o los expertos para su reparación son escasos en el mercado local, etc. Los casos más dramáticos son aquellos en los que la entrega de los bienes y/o servicios producen algún daño o empeoramiento de la condición de bienestar que se deseaba satisfacer, en vez de una mejora. Por tanto, más barato es evaluar que invertir ciegamente en la provisión de servicios y bienes sin saber su impacto.

³³ Pueden consultarse los informes anuales del Panorama Social de CEPAL o el Informe IPES de BID.

³⁴ El método ingreso-consumo o MIC, presenta una nueva forma de elaborar mediciones sobre la base de la línea de pobreza. Uno de sus principales aportes es el de identificar a través de analizar de manera simultánea el ingreso y el consumo de los hogares, a los individuos o grupos familiares que se encuentran en un alto riesgo de caer en pobreza y en consecuencia si no se atienden acérán, irremediablemente e estratos de pobreza extrema o moderada. En: Tuirán Gutiérrez, Alejandro y Alejandro Medina Giopp (2001) El MIC estrategia para mejorar las estimaciones por línea de pobreza y elaboración de recomendaciones para enfrentarla. Material mimeografiado.

Mejía, 1993), más aún cuando éstos pueden resultar controversiales, ya que fijan responsabilidades. Esta concepción, rígida y retrógrada, rompe la posibilidad de lograr retroalimentación y aprendizaje para directivos del programa, para los operadores, así como para los grupos que dan apoyo político y para la comunidad que recibe los bienes y/o servicios del programa.

En efecto, las experiencias de no hacer públicas las evaluaciones de impacto de los programas es una realidad cotidiana en América Latina, lo que frena la posibilidad de generar la capacidad de rendición de cuentas (Valadez & Bamberger, 1994). Afortunadamente, la expansión del movimiento de modernización de la gestión pública, inspirado en las reformas de los países angloamericanos, ha logrado posicionar el tema de la evaluación “publicitada” como uno de los asuntos principales de la modernización a través de diversos Sistemas Integrados (Godoy & Rangel, 1997). Las iniciativas son diversas e interesantes. A modo de ejemplo:

- A inicios de los años 90 el Comité Interministerial de Modernización de la Gestión Pública de Chile lanzó su programa “Metas Ministeriales”, con el que se apoyaba a los diversos servicios públicos para fijar con claridad metas e indicadores de impacto y seguimiento a su gestión. En una segunda etapa otra iniciativa reforzó las metas ministeriales, los “Compromisos de Modernización de la Gestión Pública” con fuerte énfasis en la difusión de resultados. En una tercera etapa se creó el Sistema Integrado de Formulación, Evaluación y Monitoreo de Programas y Proyectos Sociales (SIFEM).
- En Colombia se instaló el Sistema de Evaluación de la Gestión Pública denominado SINERGIA, con estrategias similares a las iniciativas del SIFEM de Chile.
- Costa Rica implementó un Sistema Nacional de Evaluación (SINE) y Compromisos de Modernización, como esfuerzos para apuntalar una cultura de evaluación en la gestión pública.
- Una de las iniciativas más recientes la constituye el Sistema de Evaluación por Resultados (SISER) de Bolivia.
- En el Perú, el Sistema Nacional de Inversión Pública (SNIP) constituye la más reciente preocupación por la difusión pública de los resultados de los proyectos de inversión social.

Aunado a estas iniciativas, los avances en la tecnología de información han permitido transparentar los denominados procesos transversales de la gestión, tales como la administración de recursos financieros, humanos y las adquisiciones, lo que se han denominado genéricamente como sistemas integrados de administración financiera (SIAF); Sistemas integrados de adquisición de bienes y servicios (SIAByS) y Sistemas Integrados de Administración del Personal (SIAP). Finalmente, iniciativas como las procuradurías sociales, el *Ombudsman* (defensor del Pueblo) y otras refuerzan la posibilidad de transitar hacia una gestión pública orientada hacia resultados o creación de valor público, fortaleciendo los mecanismos de rendición de cuentas.

Todo programa social algo positivo deja:

Este es, quizá, uno de los mitos más arraigados que afecta la cultura gerencial de los funcionarios en el área social de la región. El razonamiento es el siguiente: “*todo gasto en proyectos sociales, aunque no se evalúe, o aunque no se apliquen criterios rigurosos en la asignación de los recursos, algo positivo debe dejar*”. Se cree que invertir en lo social siempre algo positivo dejará, por lo que gastar en proyectos sociales es bueno *per se*. Esto supone que siempre y en todas las condiciones, los proyectos logran impactos positivos. Desafortunadamente, este razonamiento es irreal. Evaluaciones de impacto realizadas en programas de naturaleza diversa (vivienda, apoyo nutricional, desarrollo agroforestal, educación, salud y otros muchos) demuestran que en no pocas ocasiones la situación de los beneficiarios (esto es de quienes reciben bienes y servicios del proyecto o proyectos) es peor que la de quienes estaban en condiciones equivalentes a ellos al inicio del programa. Así, en síntesis, el programa puede deteriorar las condiciones de bienestar de la población objetivo en lugar de beneficiarlos; por ello es ineludible la evaluación de impacto.

Por ejemplo, uno de los principales objetivos de un proyecto de vivienda en El Salvador, durante los años 70, fue incrementar el nivel de ingreso de las familias participantes. Para probar el impacto de este programa, aparentemente exitoso, las futuras familias beneficiadas y el grupo de control fueron entrevistados en 1976, poco antes de que el proyecto comenzara y con ello se obtuvo una línea base con las características sociales y demográficas de los hogares. Las familias fueron otra vez entrevistadas en 1978, poco después de que los beneficiarios se mudaran a sus nuevas casas y, posteriormente en 1980 cuando el proyecto estaba totalmente establecido. Para evaluar el impacto del proyecto en el ingreso y desempleo de los jefes de hogar, se calculó el ingreso promedio para los beneficiarios en 1976 y nuevamente en 1980. La tabla siguiente muestra los resultados.

Tabla N° 2.11. Ingreso mensual promedio de las familias que participan en el proyecto de vivienda de El Salvador y para un grupo de Control, 1976-1980 (Fuente: Valadez & Bamberger, 1994).

Grupos	Antes de que el proyecto comience (1976)	Después de que el proyecto ha funcionado por dos años (1980)	Variación absoluta	Variación porcentual
Grupo con proyectos beneficiados	355.0	569.6	234.5	70.0
Grupo de Control	258.3	451.0	172.7	74.6

Este ejemplo demuestra claramente la importancia del grupo de control. Si solo el grupo de beneficiarios se hubiese estudiado (como es frecuente), se pudo asumir que el proyecto produjo un impacto muy significativo sobre el ingreso de los hogares participantes, el que se incrementó en 70% en un periodo de cuatro años. Sin embargo, una vez conocido que el ingreso del grupo de control se incrementó a 74.6% sobre el mismo periodo, es claro que el proyecto no tuvo impacto positivo sobre los ingresos, de hecho pudo tener un pequeño efecto negativo. Queda claro que no todo proyecto social deja algo positivo (Medina Giopp, 2002).

No se puede evaluar porque los resultados se producen en el largo plazo:

Algunos funcionarios afirman que la evaluación de impacto no resulta viable porque los cambios que producirá el proyecto se observarán en el largo plazo. Sin embargo, esto no es del todo cierto. En los casos, por ejemplo, de programas que buscan reducir la tasa de embarazo en adolescentes; desarrollar prácticas higiénicas al preparar alimentos y para tratamiento del agua (cloración, mineralización, hervido, etc.); uso de cinturón de seguridad; mantenimiento de infraestructura social (caminos vecinales, escuelas, canales de riego, etc.); preparación de alimentos balanceados; uso de jeringas descartables; inversión en actividades productivas; utilización de una técnica agrícola específica, etc., las evaluaciones de impacto se realizan a corto y mediano plazo, pues contempla el análisis de “efectos” entendidos como “el grado de uso” de los bienes y/o servicios por parte de los beneficiarios que los reciben. Es decir, es relevante identificar el grado en que los cambios de comportamiento se han alcanzado, ya que si no se ha avanzado en este sentido la posibilidad de tener un impacto significativo en la calidad de vida tiende a ser nula.

Así, entonces, la evaluación de impacto no sólo espera a medir los cambios en las condiciones de bienestar de la población que ha estado sujeta a recibir productos por el periodo determinado de tiempo que implica el programa o proyecto, sino que puede comenzar con la evaluación de los efectos o también denominados como “*condicionantes del impacto*” ya que de no darse las probabilidades de alcanzarlo, se reducen (Medina Giopp, 2002).

Un proyecto es bueno *per se*, sin importar la magnitud del problema a atender:

Si un proyecto es exitoso al demostrar que gracias a su intervención se mejoran las condiciones de bienestar de los beneficiarios, el proyecto indudablemente contribuye de manera positiva a la solución de un problema. Sin embargo es importante considerar que los proyectos se articulan en programas y éstos en políticas y, por tanto, no se puede afirmar que la inferencia sea exitosa *per se*. Por ejemplo, suele ocurrir que a nivel de proyecto se conocen bien los bienes o servicios que se entregan y si éstos producen (o no) impacto, pero no se conoce la base de la población que posee la carencia. Bajo esta lógica, un proyecto puede ser considerado exitoso ya que genera impacto en los beneficiarios, pero si estos constituyen un porcentaje muy bajo del total de la población carente en el país, por ejemplo 2%, entonces es fácilmente cuestionable dicho éxito y se puede poner en duda la efectiva articulación del proyecto a programas y políticas.

2.3.4. Técnicas de evaluación de impacto

La evaluación de impacto hace uso de diversas técnicas de investigación. Algunos autores han realizado algunas clasificaciones basándose en uno u otro aspecto. Abdala (2001), por ejemplo, hace una clasificación de acuerdo a enfoques y las ordena en tres: las pseudos-evaluaciones, evaluaciones exclusivamente

cuantitativas (evaluaciones experimentales y cuasi-experimentales) y las evaluaciones verdaderas. Esta clasificación, similar con la de Baker (2000) y Diez de Medina (2003), resalta la conveniencia de aplicar métodos combinados para mejorar los resultados de las evaluaciones de impactos. En la tabla siguiente se puede ver esta clasificación y las características de cada enfoque.

Tabla N° 2.12. Enfoques de evaluación de impacto (Fuente: Abdala, 2001).

Enfoque	Descripción
Pseudo evaluaciones	<ul style="list-style-type: none"> • En ellas se cumple un proceso evaluatorio pero con fuerte injerencia de alguna de las partes comprometidas con el proyecto y que presiona para que aparezcan resultados preestablecidos. • El evaluador no actúa neutralmente.
Evaluaciones experimentales y cuasi-experimentales	<ul style="list-style-type: none"> • Son evaluaciones exclusivamente cuantitativas. Este diseño exige que se constituya un grupo de control. Se mide el impacto por comparación estadísticas entre el grupo de control y el beneficiario de las acciones del programa. • Se mide el incremento del bienestar de los beneficiarios.
Evaluaciones verdaderas mixtas cuanti-cualitativas	<ul style="list-style-type: none"> • Las evaluaciones mixtas permiten agregar elementos cualitativos como: cambios actitudinales, cambios psicosociales, necesidad y satisfacción con el programa, autoestima, empleabilidad. • Pueden medirse impactos sobre otros actores. Se utilizan técnicas como estudios de casos, observaciones, entrevistas en profundidad. • El aporte mixto ayuda a entender más el por qué unos programas son exitosos y otros fracasan.

A continuación se describe, brevemente, las técnicas de evaluación de impacto, siguiendo la presentación de Baker (2000).

2.3.4.1. Evaluaciones cuantitativas

Las evaluaciones cuantitativas son aquellas que miden el impacto por comparación entre un grupo de control y el grupo beneficiario en una intervención. Estas evaluaciones pueden ser experimentales o cuasi experimentales.

Evaluaciones experimentales:

Los diseños experimentales son los procedimientos de evaluación más sólidos. Estos diseños usan grupos de control para medir el impacto de los programas. Distribuyen aleatoriamente a los participantes en grupos de tratamiento y de control que son estadísticamente equivalentes entre sí, para determinar los efectos del mismo. Se trata de un resultado muy convincente porque, en teoría, los grupos de control generados mediante asignación aleatoria sirven como un escenario contrafactual perfecto, sin los dificultosos problemas de sesgo de selección.

Aunque los diseños experimentales son –metodológicamente hablando- el método óptimo para estimar el impacto de un proyecto, en la práctica tiene una serie de limitaciones aplicativas (Baker, 2000) y que cuestionan su validez:

1. La aleatorización podría ser poco ética debido a la negación de beneficios o servicios a miembros de la población que serían calificados como grupo control (Fernández-Ballesteros, 1996).
2. Puede ser políticamente difícil proporcionar una intervención a un grupo y no a otro. El grupo control podría reclamar y sabotear la intervención con el grupo experimental.
3. El alcance del programa podría significar que no hubiera grupos sin tratamiento, como en el caso de un proyecto o cambio de política de amplio alcance (nivel nacional).
4. Durante el experimento los individuos de los grupos de control podrían cambiar ciertas características que los identifican, lo que podría invalidar o contaminar los resultados. Por ejemplo, si las personas se trasladan hacia un área de proyecto o salen de él, podrían ingresar o salir del grupo de tratamiento o de control. Alternativamente, las personas a las que se les niega el beneficio del programa pueden buscarlo a través de otras fuentes, o bien aquellas a las que se les ofrece un programa podrían no aceptar la intervención.
5. Podría resultar difícil garantizar que la asignación sea realmente aleatoria. Un ejemplo de esto podrían constituirlo los administradores que excluyen solicitantes de alto riesgo para lograr mejores resultados.
6. Y, por último, los diseños experimentales pueden ser costosos y prolongados en ciertas situaciones, especialmente en la recopilación de nuevos datos.

Pese a estas limitaciones de aplicación, lo cierto es que con una planificación cuidadosa se pueden abordar algunos de ellos al momento de implementar los diseños experimentales. He aquí algunas experiencias:

- Una forma es la selección aleatoria de los beneficiarios (Baker, 2000). Esto se puede aplicar para proporcionar un mecanismo de distribución políticamente transparente y la base de un diseño de evaluación sólido, puesto que las restricciones de presupuesto o de información con frecuencia hacen imposible identificar y llegar de manera precisa a los beneficiarios más calificados.
- Una segunda forma es ingresar los grupos de control al programa en una etapa posterior, una vez que se ha diseñado y se ha iniciado la evaluación. Esto se aplicó en la evaluación de un programa de nutrición en Colombia, proporcionando la ventaja adicional de abordar preguntas con respecto al tiempo necesario para que el programa sea eficaz para reducir la desnutrición (McKay et al, 1978).
- Por último, se puede aplicar la aleatorización dentro de un subconjunto de beneficiarios igualmente calificados, llegando al mismo tiempo a todos los más calificados y negando los beneficios a los menos calificados, como se hizo con los proyectos de educación en la región del Chaco para la evaluación del fondo social en Bolivia (Pradhan, Rawlings y Ridder, 1998). Sin embargo, si se implementa esta última sugerencia, se debe tener en cuenta que los resultados que arroje la

evaluación serán válidos para el grupo del cual se tomó la muestra generada aleatoriamente.

Evaluaciones cuasi-experimentales:

Los diseños cuasi-experimentales se utilizan cuando es imposible crear grupos de control y tratamiento (Shadish, Cook & Campbell, 2002). Estas técnicas generan grupos de comparación que se asemejan al grupo de tratamiento, al menos en las características observadas, usando metodologías econométricas que incluyen métodos de pareo (Matching), métodos de doble diferencia, métodos de variables instrumentales o comparaciones reflexivas. Cuando se usan estas técnicas, los grupos de tratamiento y de comparación por lo general se seleccionan después de la intervención usando métodos no aleatorios. Por lo tanto, se deben aplicar controles estadísticos para abordar las diferencias entre los grupos de tratamiento y de comparación y emplear técnicas sofisticadas de pareo para crear un grupo de comparación que sea en lo posible lo más similar al grupo de tratamiento. En algunos casos también se selecciona un grupo de comparación antes del tratamiento, aunque la selección no es aleatoria.

La ventaja principal de los diseños cuasi-experimentales es que se pueden basar en fuentes de datos existentes y, por lo tanto, a menudo son más rápidos y menos costosos en implementar. Además, se pueden realizar una vez que el programa se ha implementado, a condición de que existan suficientes datos. Las desventajas principales de las técnicas cuasi-experimentales son que (a) con frecuencia se reduce la confiabilidad de los resultados, puesto que la metodología es menos sólida estadísticamente, (b) los métodos pueden ser estadísticamente complejos y (c) conllevan un problema de sesgo de selección (Anguera et al, 1995). Al generar un grupo de comparación en lugar de asignarlo aleatoriamente, hay muchos factores que pueden afectar la confiabilidad de los resultados (Cook y Campbell, 1979; Cook, Campbell y Peracchio, 1990). La complejidad estadística requiere conocimientos especializados considerables en el diseño de la evaluación y el análisis e interpretación de los resultados. Esto no siempre es posible, especialmente en las condiciones de los países en desarrollo (Baker, 2000).

El tercer problema de sesgo se relaciona con la medida en la cual los subgrupos de una población beneficiaria participan en forma diferenciada en el programa, afectando así la muestra y, finalmente, los resultados. Hay dos tipos de sesgo: aquellos causados por las diferencias en los elementos observables o algún elemento de los datos y aquellos causados por las diferencias en los elementos no observables (no en los datos), lo que con frecuencia se denomina sesgo de selección. Un sesgo observable podría incluir los criterios de selección mediante los cuales se escoge a un individuo, como ubicación geográfica, asistencia a la escuela o participación en el mercado laboral. Los no observables que pueden sesgar los resultados de un programa podrían incluir la capacidad individual, la disposición al trabajo, los vínculos familiares y un proceso subjetivo (con frecuencia guiado por la política) de seleccionar individuos para un programa. Ambos tipos de sesgos pueden generar resultados inexactos, como subestimar y sobrestimar los efectos reales del programa, efectos negativos cuando los efectos reales del programa son positivos (y viceversa) y efectos estadísticamente

insignificantes cuando los efectos reales del programa son relevantes y viceversa. (Véase, por ejemplo, La Londe, 1986, Fraker & Maynard, 1987, LaLonde & Maynard, 1987, y Friedlander & Robins, 1995). Es posible controlar por el sesgo mediante técnicas estadísticas, como variables de comparación e instrumentales, pero es muy difícil eliminarlo por completo, siendo así un importante desafío para los investigadores en el campo del análisis de los efectos.

Entre las técnicas de diseño cuasi-experimental en general se considera que las técnicas de comparación pareada son la alternativa sub-óptima al diseño experimental. Gran parte de la literatura sobre metodologías de evaluación se centra en el uso de este tipo de evaluaciones, lo que indica el frecuente uso de las comparaciones pareadas y los numerosos desafíos que plantea el contar con grupos de comparación poco adecuados.

En los últimos años se han producido significativos avances en las técnicas de correspondencia de puntuación de la propensión (Rosenbaum & Rubin, 1985; Jalan & Ravallion, 1998). Este método es muy atractivo para los evaluadores que tienen restricciones de tiempo y no tienen la ventaja de contar con datos básicos (de referencia), dado que se pueden usar con una simple sección transversal de datos. Sin embargo, para esta técnica se debe contar con los datos adecuados, porque se basa en tomar nuevas muestras de los beneficiarios durante la aplicación en terreno de una encuesta más amplia, “haciéndolos corresponder” luego con un grupo de comparación seleccionado de la muestra básica más amplia de la iniciativa global, frecuentemente una encuesta domiciliaria a nivel nacional. Dado el crecimiento de las aplicaciones de grandes encuestas en los países en desarrollo, como las encuestas sobre las condiciones de vida de propósitos múltiples, este método de evaluación es particularmente prometedor (Jalan & Ravallion, 1998).

2.3.4.2. Evaluaciones cualitativas

Para realizar una evaluación del impacto también se usan técnicas cualitativas, en un intento por determinar el efecto basándose en algo diferente al escenario contrafactual (Mohr, 1995). En este caso, se trata de comprender los procesos, comportamientos y condiciones como las perciben los individuos o grupos estudiados (Valadez & Bamberger, 1994). Por ejemplo, los métodos cualitativos y, en particular, la observación de los participantes, pueden proporcionar información sobre las formas en que los beneficiarios perciben un proyecto y cómo se ven afectados por éste. Puesto que medir el escenario contrafactual es esencial para las técnicas de análisis de los efectos, los diseños cualitativos en general se han usado en conjunto con otras técnicas de evaluación.

Algunas de las técnicas utilizadas en este tipo de evaluaciones son las que contienen la evaluación rural rápida o la planificación participativa (The World Bank Participation Sourcebook, 1996). La principal ventaja de las evaluaciones cualitativas es su flexibilidad y adaptabilidad a las necesidades de la evaluación. Permiten mejorar una evaluación de impacto al permitir mayor comprensión de las percepciones y prioridades de las partes interesadas y los diferentes factores que puedan haber afectado a los resultados de un programa (Baker, 2000).

Sin embargo, existen aún serias desventajas, entre ellas: la subjetividad involucrada en la recopilación de datos, la falta de un grupo de comparación y la falta de solidez inferencial dados los reducidos tamaños de las muestras. Además, la validez y fiabilidad de los datos cualitativos dependen en gran medida de la habilidad metodológica, sensibilidad y capacitación del evaluador (Miles & Huberman, 1994; Taschereau, 1998). Si el personal en terreno no es sensible a las normas y prácticas sociales y culturales específicas y a los mensajes no verbales, los datos reunidos pueden ser mal interpretados. Y por último, sin un grupo de comparación, es imposible determinar el escenario contrafactual y, por lo tanto, la causalidad del efecto del proyecto (ILPES & CEPAL, 2003; Abdala, 2000).

2.3.4.3. Evaluaciones mixtas cuali - cuantitativas

Con este enfoque se combina datos cuantitativos (de muestras estadísticas que son más adecuadas para evaluar la causalidad usando métodos econométricos o para establecer conclusiones que se puedan generalizar) con información obtenida por métodos cualitativos sobre las perspectivas de los beneficiarios, la dinámica de algunos procesos o los motivos que expliquen algunos resultados obtenidos por medio de métodos cuantitativos (Abdala, 2000; ILPES & CEPAL, 2003).

En efecto, aunque existe abundante literatura donde se compara los métodos cuantitativos con los cualitativos en la evaluación del impacto, cada vez se acepta más la integración de los dos enfoques (Rao & Woolcock, 2003). En efecto, las evaluaciones de impacto que se basan en datos cuantitativos de muestras estadísticamente representativas son más adecuadas para evaluar la causalidad usando métodos econométricos o llegando a conclusiones que se pueden generalizar. Sin embargo, los métodos cualitativos permiten estudiar cabalmente los temas, casos o hechos seleccionados y pueden proporcionar información decisiva sobre las perspectivas de los beneficiarios, la dinámica de un determinado programa o los motivos de ciertos resultados observados en un análisis cuantitativo.

Actualmente se afirma que la integración de las evaluaciones cuantitativas y cualitativas puede ser el mejor vehículo para satisfacer las necesidades de evaluación de un proyecto (Baker, 2000). Al combinar los dos enfoques, los métodos cualitativos se pueden usar para informar las preguntas clave sobre la evaluación del impacto, examinar el cuestionario o la estratificación de la muestra cuantitativa y analizar el marco social, económico y político dentro del cual se lleva a cabo un proyecto. Los métodos cuantitativos, en tanto, se pueden usar para informar las estrategias de recopilación de datos cualitativos, diseñar la muestra para informar la medida en que los resultados observados en el trabajo cualitativo son aplicables a una población de mayor tamaño al usar una muestra estadísticamente representativa. Y finalmente, el análisis estadístico se puede usar para controlar las características de los hogares y las condiciones socioeconómicas de diferentes áreas de estudio, con lo que se eliminan las explicaciones alternativas de los resultados observados (Baker, 2000).

Existen varias ventajas en usar enfoques integrados en la investigación (Bamberger, 2000), los que también se aplican a las evaluaciones de impacto. Entre ellas:

- Se puede incorporar controles de coherencia con la ayuda de procedimientos de triangulación que permiten realizar dos o más estimaciones independientes de las variables clave (como ingreso, opiniones sobre los proyectos, motivos para usar o no los servicios públicos y efecto específico de un proyecto).
- Se pueden obtener diferentes perspectivas. Por ejemplo, aunque los investigadores podrían considerar el ingreso o el consumo como indicadores clave del bienestar de un hogar, los estudios de casos podrían revelar que las mujeres están más preocupadas por la vulnerabilidad (definida como la falta de acceso a sistemas de apoyo social en tiempos de crisis), impotencia o exposición a la violencia.
- Los análisis se pueden conducir en diferentes niveles. Los métodos de encuestas pueden proporcionar buenas estimaciones del bienestar individual, familiar y a nivel de la comunidad, pero son mucho menos eficaces para analizar los procesos sociales (conflicto social, motivos para usar o no usar los servicios, etc.) o para el análisis institucional (qué tan eficazmente funcionan los servicios de salud, educación, crédito y otros y cómo son percibidos por la comunidad). Por el contrario, existen muchos métodos cualitativos que están diseñados para analizar temas como el proceso social, comportamiento institucional, estructura social y conflictos.
- Se puede proporcionar oportunidades de retroalimentación para interpretar los resultados. Los informes de encuestas a menudo incluyen referencias a aparentes incoherencias en los resultados o a diferencias interesantes entre las comunidades o grupos que no se pueden explicar con los datos. En la mayoría de las investigaciones cuantitativas, una vez que finaliza la fase de recolección de datos no es posible regresar al terreno para verificar esos aspectos. La mayor flexibilidad de la investigación cualitativa significa que a menudo es posible regresar al terreno para reunir datos adicionales. De hecho, los investigadores de encuestas también utilizan métodos cualitativos para verificar los valores atípicos (respuestas que se desvían de los patrones generales).

En la práctica, la integración de los métodos cuantitativos y cualitativos se debe realizar durante cada paso de la evaluación del impacto.

2.3.4.4. Otros enfoques de evaluación

Baker (2000) documenta otros enfoques que son pertinentes al momento de analizar la evaluación de las repercusiones de los proyectos en la pobreza: (a) los enfoques para medir el efecto de los programas de ajuste estructural, (b)

evaluaciones basadas en la teoría, y c) análisis de la eficacia en función de los costos.

Evaluación de ajuste estructural: Este tipo de evaluaciones se aplica cuando se modifica, mediante políticas integrales, la estructura económica de una nación. Sin embargo, en la práctica se encuentran limitaciones: a) los cambios de políticas tienen un efecto a nivel de toda la economía, haciendo imposible encontrar grupos de comparación, b) debido a factores exógenos, retrasos, retroalimentaciones y sustituciones, cualquier cambio en el bienestar de los pobres debe interpretarse con extrema prudencia, c) es difícil predecir lo que habría sucedido si no se hubiera realizado el ajuste, esto es, qué políticas alternativas habría seguido un gobierno y cómo habría afectado a los pobres (Killick, 1995; Poppele, Summarto & Pritchett, 1999; Bourguignon, De Melo & Suwa, 1991; Sahn, Dorosh & Younger, 1996).

Evaluación basada en la teoría: La premisa de las evaluaciones basadas en la teoría es que los programas y proyectos se basan en teoría explícita o implícita acerca de cómo y por qué funcionará un programa. Entonces, la evaluación se basará en apreciar cada teoría y los supuestos acerca de un programa durante la fase de implementación, en lugar de hacerlo en un momento intermedio o cuando el proyecto ha finalizado. Al diseñar la evaluación, la teoría implícita se presenta en la forma de muchos micropasos, creándose entonces los métodos para que la recolección y el análisis de datos hagan un seguimiento de la exposición de los supuestos. Si los sucesos no resultan como se esperaba, la evaluación puede afirmar con un cierto nivel de confianza dónde, por qué y cómo ocurrió la falla (Baker, 2000).

El enfoque se centra en las respuestas de las personas a las actividades del programa. Las teorías dirigen la atención del evaluador a los probables tipos de efectos en el corto y largo plazo. Algunas de las ventajas son, en primer lugar, que la evaluación proporciona indicios oportunos de la eficacia del programa durante la implementación del proyecto. Si hay fallas durante esta fase, es posible solucionarlas durante el proceso. En segundo lugar, el enfoque permite explicar cómo y por qué se produjeron los efectos. Si los sucesos resultan como se esperaba, la evaluación puede indicar con cierta confianza cómo se generaron los efectos. Siguiendo la secuencia de las etapas, es posible hacer un seguimiento de los micropasos que llevan desde los aportes del programa hasta los resultados. Las falencias del enfoque son similares a muchas de las demás metodologías: a) la identificación de los supuestos y teorías puede ser inherentemente complejo; b) los evaluadores pueden tener problemas para medir cada paso, a menos que cuenten con los instrumentos y datos adecuados, c) se podrían enfrentar problemas al examinar la iniciativa, porque los planteamientos teóricos podrían ser demasiado generales y estar contruidos en forma demasiado inexacta para permitir un estudio bien definido y d) podría haber problemas de interpretación que dificultarían la generalización a partir de los resultados (Weiss, 1998; Baker, 2000).

Análisis de eficacia en función de los costos: Aunque este tipo de análisis no se relaciona estrictamente con la medición de los efectos, permite que las autoridades

responsables midan la eficiencia de los programas comparando intervenciones alternativas sobre la base del costo de producir un resultado determinado. Puede aumentar enormemente las implicaciones de política de la evaluación de impacto y, por lo tanto, también se debería incluir en el diseño de cualquier evaluación de impacto³⁵.

Con el análisis de costos-beneficios se intenta medir, en términos monetarios, la eficiencia económica de los costos de un programa versus sus beneficios. Para muchos proyectos, especialmente en los sectores sociales, no es posible medir todos los beneficios en términos monetarios. Por ejemplo, los beneficios de que un programa proporcione insumos escolares (libros de texto, mobiliario para la sala, programas preescolares) serían el aumento del aprendizaje. En este caso, en lugar de medir los resultados monetarios, se podrían usar las puntuaciones del progreso en materia de aprendizaje para cuantificar los beneficios. Para ello se necesitaría realizar un análisis de la eficacia en función de los costos. Los conceptos para ambos tipos de análisis son los mismos.

Los principales pasos del análisis de costo-beneficio y eficacia en función de los costos son la identificación de todos los costos y beneficios del proyecto y luego calcular una razón costo-eficacia. Al calcular los costos, se incluye el valor de la intervención misma, al igual que todos los demás costos, como costos administrativos, de entrega, inversión, el valor monetario de los bienes o servicios proporcionados en forma gratuita, los costos sociales (como el deterioro ambiental y los riesgos para la salud). Los beneficios pueden ser monetarios, como el aumento en el ingreso, o bien el número de unidades entregadas, valoraciones de las pruebas o mejoramientos de salud. Cuando no se pueden cuantificar los beneficios, generalmente se usan indicadores subjetivos, como sistemas de clasificación o ponderación; sin embargo, este enfoque es complicado para interpretar valoraciones subjetivas (Baker, 2000).

Una vez determinados los costos y beneficios, la razón eficacia en función de los costos (R) es entonces $R = \text{costo/unidad (o beneficio)}$. Esta razón se puede comparar en todas las intervenciones para medir la eficiencia. En teoría, esta técnica es bastante directa. En la práctica, sin embargo, hay muchos riesgos involucrados en la identificación y cuantificación de los costos y beneficios (Baker, 2000). Al respecto, es importante asegurarse que se seleccionen los indicadores adecuados, que las metodologías y supuestos económicos usados sean coherentes en todas las razones y que las razones de hecho sean comparables.

2.3.4.5. ¿Cómo elegir el método de evaluación de impacto?

Dada la multiplicidad de los tipos de proyectos, preguntas de la evaluación, disponibilidad de datos, restricciones de tiempo y circunstancias, cada estudio de evaluación de impacto será diferente y requerirá cierta combinación de las metodologías adecuadas, tanto cuantitativas como cualitativas. El evaluador debe

³⁵ Un estudio completo del análisis de costos-beneficios y eficacia en función de los costos se encuentra en el *Handbook on Economic Analysis of Investment Operations* del Banco Mundial (1996).

examinar con cuidado las opciones metodológicas al diseñar el estudio, con el objetivo de producir los resultados más sólidos posibles (Baker, 2000).

En general, las metodologías elegidas para la evaluación del impacto no son mutuamente excluyentes. Desde luego, en las evaluaciones más robustas con frecuencia se combinan métodos para asegurar su solidez y prever las contingencias en su implementación (Subbarao et al, 1999; GAO, 1991)³⁶.

En general, para saber cuál es el método de evaluación de impacto apropiado para cada caso, es recomendable que el evaluador considere los siguientes criterios:

- **Riesgos de perjudicar la validez:** El criterio más importante para valorar un método es la validez de sus estimaciones del impacto del programa. Aunque todos los métodos son vulnerables a algunos riesgos a la validez, varían considerablemente en términos del número y tipos de riesgos a las cuales están sujetos. En el capítulo IV se analiza en detalle las amenazas peculiares para cada tipo de diseño. Sin embargo, en la Tabla 2.13 se presenta una comparación de diseños según su calidad metodológica (Bamberger, 2006).
- **Habilidad para aislar los efectos del programa:** Idealmente, las medidas del impacto del programa incluyen solo resultados directamente atribuibles al programa. En la mayoría de los escenarios, factores como las fuerzas de desarrollo económico, programas sociales múltiples, estructura demográfica cambiante y la presencia de actividades de planificación familiar fuera del programa complican los intentos de medir el impacto del programa. Los métodos difieren en la medida en que el evaluador puede aislar los efectos del programa de las influencias de otros factores. Los métodos preferibles son aquellos que sean más eficientes para aislar estos efectos.
- **Costos:** Esto se refiere a los costos de la recolección y análisis de datos. En igualdad de condiciones, se prefieren los métodos de menor costo. No debe olvidarse que las limitaciones presupuestarias influyen en el número de entrevistas que pueden realizarse, la capacidad de combinar la recopilación y el análisis de datos cualitativos y cuantitativos, el tipo de experiencia profesional del equipo de investigación y el análisis que puede llevarse a cabo (Bamberger, 2006).
- **Datos requeridos:** Los métodos varían considerablemente en cuanto a los datos requeridos. Aparte de las diferencias en el volumen de datos requeridos, algunos métodos necesitan datos que son muy difíciles de recolectar y/o son más vulnerables a errores de medición que otros métodos. Esto puede aumentar el riesgo de que el error de medición

³⁶ También es recomendable incorporar el análisis de costo-beneficio o eficacia en función de los costos (Baker, 2000). Esta metodología permite que las autoridades responsables comparen intervenciones alternativas sobre la base del costo de producir un resultado determinado. Esto es de particular importancia en el marco de los países en desarrollo, en que los recursos son extremadamente limitados. Por último, es ideal combinar métodos cuantitativos y cualitativos, porque proporcionará el efecto cuantificable de un proyecto y una explicación de los procesos e intervenciones que originaron estos resultados.

oscurezca los efectos de los programas o exagere la magnitud del impacto logrado realmente. Cuando la evaluación se basa en datos secundarios o cuando los datos proceden de estudios realizados por otros organismos (estudios concatenados o sincronizados), las limitaciones de información pueden mermar la compatibilidad de la cobertura o cronología de la muestra o la posibilidad de que los datos cubran las variables necesarias o las definan adecuadamente (Bamberger, 2006).

- **Percepción de la naturaleza de las trayectorias causales:** Los métodos varían mucho en relación a la cantidad de información que proporcionan sobre cómo los insumos son transformados a productos y resultados como parte del proceso de medición del impacto. Aunque no se requiere para la medición del impacto, esta información de los mecanismos causales provee una percepción muy útil de cómo pueden mejorarse los programas en ciclos de programa subsiguientes.
- **Requisitos en el grado de control del programa:** Los experimentos diseñados como tales (los experimentos al azar y en menor grado los estudios cuasi-experimentales) proveen la evidencia más importante del impacto del programa, pero también requieren condiciones de mayor control en la forma en la cual el programa es evaluado y la forma en que se efectúan otras intervenciones. Los estudios no experimentales no requieren que los programas sean implementados de manera específica para proveer mediciones válidas del impacto del programa, pero generalmente requieren mayor cantidad de datos y análisis más complejos para producir resultados válidos.
- **Recursos y destrezas técnico-estadísticas requeridas:** Aunque todos los métodos considerados requieren de conocimiento y destrezas básicas en las áreas de la investigación y la estadística, algunos métodos y enfoques requieren destrezas relativamente avanzadas y en algunos casos un *software* especializado (Ravallion, 2005). Además, algunos métodos están diseñados específicamente para la medición de ciertos tipos de resultados y debe tenerse en cuenta esa relación.
- **Tiempo:** Las limitaciones temporales tienen que ver con el momento en que se inicia y termina la evaluación, el tiempo que pueden pasar los investigadores sobre el terreno y el tiempo disponible para recibir información de las partes interesadas. Debe tenerse en cuenta la disponibilidad del tiempo y la temporalización del programa (Bamberger, 2006; Bedi, Bhatti, Gine, Galasso, Goldstein & Legovini, 2006).

En la siguiente tabla se presenta las opciones de diseño para reducir los costos en la obtención de datos (mediante encuestas) en un estudio de EI, considerando también la robustez del diseño para obtener inferencias causales confiables.

Tabla N° 2.13. Opciones de diseño para reducir los costos de la obtención de datos³⁷ (Fuente: Bamberger, 2006).

Diseño de la evaluación	Referencia	Tratamiento	Mitad de período	Después de la intervención	Porcentaje de ahorro con respecto al Diseño 1	Observaciones
Diseño de referencia: diseño robusto de evaluación de impacto de uso general						
1. Diseño con grupo de beneficiarios y de control antes y después de la intervención con asignación aleatoria y no aleatoria.	P1 C1	X		P2 C2	0	Diseño más sólido de todos en la mayoría de los casos reales en que la evaluación comienza al mismo tiempo que el proyecto.
Diseños progresivamente menos robustos						
2. Diseño aplazado de grupo de comparación antes y después de la intervención. La evaluación no comienza hasta mediados del período de ejecución.		X	P1 C1	P2 C2	0	Diseño útil cuando la evaluación comienza a mediados del período. Si la ejecución se retrasa, no será mucho más débil que el Diseño 1. Eventual ahorro moderado porque el período del contrato de evaluación es más breve.
3 Grupo de beneficiarios antes y después de la intervención y grupo de comparación sólo después de ésta.	P1	X		P2 C2	25	Aunque es más débil que los diseños anteriores, es relativamente sólido, ya que permite hacer comparaciones a lo largo del tiempo y análisis transversales con posterioridad al proyecto.
4. Grupo de beneficiarios y de comparación posteriores a la intervención sin datos de referencia.				P1 C1	50	Diseño frecuente cuando la evaluación comienza hacia el final del ciclo del proyecto o cuando éste ha terminado. La falta de datos de referencia dificulta el control de las diferencias iniciales entre los dos grupos, pero eso se puede resolver en parte con puntuaciones de propensión. Sin embargo, el uso de puntuaciones de propensión puede requerir muestras más grandes.

³⁷ La reducción de costos se basa en la hipótesis de que se necesitarán encuestas u otros métodos de obtención de datos primarios. Las reducciones estimadas de los costos de recopilación de datos, en comparación con el Diseño 1, dependen del número de puntos de obtención de datos (antes y después de la intervención, y grupo de beneficiarios y de comparación) eliminados. En el caso de encuestas concatenadas o sincronizadas, o cuando la evaluación se basa en datos secundarios, el ahorro puede ser mayor.

Diseño de la evaluación	Referencia	Tratamiento	Mitad de período	Después de la intervención	Porcentaje de ahorro con respecto al Diseño 1	Observaciones
Dos diseños muy utilizados que, por falta de una hipótesis aceptable, casi nunca son considerados como diseños de impacto de calidad						
5. Análisis del grupo de beneficiarios antes y después de la intervención sin grupo de comparación.	P1			P2	50	Diseño frecuente cuando sólo se compilan datos sobre el grupo de beneficiarios. Metodológicamente débil, ya que el uso de P1 como hipótesis requiere suposiciones audaces sobre la variación de los efectos con el paso del tiempo y las variables individuales no observables.
6. Grupo de beneficiarios después de la intervención sin datos de referencia ni grupo de comparación.				P1	75-90	Es el diseño más débil de todos, pero se utiliza con frecuencia cuando hay que realizar evaluaciones hacia el final del proyecto y con poco tiempo y recursos. Se emplean métodos cualitativos, los registros del proyecto y datos secundarios agregados para estimar la hipótesis.
<p>Símbolos: P = grupo de beneficiarios de selección aleatoria o no aleatoria. C = grupo de control o comparación de selección aleatoria o no aleatoria (no equivalente). P1, P2, C1, C2 indican la primera y la segunda utilización del instrumento de encuesta con los grupos beneficiarios y de comparación, respectivamente.</p>						

2.3.5. Principios metodológicos de la evaluación de impacto

Basado en la revisión bibliográfica (Ej. Blomquist, 2003; Ezemanari et al., 1999; Mohr, 1995), se puede plantear la existencia de algunos principios metodológicos que rigen el proceso de evaluación de impacto. Se presenta los cinco principales.

Principio 1. Determinar la hipótesis de desarrollo:

Desde el punto de vista estrictamente económico, un proyecto es un proceso de transformación de insumos (o recursos) en productos. De lo dicho, un programa es más exitoso cuanto más valor agrega a los insumos. Así, por regla general, sólo deberían emprenderse aquéllos proyectos que crean valor para la sociedad (Medianero, 2001), lamentablemente en los países en desarrollo ésta suele ser una situación más bien excepcional, ya que la continuidad o no de un proyecto se decide bajo criterios subjetivos o políticos. Así, los programas sociales establecen, implícitamente, una relación de causa a efecto, que la evaluación de impacto deberá confirmar o recusar. Por ello, es un principio metodológico considerar que las hipótesis contenidas en el proyecto (hipótesis sobre el desarrollo o hipótesis de cambio) son proposiciones sujetas a comprobación empírica, a través precisamente de la evaluación de impacto del programa. Este principio se fundamenta, en mi entender, en la teoría de la causación (inferencia causal).

Principio 2. Enfocar los resultados y no los procesos:

Este principio estipula que desde el punto de vista de una evaluación de impacto, no importan las actividades y el cúmulo de indicadores de proceso a través de los cuales se realiza el programa, sino los productos generados a través de los procesos internos.

En efecto, las actividades y los insumos requeridos para su realización son un asunto interno, que bien puede ser importante para la gerencia diaria del proyecto, y que de hecho es parte importante del monitoreo de procesos, pero que carece de relevancia desde la perspectiva de la evaluación de impacto.

Las burocracias tradicionales e ineficientes suelen “medir” sólo los insumos (lo más común es medir el cumplimiento en la ejecución del presupuesto como si eso fuera un mérito por sí mismo) o, a lo mucho, la entrega de algunos materiales (libros de texto escolares, alimentos, etc.). Una evaluación de impacto es un análisis de los resultados del proyecto y éstos se pueden expresar en tres niveles sucesiva y secuencialmente superiores: a) productos, b) efectos e c) impactos.

1. Los **productos** son los bienes o servicios que un proyecto entrega a la población o que han sido producidos a fin de lograr determinados efectos en las personas, instituciones o sistemas. Por ejemplo, el Programa Nacional Wawa Wasi contempla la entrega de los siguientes productos: (a) atención integral de los niños mediante los servicios de salud, alimentación y estimulación; (b) capacitación de todos los participantes en la ejecución del programa; (c) promoción

- y difusión de la cultura de crianza; (d) financiamiento de la construcción de los Wawa Wasis comunales y equipamiento; y, (e) gestión de programa.
2. Los **efectos** son los cambios en las condiciones de vida de la gente, de las instituciones o del entorno para el desarrollo, que se espera que ocurran como consecuencia de la combinación de los productos entregados. Los efectos son los resultados que se esperan a nivel del propósito (dentro del contexto del marco lógico). El propósito por definición especifica el cambio o resultado final deseado, el lugar donde ocurrirá dicho cambio y la población que será afectada con dicho cambio, incluyendo la especificación de cualquier diferencia sociodemográfica.
 3. Los **impactos** describen los objetivos nacionales o sectoriales de desarrollo que el proyecto pretende lograr, o más precisamente, a los cuales pretende contribuir. Bajo el enfoque del marco lógico, el impacto se refiere al “fin” del proyecto. En el caso del Programa Nacional Wawa Wasi, el efecto esperado consiste en *"promover en todo el país servicios y acciones orientadas a generar condiciones favorables para el desarrollo integral de niños y niñas, particularmente aquellos en situación de riesgo"*. Por su parte, el impacto esperado está expresado en la frase siguiente: *"Contribuir con el desarrollo humano de niñas y niños del país, mejorando sus niveles de bienestar y promoviendo su desarrollo integral"*.

Las evaluaciones de impacto se centran, pues, en los resultados, entendidos éstos como los productos, efectos e impactos. Sin embargo, por lo general, las evaluaciones de impacto enfatizan más en el análisis de los efectos e impactos propiamente dichos, que los productos del proyecto³⁸. En la siguiente figura esquematiza la relación:

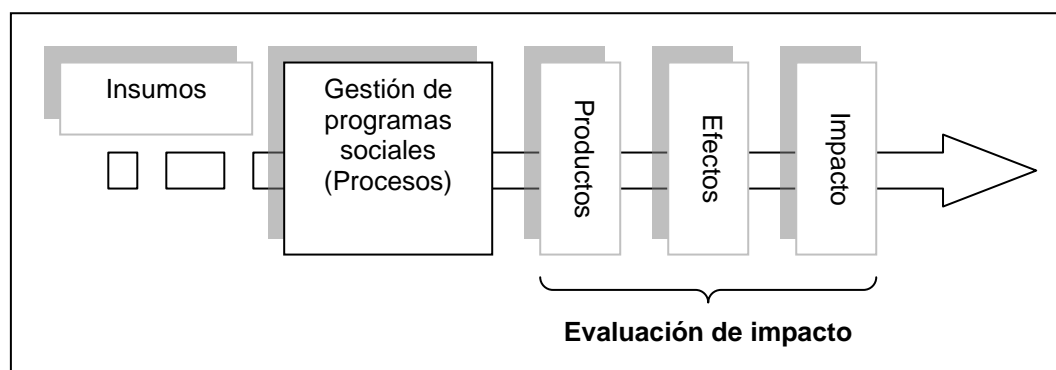


Figura N° 2.3. La conversión de insumos en resultados (Fuente: Elaboración propia).

Principio 3. Usar indicadores objetivamente verificables.

Dada la variable complejidad que presentan los distintos proyectos de inversión pública, resultaría demasiado controversial evaluarlos tomando como

³⁸ En ocasiones se habla de impacto a nivel proyecto (propósito) e impactos a nivel global (fin), pero esta distinción es propia del Marco Lógico, donde se diferencian ambos aspectos.

base las proposiciones literales de los objetivos, ya sea a nivel de fin, propósito o productos. Debido a ello, los evaluadores recurren a expresiones cuantitativas de los objetivos, las cuales, bajo el enfoque del marco lógico, son conocidas como “indicadores objetivamente verificables”.

En el contexto del marco lógico, los indicadores constituyen el medio para establecer qué condiciones serían las que señalen el logro de los objetivos del proyecto y por ende reduzcan la ambigüedad y la subjetividad en torno al grado de éxito de las intervenciones. En tanto variables, los indicadores presentan dos características esenciales: (a) son características observables de algo; (b) son verificables objetivamente, esto es, son verificables por medios externos al objetivo que pretenden medir. Los indicadores son formas operativas de las variables. Debido a la importancia de que un indicador pueda verificarse en forma objetiva, independientemente de si es directo o indirecto, junto a la especificación de indicadores se deben seleccionar los medios o fuentes apropiados de verificación. El principio fundamental reza así: *“si un indicador no es verificable por ningún medio, entonces encuéntrese otro indicador”*³⁹. Por tanto, los tres principios que deben guiar la selección de indicadores son los siguientes: a) Los indicadores deben ser significativos y relevantes, b) los datos requeridos para hacer cálculos deben ser factibles de una recopilación oportuna y económica y c) los indicadores y su cálculo deben estar acordes con la capacidad institucional de la entidad ejecutora.

Principio 4. Construir la línea base:

La línea base es la primera medición de todos los indicadores seleccionados para medir los objetivos de un programa. Se realiza al inicio del proyecto para contar con una “base” que permita luego cuantificar los cambios netos producidos por la intervención.

Debido a diversas circunstancias, en los procesos de evaluación de programas pueden presentarse tres situaciones tipo:

1. El proyecto tiene línea de base, formulada en la etapa de pre-inversión. Es factible por ende realizar una evaluación de impacto.
2. Durante la preparación del proyecto, no se elaboró la línea de base. Es posible hacer retroactivamente la línea base y por lo tanto es también factible realizar una evaluación de impacto utilizando metodología de pareo (matching).
3. El proyecto no tiene línea de base. No es posible elaborar una línea base retroactiva. La solución sería transformar las observaciones cualitativas en indicadores de resultados.

La construcción de la línea base es un elemento crucial en el diseño de las evaluaciones de impacto, pues sólo de ese modo se podrá apreciar objetivamente el impacto del proyecto, comparar las situaciones antes y después del proyecto, así como las situaciones con y sin proyecto.

³⁹ El marco lógico incentiva al evaluador a definir clara y explícitamente que es lo que indicará que la ejecución del proyecto ha sido exitosa. De este modo se evitan las interpretaciones subjetivas de los logros del proyecto, y éste gana en cuanto a evaluabilidad; es decir, capacidad para ser evaluado objetivamente durante y después de su ejecución.

Principio 5. Construir el escenario contrafactual:

La evaluación del impacto tiene por objeto determinar si un proyecto ha producido los efectos deseados en la población beneficiaria y si estos efectos son atribuibles a la intervención (Baker, 2000). Las evaluaciones de impacto, por lo general, también permiten examinar consecuencias no previstas en los beneficiarios, ya sean positivas o negativas. Tal como se mencionó en subtítulos anteriores, algunas de las preguntas que se abordan en la evaluación de impacto incluyen las siguientes: ¿Cómo afectó el proyecto a los beneficiarios?, ¿Algún mejoramiento fue el resultado directo del proyecto o se habría producido de todas formas?, ¿Se podría modificar el diseño del proyecto para mejorar sus repercusiones? Sin embargo, estas preguntas no se pueden medir simplemente a través de los indicadores establecidos en el proyecto. Puede haber otros factores o sucesos que estén correlacionados con los resultados, sin ser causados por el proyecto, por eso, para asegurar el rigor metodológico, la evaluación de impacto debe estimar el escenario contrafactual o escenario simulado alternativo; es decir, lo que habría ocurrido si el proyecto nunca se hubiera realizado. De lo dicho, determinar el escenario contrafactual es esencial para el diseño de la evaluación y, hasta ahora, sólo existe dos metodologías apropiadas para ello: los diseños experimentales (aleatorios) y los diseños cuasiexperimentales (no aleatorios).

Estos cinco principios metodológicos tienen una relación funcional sistémica, donde siempre la hipótesis de desarrollo (Principio 1) se centra en los resultados del programa y no en sus procesos (Principio 2). Esta hipótesis se verifica utilizando indicadores objetivos (Principio 3) y una línea base antes y después (Principio 4) en conjunto con un escenario contrafactual (Principio 5). Esta relación se gráfica en la siguiente figura.

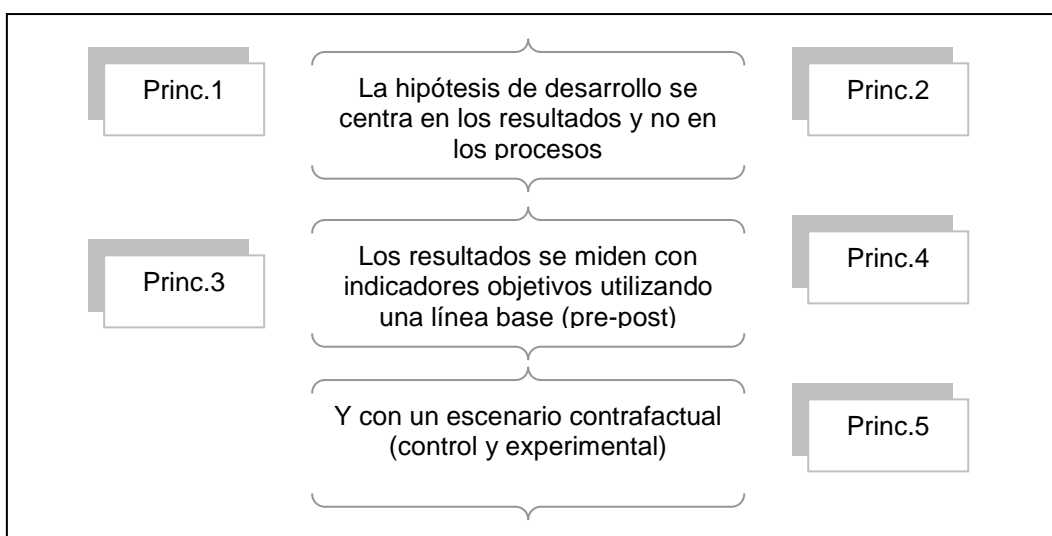


Figura N° 2.4. Relación entre principios metodológicos de EI (Fuente: Elaboración propia).

2.3.6. Los indicadores de evaluación de impacto

Tal como se mencionó en el principio metodológico 3, los indicadores constituyen herramientas indispensables de información relevante y oportuna para la toma de decisiones, el seguimiento de los compromisos y la rendición de cuentas. Asimismo, son elementales para evaluar, dar seguimiento y predecir tendencias de la situación de un país, un Estado o una región, valorar el desempeño institucional y llevar a cabo la comparación del desarrollo económico y social (Sandoval & Richard, 2003).

En sentido estricto, Mondragón Pérez (2002) señala que no existe una **definición** oficial por parte de algún organismo nacional o internacional, sólo algunas referencias que los describen como *“herramientas para clarificar y definir, de forma más precisa, objetivos e impactos (...) son medidas verificables de cambio o resultado (...) diseñadas para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso (...) con respecto a metas establecidas, facilitan el reparto de insumos, produciendo (...) productos y alcanzando objetivos”* (p.52).

Otra definición usada (en el sentido social) por diferentes organismos y autores, es que *“los indicadores sociales (...) son estadísticas, serie estadística o cualquier forma de indicación que nos facilita estudiar dónde estamos y hacia dónde nos dirigimos con respecto a determinados objetivos y metas, así como evaluar programas específicos y determinar su impacto”* (Horn, 1993, p. 147). La construcción de indicadores requiere un marco legal, programático y normativo (nacional o internacional) que establece las necesidades de información para medir o analizar la situación de la economía, la sociedad, la población o el medio ambiente, respecto a determinados valores o metas perseguidos.

Sobre este punto, es importante mencionar que la ONU se dio a la tarea de fomentar la generación de estadísticas e indicadores⁴⁰ por lo que poco a poco, cada nación ha ido conformando una infraestructura estadística que le ha permitido la descripción, comparación y evaluación de su situación con respecto al resto del mundo. Las áreas en las que se han logrado importantes avances en la identificación de indicadores son: demografía, economía y pobreza, educación, salud y nutrición, trabajo y empleo; algunos temas están aún en proceso (como los de medio ambiente) y otros han presentado problemas (por ejemplo, los derechos humanos y el buen gobierno).

Los indicadores facilitan la vigilancia y la evaluación sistemática, por ello los indicadores son parte del diseño del programa desde sus fases de iniciación y planificación (OMS, 2003).

⁴⁰ La Comisión de Estadística tiene un conjunto mínimo de 15 indicadores de cada país para el seguimiento de las cumbres mundiales; el Comité de Asistencia para el Desarrollo de la OCDE tiene trabajados 21 indicadores de las metas de desarrollo internacional; la Comisión de Desarrollo Sustentable de las Naciones Unidas maneja 57 indicadores, y la ONU posee un conjunto básico de indicadores sociales que dan cuenta del acceso de la población a los servicios. La OCDE desarrolla y fomenta una segunda generación de indicadores que muestran no sólo los resultados de la implantación de políticas, sino también los costos, insumos y procesos que participan en la efectividad de las políticas públicas.

Por ello, dentro de las **características** que un indicador debe cumplir se encuentran las siguientes:

- Estar inscrito en un marco teórico, que le permita asociarse firmemente con el evento al que el investigador pretende dar forma.
- Ser específicos, es decir, estar vinculados con los fenómenos económicos, sociales, culturales o de otra naturaleza sobre los que se pretende actuar; por lo anterior, se debe contar con objetivos y metas claras, para poder evaluar qué tan cerca o lejos se está de los mismos y proceder a la toma de decisiones pertinentes.
- Ser explícitos, de tal forma que su nombre sea suficiente para entender si se trata de un valor absoluto o relativo, de una tasa, una razón, un índice, etc., así como a qué grupo de población, sector económico o producto se refieren y si la información es global o está desagregada por sexo, edad, años o región geográfica.
- Estar disponibles para varios años, con el fin de que se pueda observar el comportamiento del fenómeno a través del tiempo, así como para diferentes regiones y/o unidades administrativas.
- La comparabilidad es un insumo que permite fomentar el desarrollo social o económico de acuerdo con lo que tenemos respecto a los demás o a otros momentos; de hecho, nociones socioeconómicas como producción, pobreza y trabajo sólo son comprensibles en términos relativos.
- Deben ser relevantes y oportunos para la aplicación de políticas, describiendo la situación prevaleciente en los diferentes sectores de gobierno, permitiendo establecer metas y convertirlas en acciones. Los indicadores no son exclusivos de una acción específica; uno puede servir para estimar el impacto de dos o más hechos o políticas, o viceversa. Para tener una evaluación completa de un sector o un sistema, se requiere de un conjunto de indicadores que mida el desempeño de las distintas dependencias o sectores y proporcione información acerca de la manera como éstos trabajan conjuntamente para producir un efecto global.
- Ser claro y de fácil comprensión para los miembros de la comunidad, de forma que no haya duda o confusión acerca de su significado, y debe ser aceptado, como expresión del fenómeno que se desea medir. Para cada indicador debe existir una definición, fórmula de cálculo y meta, datos necesarios para su mejor entendimiento y socialización. Por lo anterior, es importante que el indicador sea confiable, exacto en cuanto a su metodología de cálculo y consistente.

- Que la recolección de la información permita construir el mismo indicador de la misma manera y bajo condiciones similares, año tras año, de modo que las comparaciones sean válidas.
- Técnicamente debe ser sólido, es decir, válido, confiable y comparable, así como factible, en términos de que su medición tenga un costo razonable.
- Ser sensible a cambios en el fenómeno, tanto para mejorar como para empeorar. Ejemplos de cómo un indicador puede desviarse de medir la efectividad de una política, se encuentran de manera frecuente en el tema de medio ambiente, como cuando se tuvo que descartar el de superficie afectada por incendios para medir la efectividad de las acciones de gobierno, debido a que su comportamiento también depende de factores naturales no controlables.

Si bien la selección de indicadores no debe depender de la disponibilidad de información proveniente de encuestas, censos y/o registros administrativos, sino de los objetivos fijados en los programas y proyectos de gobierno (dentro de los cuales se inscriben), es importante considerar el costo-beneficio del tiempo y los recursos necesarios para su construcción, por lo que, de preferencia, debe ser medibles a partir del acervo de datos disponible.

Los indicadores no están exentos de **limitaciones** para su conformación; entre la problemática que se puede presentar en su identificación e integración están las siguientes:

- Para la selección de los más adecuados para cada objetivo, existen diferentes actores, como los sectores público, privado, social o académico que no comparten las mismas necesidades de información ni persiguen las mismas metas, lo cual dificulta lograr unanimidad en su definición.
- El carácter cuantitativo hace que se generen indicadores sólo de aquello que puede ser medido en cantidad.
- El uso de indicadores presupone que las metas de instituciones y dependencias públicas, así como sus niveles de logro están disponibles e identificables en planes y programas de gobierno y que no hay contradicciones entre los mismos. Depender de objetivos que son cambiantes, por lo menos en cada nueva administración, no sólo modifica el tipo de indicadores que deben ser empleados, sino también la disponibilidad de datos para conformarlos, lo que implica un ajuste permanente de las fuentes tradicionales de información (Mondragón, 2002).

Por otro lado, existen diversas **clasificaciones tipológicas** de los indicadores de evaluación de impacto. Entre las principales, se menciona las siguientes:

a. Dependiendo del campo de conocimiento:

Según el área de conocimiento, se habla de indicadores económicos, sociales, ambientales, etcétera, cuyo fin es ser un insumo para evaluar el

acercamiento a las metas de bienestar económico, social y de conservación del medio ambiente, respectivamente. La diferencia entre ellos está en las unidades de medida que utilizan: mientras que los indicadores económicos lo hacen en unidades monetarias y/o productos, los sociales lo hacen en relación a personas; y los ambientales, principalmente, en recursos naturales (Mondragón, 2002).

b. Indicadores objetivos y subjetivos:

Si se considera la forma como se obtiene la información para construirlos, se puede diferenciar entre los indicadores objetivos y subjetivos:

- Para Mondragón Pérez (2002), los primeros se basan en evidencias externas independientes del informante (Ej. Nivel educativo), suponiendo que los métodos de captación, procesamiento y divulgación de la información son objetivos. Los segundos son juicios y reflejan percepciones y opiniones de la población con respecto a su situación, a la de la sociedad o al país.
- Para Nussbaum y Sen (1999), las palabras “objetivo” y “subjetivo” no son del todo claras, lo objetivo se refiere a informes de las condiciones reales y a la conducta evidente, en tanto que lo subjetivo consiste en la medición de actitudes o creencias.

Cuando se usan indicadores objetivos, no se pide a los informantes que evalúen si sus condiciones de vida son buenas o malas, satisfactorias o insatisfactorias. Simplemente se les pide que informen de sus condiciones de vida o de su conducta de conformidad con algunos parámetros dados (Nussbaum & Sen, 1999). Los indicadores objetivos se basan en observaciones externas y se aplican simplemente al conteo de diferentes actividades (Nussbaum & Sen, 1999).

c. Indicadores positivos o negativos:

Si lo que se pretende destacar son avances o rezagos de algún aspecto de la realidad, se habla de indicadores positivos o negativos; por ejemplo, para el tema de educación, se puede hablar de índices de alfabetismo o analfabetismo. También, existen indicadores indeterminados (como la tasa de matrícula en educación superior, de la cual, si bien se espera que aumente y alcance 100%, no necesariamente es positivo pues es imposible que toda la población alcance este nivel de estudios).

d. Indicadores absolutos y relativos:

Cuando la evaluación del indicador depende de un valor determinado (como puede ser un valor máximo o mínimo que se debe cumplir), o de la posición relativa del país o de las poblaciones con respecto a otras, entonces hablamos de indicadores absolutos y relativos. Los primeros dependen de una meta a cubrir (como puede ser un 100% de alfabetismo, asistencia escolar, población ocupada, etc.), mientras que los segundos ubican la posición de una unidad geográfica (como un país, con respecto a otras unidades), un ejemplo es el índice de desarrollo humano.

e. Indicadores que se aplican en la administración pública:

Otros esquemas que se aplican en la administración pública organizan a los indicadores en: de gestión y de resultado, de insumos y productos, y otros más complejos que van desde costos, insumos, procesos, productos y resultados. Esta clasificación se asocia al enfoque del Marco Lógico. Tal como se mencionó anteriormente, bajo el enfoque del marco lógico, los indicadores se clasifican en cuatro grupos básicos:

1. Indicadores de Impacto. Estos incluyen medidas de desempeño para los objetivos del nivel más alto a los cuales apunta un proyecto. Por este motivo, los indicadores de este nivel pueden ir más allá del alcance del proyecto (Ej. Impacto real/Impacto programado, incidencia de abandono infantil, índice de violencia intrafamiliar, % de deserción escolar, rendimiento escolar, desocupación juvenil, prevalencia de drogadicción, incidencia de embarazo precoz, prevalencia de ETS/SIDA).
2. Indicadores de Efecto. Estos indicadores muy a menudo definen el cambio en el comportamiento de los beneficiarios del proyecto o el cambio en la manera en que funcionan las instituciones como resultado del proyecto. En consecuencia, la definición de estos indicadores puede ser difícil (Ej. cantidad de familias que demandan atención, porcentaje de jóvenes que utilizan las herramientas entregadas para buscar trabajo, cantidad de jóvenes participando en campañas de prevención de drogas, variación en volumen de uso de preservativos).
3. Indicadores de Producto. Establecen los marcos de referencia para la evaluación de los resultados del proyecto, ya que corresponde a la institución ejecutora producir los resultados esperados.
4. Indicadores de Proceso. Son los indicadores a nivel de actividades y que generalmente se asocian a los recursos programados para el proyecto. Tomados en conjunto producen una evaluación de desempeño, ya que los costos están directamente relacionados con las actividades. Los recursos necesarios definidos en estos indicadores se emplean para analizar la eficacia o efectividad en función de los costos.

f. Indicadores de contexto:

Aunque no reflejan de forma directa la situación del sector que se quiere evaluar, son parte del ambiente que afecta la situación social, económica o ambiental y pueden modificar el comportamiento de los fenómenos bajo observación. Es común que se considere al producto interno bruto (PIB) *per cápita*, a la tasa de fecundidad y a la de crecimiento de la población, entre otros, como indicadores de contexto.

g. Indicadores simples o sintéticos:

Dependiendo del tipo de medición o procedimiento estadístico que se requiere para su obtención, los indicadores pueden ser clasificados como simples (si se trata de una estadística univariada y/o poco complicada), o sintéticos (si se habla de un agregado que sintetiza la situación global de un sector determinado y que incluye varios componentes del mismo).

En este sentido, es importante aclarar que un indicador no debe ser, necesariamente, una estadística derivada, es decir, un valor absoluto (como el monto de la población), ya que puede constituir un indicador cuando se le inscribe en una serie de tiempo que nos permite evaluar si su evolución se está acercando o alejando de las expectativas esperadas.

2.3.7. Procedimiento general de la evaluación de impacto

Como todo procedimiento de investigación, siempre se requiere un plan. Por eso, emprender una evaluación de impacto sin la planificación adecuada puede generar demasiados costos, ya que en cada etapa del proceso surgirán problemas de implantación y se requerirán recursos adicionales para superarlos. Por ello, es importante contar con un estudio bien diseñado, un equipo comprometido y muy competente y una buena comunicación entre ellos (Kellogg Foundation, 1998).

Lo ideal sería que la evaluación sea incorporada desde la etapa inicial del diseño del programa, pues solo así será posible obtener resultados oportunamente, de modo que las conclusiones se puedan utilizar para ajustar componentes específicos durante el programa (OPS, 2001). Hacerlo así permite garantizar un buen diseño ya que se puede implementar una línea base y determinar grupos controles clave o controlar covariantes importantes (Bedi, Bhatti, Gine, Galasso, Goldstein & Legovini, 2006).

Sin considerar el tamaño, el tipo de programa ni la metodología usada para la evaluación, se deben seguir varios pasos clave que se resumen a continuación (véase la tabla siguiente). La secuencia de estos pasos es decisiva, especialmente para asegurar la recolección de los datos necesarios antes que comience la implementación del programa. La planificación oportuna permite aleatorizar, crear comparaciones pareadas *ex ante*, recolectar datos básicos e identificar encuestas futuras que se podrían usar en un enfoque de correspondencia de puntuación de la propensión (matching).

Tabla 2.14. Procedimiento general de la evaluación de impacto

1.	Determinar si realizar o no una evaluación
2.	Aclarar los objetivos de la evaluación
3.	Examinar la disponibilidad de datos
4.	Diseñar la evaluación
5.	Formar el equipo de evaluación
6.	Si se recopilan datos: <ul style="list-style-type: none">• Diseñar y seleccionar muestras• Elaborar instrumentos de recopilación de datos• Reunir y capacitar personal para trabajo de campo• Realizar pruebas piloto• Recopilar datos• Administrar datos y acceder a ellos
7.	Recopilar datos actuales
8.	Analizar los datos
9.	Redactar los resultados y analizarlos con las autoridades responsables y otras partes interesadas
10.	Incorporar los resultados en el diseño del proyecto.

A continuación se describe cada paso del proceso.

2.3.7.1. Determinar la realización de la evaluación

Como primer paso, hay que determinar si se requiere o no una evaluación del impacto. Como se señaló anteriormente, la evaluación del impacto se diferencia de otras evaluaciones porque se centra en examinar la causalidad. Dada la complejidad de realizar una evaluación de impacto, se deben examinar sus costos-beneficios y considerar si sería más adecuado algún otro método. En todo caso, la información más importante para decidir si se realizará una evaluación de impacto es la referida al apoyo político y financiero (Baker, 2000).

Es más factible realizar la evaluación del impacto cuando el proyecto es innovador, pues supone importantes asignaciones de recursos, comprende intervenciones bien definidas y, además, podrá extenderse basándose en los resultados de la evaluación.

Otra importante consideración es asegurar que el programa que se evaluará se encuentre en una etapa suficientemente avanzada para ser sometido a una evaluación de impacto. Los proyectos piloto y las reformas incipientes con frecuencia tienden a sufrir modificaciones de su contenido y administración. Estos cambios pueden perjudicar la coherencia de la evaluación, especialmente de los diseños experimentales y cuasi-experimentales. Cuando las políticas que se evaluarán aún se están definiendo, es recomendable no usar una evaluación del impacto.

En todo caso, el reto para el evaluador y el cliente consiste en decidir si es posible realizar una evaluación de impacto de calidad con las limitaciones existentes en la práctica, y elegir el diseño más sólido posible en función de una serie concreta de restricciones de presupuesto, tiempo e información. Por ejemplo, ¿A partir de qué punto se considera que la muestra es demasiado pequeña, o su

cobertura demasiado limitada, para que el análisis estadístico sea válido? ¿Cuáles son los criterios para evaluar si los datos secundarios son adecuados para estimar las condiciones iniciales? ¿Y cuándo es posible formular una hipótesis válida a falta de un estudio de referencia? (Bamberger, 2006).

2.3.7.2. Aclarar los objetivos de la evaluación

Determinada la ejecución de la evaluación del impacto, es necesario establecer objetivos claros. Los objetivos claros son esenciales para identificar las necesidades de información, definir los indicadores de resultados y efectos y crear una estrategia de evaluación sólida que proporcione respuestas precisas (Baker, 2000).

Aunque aparentemente el planteamiento del objetivo pareciera ser una de las partes más fáciles del proceso de evaluación, puede ser en extremo difícil. Por ejemplo, los planteamientos que son demasiado amplios son difíciles de evaluar. El planteamiento del objetivo de la evaluación del PROBECAT de México, de que la evaluación busca examinar el “*efecto del programa de capacitación PROBECAT en los resultados del mercado laboral*”, sería más preciso si se limitara al efecto de PROBECAT en las horas trabajadas, los ingresos por hora, el salario mensual y el tiempo para conseguir el primer empleo para diferentes tipos de trabajadores. La evaluación de PROGRESA de México es un buen ejemplo de la creación de un esquema claro y el delineamiento de múltiples objetivos desde el comienzo, con un análisis separado para cada componente y con los objetivos detallados en subcategorías. Esto fue especialmente importante, porque la intervención fue bastante compleja al tener que abordar no sólo el efecto del programa, sino también aspectos de la focalización y secuencia cronológica de las operaciones del programa.

Revisar otros componentes de evaluación, como la eficacia en función de los costos o las evaluaciones del proceso, también pueden ser importantes objetivos de un estudio y pueden complementar la evaluación del impacto. La eficacia en función de los costos puede constituir una preocupación muy especial para las autoridades responsables, cuyas decisiones serán reducir, expandir o reformar la intervención que se está evaluando. En temas relacionados con la entrega de servicios, una evaluación de los procesos podría ser pertinente para evaluar los procedimientos, dinámica, normas y restricciones bajo las cuales se lleva a cabo un determinado programa.

2.3.7.3. Examinar la disponibilidad de los datos

Se pueden emplear muchos tipos de datos para realizar estudios de evaluación del impacto. Éstos pueden incluir una gama que fluctúa desde encuestas hasta entrevistas cualitativas. En muchos casos, la evaluación del impacto aprovechará cierto tipo de datos existentes o se apoyará en alguna encuesta en curso, lo que puede significar un considerable ahorro de costos. Sin embargo, con este enfoque pueden surgir problemas de secuencia cronológica del esfuerzo de

recolección de datos y con la flexibilidad del diseño del cuestionario. En el cuadro siguiente se destacan algunos puntos clave para identificar las fuentes de datos disponibles para la evaluación de impacto.

Tabla N° 2.15. Puntos clave para identificar las fuentes de datos disponibles para la evaluación de impacto (Fuente: Baker, 2000).

-
- Conocer bien el programa. Es arriesgado iniciar una evaluación sin conocer bastante sobre los detalles administrativos e institucionales del programa; esta información en general proviene de la administración del programa.
 - Recopilar información sobre los “hechos estilizados” pertinentes al entorno. Los hechos pertinentes podrían incluir el mapa de pobreza, la forma en que funciona el mercado laboral, las principales divisiones étnicas y otros programas públicos pertinentes.
 - Ser ecléctico con respecto a los datos. Las fuentes pueden abarcar entrevistas informales no estructuradas con participantes en el programa y datos cuantitativos de muestras representativas. Sin embargo, es muy difícil plantear preguntas de un escenario contrafactual en las entrevistas o grupos representativos; se debe intentar preguntar a alguien que actualmente participe en un programa público: “¿Qué estaría haciendo ahora si este programa no existiera?” Puede ser valioso conversar con los participantes del programa, pero es poco probable que ello por sí sólo proporcione una evaluación creíble.
 - Asegurarse de que existen datos sobre los indicadores de resultados y variables explicativas pertinentes. Estas últimas deben abordar la heterogeneidad en los resultados basada en la participación en el programa. Dependiendo, por ejemplo, si se cuenta con cierto nivel de instrucción, los resultados pueden ser diferentes. Podría ser imposible observar el efecto del programa a menos que se controle por esa heterogeneidad.
 - Dependiendo de los métodos usados, también se podrían necesitar datos sobre variables que influyen en la participación, pero no influyen en los resultados dada la participación. Estas variables instrumentales pueden ser valiosas para separar los efectos probablemente causales de los programas no aleatorios.
 - Los datos sobre resultados y otras variables explicativas pertinentes pueden ser cuantitativos o cualitativos. Pero debe ser posible organizar la información en cierto tipo de estructura sistemática de datos. Un ejemplo simple y común es que se tienen valores de diversas variables, incluido uno o más indicadores de resultados para diversas unidades de observación (individuos, hogares, empresas, comunidades).
 - Frecuentemente las variables sobre las que se tienen datos y las unidades de observación que se usan se seleccionan como parte del método de evaluación. Estas elecciones deben estar vinculadas al conocimiento previo sobre el programa (desde luego, sus objetivos, pero también cómo se maneja) y el escenario en que se introdujo.
 - La fuente específica de los datos sobre los resultados y sus determinantes, incluida la participación en el programa, en general proviene de datos de encuestas de algún tipo. La unidad de observación podría ser el hogar, empresa o área geográfica, dependiendo del tipo de programa que se está estudiando.
 - Los datos de las encuestas con frecuencia se pueden complementar con otros datos útiles sobre el programa (como de la base de datos de supervisión del proyecto) o escenario (como de las bases de datos geográficas)
-

Al nivel más básico, se requerirán datos sobre el universo de la población de interés como base para determinar los tamaños de las muestras, crear el marco de muestreo y seleccionar la muestra. Otros tipos de datos que pueden estar disponibles en un país determinado y que se pueden emplear para distintas evaluaciones de impacto incluyen (Valadez & Bamberger, 1994): encuestas ingresos y gastos familiares, encuestas sobre las condiciones de vida (ENCOVI), encuestas nacionales de Hogares (ENAH0); encuestas nacionales demográficas y salud (ENDES); encuestas del mercado laboral, registros de cooperativas, asociaciones de crédito y otras instituciones financieras, registros escolares sobre asistencia, repetición y rendimiento en exámenes; registros de salud pública sobre mortalidad infantil, incidencia de diferentes enfermedades infecciosas, número de

mujeres que busca orientación sobre anticoncepción o consumo de preservativos, encuestas especializadas aplicadas por universidades, organizaciones no gubernamentales (ONGs) y grupos asesores, vigilancia de datos de administradores de programas y estudios de casos de proyectos (Baker, 2000).

2.3.7.4. Diseñar la evaluación

La elección de metodologías dependerá de las preguntas, secuencia cronológica, restricciones presupuestarias y capacidad de implementación de la evaluación. Generalmente, se contrapesan los pros y contras de los distintos tipos de diseño, para determinar las metodologías que son más adecuadas y las técnicas cuantitativas y cualitativas que se pueden integrar para complementarse entre sí (Baker, 2000).

Incluso después que se ha determinado el diseño de la evaluación y se ha incorporado al proyecto, los evaluadores deben estar preparados para ser flexibles y modificar el diseño a medida que se implementa el proyecto. Al definir el diseño, también es importante determinar cómo la evaluación del impacto se ajustará a la estrategia de control y evaluación más amplia aplicada a un proyecto.

Algunos diseños demoran más en implementarse que otros. La asignación aleatoria y los métodos antes y después tardan más que los enfoques de comparación pareada *ex post*. Al usar enfoques antes y después que utilizan evaluaciones básicas (de referencia) y seguimiento, se debe disponer de suficiente tiempo para que el último miembro del grupo de tratamiento reciba la intervención y luego por lo general de más tiempo para que se materialicen y se observen los efectos posteriores al programa. Grossman (1994) sugiere que en general hay que dejar transcurrir entre 12 y 18 meses después de incorporar la muestra en la intervención para examinar los efectos. En los proyectos del Banco Mundial con niveles de referencia, la espera para que se implemente la intervención y que se materialicen los resultados puede tomar años (Baker, 2000).

Una consideración final en el diseño de la evaluación es la capacidad de implantación del equipo de evaluación. Los aspectos de la implantación pueden ser bastante difíciles, especialmente en los países en desarrollo, donde hay poca experiencia en las evaluaciones de programas e investigaciones aplicadas. La composición del equipo de evaluación es muy importante, al igual que la experiencia de los miembros de equipo en diferentes tipos de metodologías y su capacidad relativa a otras actividades de evaluación. Esto se aplica en particular cuando se trabaja con organismos del sector público con múltiples responsabilidades y limitado personal. Es importante conocer la carga de trabajo, para evaluar no sólo la forma en que esto afectará la calidad de la evaluación que se está realizando, sino también el costo de oportunidad de la evaluación con respecto a otros esfuerzos de los cuales está encargada el equipo. Alternativamente, puede ser preferible contratar una empresa privada para que realice la evaluación y que se dedique exclusivamente a ello (Baker, 2000).

Por otro lado, cuando se realizan evaluaciones de impacto con limitaciones de presupuesto, tiempo o información, aumenta la dificultad de manejar las cuatro series de amenazas a la calidad del diseño y la validez de las conclusiones. Aunque

esas cuatro series de amenazas a la validez de las conclusiones de la evaluación pueden afectar a todas las evaluaciones, son más difíciles de gestionar en las situaciones reales. Estas son (Bamberger, 2006):

- **Amenazas a la calidad general del diseño y la ejecución de la evaluación:** Las limitaciones de recursos pueden restringir la atención prestada al diseño de la evaluación, al desarrollo y a la verificación de instrumentos, así como a las consultas a clientes; asimismo, pueden existir presiones para recortar los costos de la recopilación de datos mediante la contratación de entrevistadores más económicos, o reduciendo la formación y la supervisión de los entrevistadores. Por otra parte, puede ser más difícil utilizar enfoques de métodos mixtos y triangulación para controlar la calidad y comprobar plenamente la idoneidad de las fuentes de datos secundarios.
- **Amenazas al análisis estadístico:** Las limitaciones dificultan la toma de medidas para fortalecer el marco muestral o abordar los sesgos de muestreo. Puede haber, asimismo, presiones para reducir el número de puntos de recopilación de datos (por ejemplo, eliminar datos de referencia o grupos de comparación). También puede haber presiones para reducir el tamaño de la muestra, lo que disminuye la potencia de la prueba estadística y limita la posibilidad de realizar un análisis desagregado.
- **Coherencia teórica e idoneidad del marco hipotético:** Las limitaciones dificultan la realización de estudios exploratorios, la consulta a clientes y los seminarios necesarios para elaborar una teoría sobre el programa que explique cómo se espera que éste alcance sus objetivos y cómo afectan las variables contextuales y el proceso de ejecución del proyecto a la magnitud y la distribución de los impactos. Las limitaciones debilitan también el marco hipotético al eliminar puntos de recopilación de información o reducir el acceso a datos secundarios que pueden fortalecer el grupo de comparación.
- **Posibilidad de generalizar los resultados:** Cuando la evaluación no puede manejar los sesgos muestrales o analizar factores contextuales influyentes, aumenta el riesgo de obtener conclusiones erróneas sobre la posibilidad de repetir el proyecto.

En la tabla siguiente, desarrollada por Bamberger (2006), se describen los problemas comunes de estas cuatro categorías y se especifican las limitaciones que suelen contribuir a cada problema:

Tabla N° 2.16. Repercusiones de las limitaciones de presupuesto, tiempo e información en la calidad de una evaluación de impacto (Basado en Bamberger, 2006, con modificaciones).

Problemas (amenazas a la validez) causadas por las distintas limitaciones	Limitaciones que contribuyen a cada problema		
	Presupuesto	Tiempo	Información
A. Problemas que afectan a la calidad general del diseño de la evaluación (amenazas a la validez interna)			
Atención insuficiente a la planificación, la consulta a los clientes y el desarrollo de la relación con consultores locales	X	X	
Atención insuficiente al desarrollo y a la verificación de instrumentos	X	X	
Falta de tiempo para el seguimiento de los resultados de la evaluación	X	X	

Problemas (amenazas a la validez) causadas por las distintas limitaciones	Limitaciones que contribuyen a cada problema		
	Presupuesto	Tiempo	Información
Exclusión de grupos difíciles de alcanzar y sobre los cuales es difícil obtener información	X	X	X
Menor aplicación de enfoques de métodos mixtos, por lo que no pueden utilizarse comprobaciones de coherencia por triangulación	X	X	
Presión para encontrar entrevistadores más baratos y menos recursos para la formación y la supervisión	X	X	
Más dependencia de métodos cualitativos rápidos	X	X	
Mayor dificultad de comprobar la idoneidad de los datos secundarios	X	X	X
B. Problemas que afectan al diseño de la muestra y al análisis estadístico (amenazas a la validez estadística)			
Menor posibilidad de aplicar enfoques de métodos mixtos	X	X	X
Menos recursos para mejorar la calidad del marco muestral	X	X	X
Mayor dificultad para hacer frente a los sesgos muestrales y mejorar las correspondencias	X	X	X
Menor calidad de la aplicación de la muestra	X	X	X
Menor tamaño de la muestra — riesgo de falsos negativos	X		X
Presiones para eliminar la recopilación de datos sobre el grupo de tratamiento o el grupo de control, o el grupo de comparación posterior a la intervención.	X		X
Menor análisis desagregado	X	X	X
C. Problemas que afectan a la coherencia de la teoría y la validez del marco hipotético (amenazas a la validez de constructo)			
Menos tiempo y recursos para elaborar un modelo teórico para el proyecto, de modo que los conceptos e indicadores principales pueden estar menos definidos y tal vez no se puedan establecer las principales hipótesis o se establezcan erróneamente	X	X	X
Menor uso de enfoques basados en métodos múltiples y de la triangulación	X	X	X
Grupo de control/comparación más débil (más pequeño o con una correspondencia peor)	X		X
Datos de referencia poco fiables o inexistentes	X	X	X
D. Problemas que afectan a la posibilidad de generalizar los resultados y las recomendaciones sobre la posibilidad de repetir el proyecto en otros entornos o con otros grupos (amenazas a la validez externa)			
Falta de atención a los sesgos muestrales	X		X
Análisis débil de factores contextuales que contribuyen al éxito o al fracaso en lugares concretos	X	X	X

2.3.7.5. Formar el equipo de evaluación

La calidad y utilidad final de la evaluación del impacto se puede mejorar si desde un principio se identifica a los miembros del equipo, acuerda sus funciones y responsabilidades y se establecen mecanismos de comunicación durante los momentos clave de la evaluación.

El equipo básico se compone del administrador o director de la evaluación, los analistas (científicos sociales) y, para los diseños de la evaluación que implican la recolección de nuevos datos, un experto en muestreo, un diseñador de encuestas, un administrador del trabajo de campo y el respectivo equipo, y los administradores de datos (Grosh & Muñoz, 1996). Según el tamaño, alcance y diseño del estudio, algunas de estas responsabilidades serán compartidas o bien se agregarán otras necesidades de personal a este equipo básico.

Al formar el equipo de evaluación, también hay algunas decisiones importantes que el director de la evaluación debe tomar acerca de la capacidad

local y las disposiciones institucionales adecuadas para garantizar la imparcialidad y la calidad de los resultados de la evaluación:

1. La primera de ellas es si existe capacidad local para implementar la evaluación o parte de ella y qué tipo de supervisión y asistencia exterior se necesitará. Actualmente la tendencia es que diversos organismos internacionales (Ej. Banco Mundial, BID, etc.) supervisen las evaluaciones hechas por el personal local.
2. La segunda decisión es si trabajar con una empresa privada o con un organismo público. Las empresas privadas pueden ser más confiables con la entrega de resultados en forma oportuna, pero se pierde la creación de capacidades en el sector público. Además, las empresas privadas, comprensiblemente, a menudo están menos dispuestas a incorporar elementos a la evaluación que harán más costoso el esfuerzo.
3. Y la tercera es el grado de separación institucional que habrá que establecer entre los proveedores de la evaluación y los usuarios de ésta. Muchos especialistas (Ej. Baker, 2000; Abdala, 2000) recomiendan que el equipo que evalúe sea distinto al equipo que aplica el programa, pues así se garantiza la independencia y objetividad de los resultados; sin embargo, esta afirmación no es totalmente adecuada. Debido a la fuerte carga política y social de la evaluación de impacto, es recomendable también que participen agentes beneficiarios y ejecutores del programa por cuanto esa participación le otorga legitimidad.

2.3.7.6. Evaluación de datos

Tener datos adecuados, fiables y de alta calidad es esencial para la realización de una evaluación de impacto que sea válida. Por ello, es importante evaluar los datos que existen, antes de iniciar cualquier esfuerzo de recolección (Bamberger, 2006). En la tabla siguiente se relacionan las metodologías de evaluación básicas con los datos necesarios para cada una.

Tabla N° 2.17. Métodos de evaluación y necesidades de datos (Adaptado de Ezemenari, Rudqvist & Subbarao, 1999 y Baker, 2000).

Método	Necesidad de datos		Uso de enfoque cualitativo
	Mínima	Ideal	
Controles experimentales o aleatorizados	Corte transversal de un único proyecto con y sin beneficiarios	Encuestas básicas y de seguimiento de los beneficiarios y no beneficiarios. Permite controlar sucesos contemporáneos, además de proporcionar control para medir el efecto (permitiendo una estimación de diferencia en las diferencias)	Guiar el diseño del instrumento de la encuesta, muestreo. Identificar los indicadores. Recopilar y registrar datos con la ayuda de datos textuales, entrevistas informales o semi estructuradas, reuniones de los grupos representativos de la comunidad, observación directa, métodos participativos, fotografía, triangulación y análisis

Método	Necesidad de datos		Uso de enfoque cualitativo de datos.
	Mínima	Ideal	
Diseños experimentales no a) Controles o pareo creados	Encuesta extensa, censo, presupuesto nacional o encuesta de condiciones de vida en que se muestra a los beneficiarios.	Encuesta extensa y encuesta domiciliaria más reducida basada en el proyecto, ambas con dos momentos en el tiempo para controlar por los sucesos contemporáneos.	
b) Comparaciones reflexivas y doble diferencia	Básica y de seguimiento de los beneficiarios	Serie cronológica o panel de los beneficiarios comparables	
c) Control estadístico o variable instrumental	Datos transversales representativos de la población beneficiaria con las correspondientes variables instrumentales.	Corte transversal y series cronológicas representativas de la población beneficiaria, con las variables instrumentales correspondientes.	

Cuando la evaluación no se encarga hasta que el proyecto está a punto de finalizar o ha terminado (lamentablemente es la norma en muchos países), la información puede obtenerse por uno de los cuatro medios siguientes (White, 2006; Bamberger, 2006): a) mediante una encuesta específica para el proyecto; b) agregando un módulo especial a una encuesta en curso; c) mediante una encuesta sincronizada en la que se entreviste a la población beneficiaria, pero tomando el grupo de control de otra encuesta (encuesta nacional de hogares, etcétera), o d) basando la evaluación en datos secundarios, como la información sobre el proyecto y los grupos potenciales de comparación.

Es importante mencionar que casi todas las evaluaciones de impacto se basan en datos secundarios, incluso cuando se realizan encuestas, y en muchos casos los datos secundarios son la fuente principal o única de información. Por consiguiente, en la mayoría de las evaluaciones la cuestión no es si se debe utilizar la información secundaria, sino cómo garantizar la calidad y la idoneidad de una evaluación determinada (Bamberger, 2006).

Los datos secundarios pueden ser una forma útil de reducir costos y ahorrar tiempo, y cuando se emplean diseños de grupos de beneficiarios y de comparación después de la intervención, suelen ser la única manera de reconstruir condiciones de referencia del grupo de beneficiarios y del grupo de comparación antes del inicio del proyecto. En este diseño, y en casi todos los diseños, pueden utilizarse para fortalecer la estimación hipotética de cuál habría sido la situación de la población del proyecto si éste no se hubiera llevado a cabo.

Algunos de los tipos más habituales de datos secundarios son los siguientes (Bamberger, 2006):

- Los censos nacionales;

- Las encuestas generales de hogares como el Estudio de Medición de los Niveles de Vida;
- Las encuestas especializadas, como las encuestas demográficas y de salud;
- Los datos administrativos recopilados por departamentos ministeriales y otros organismos públicos (matrícula escolar, uso de centros de salud, precios de mercado de la producción agrícola, etcétera);
- Los estudios realizados por organismos donantes, organizaciones no gubernamentales y universidades;
- Los datos administrativos del organismo o ministerio encargado del proyecto;
- Los medios de comunicación (prensa, televisión, documentales, etcétera). Estos pueden ser útiles, entre otras cosas, para comprender el contexto económico y político local de cada lugar donde se ejecute el proyecto.

Otra importante aplicación de los datos secundarios es el metaanálisis, donde los impactos de proyectos o intervenciones comparables en este o en otros países proporcionan estimaciones de la magnitud de los efectos que cabe esperar de un proyecto bien diseñado. El metaanálisis puede resultar especialmente útil para estimar el tamaño que deben tener las muestras del grupo de beneficiarios y del grupo de comparación, ya que (si no se modifican los demás factores) cuanto menor sea el efecto previsto, menor será la potencia de la prueba estadística y mayor la muestra necesaria para detectar los impactos del proyecto, si los hubiere. Por ejemplo, en el Proyecto de nutrición integrada de Bangladesh se realizó un metaanálisis internacional para estimar el número de muertes por 1,000 nacidos vivos que sería razonable esperar salvar mediante la contratación de parteras tradicionales. Se comprobó que el intervalo era de cinco a siete muertes evitadas por 1,000 nacidos vivos. Este dato se empleó para confirmar que la meta de evitar siete fallecimientos por 1,000 nacidos vivos era realista (White, 2006). Podría haberse utilizado ese mismo análisis para calcular la magnitud del efecto esperado al estimar el tamaño de la muestra necesario para una evaluación de impacto.

Aunque los datos secundarios son muy valiosos para la evaluación, tienen limitaciones. Es probable que la información se haya recopilado para otros fines, por lo que las fuentes de información deben evaluarse cuidadosamente antes de su utilización. Existen posibles problemas que deben examinarse antes de hacer uso de datos secundarios se encuentran los siguientes (Bamberger, Rugh y Mabry, 2006):

- Hay un desfase cronológico entre el inicio el proyecto (que es cuando se necesitan los datos de referencia) y la recopilación o transmisión de los datos secundarios.
- ¿Qué semejanza hay entre la muestra y la población destinataria? Por ejemplo, ¿se incluyen en la encuesta tanto las escuelas públicas como las privadas? ¿Se tiene en cuenta el empleo del sector formal y del sector informal? ¿Se consideran hombres y mujeres, así como otros grupos de interés, como los ancianos?

- ¿Se obtuvo información sobre todas las variables importantes y los indicadores de resultados del proyecto? ¿Son los datos adecuados para la evaluación? A menudo, han de utilizarse uno o dos indicadores indirectos simples para medir indicadores complejos de resultados (por ejemplo, el uso de indicadores de servicios sanitarios prestados como indicador indirecto de los impactos en la salud, o el uso de los volúmenes y tipos de vehículos, y el número de empresas nuevas, como indicador del impacto de los caminos rurales).

De lo dicho, es importante evaluar la calidad y la exhaustividad de la información. A veces, cuando la información se recopila con fines administrativos, no hay controles de calidad y la información es incompleta, inexacta o sesgada (por ejemplo, cuando las escuelas tienen un incentivo para falsificar a su favor las tasas de matrícula o las puntuaciones de los exámenes, o la policía para no comunicar todos los delitos). Esto reviste especial importancia en el caso de las evaluaciones de impacto, ya que el incentivo de comunicar datos ficticios es mayor si la unidad que presta el servicio sabe que esos datos se utilizarán para una evaluación (Bamberger, 2006).

2.3.7.7. Decisión sobre lo que se medirá

Al planificar la evaluación, generalmente se decide qué se medirá, estableciendo los principales indicadores de resultados, efectos e impactos. En este caso, es común que se defina una jerarquía de indicadores que fluctúa desde indicadores de impacto a corto plazo hasta indicadores a largo plazo. Así se garantiza que aun cuando no se captan inicialmente los efectos finales, se podrán evaluar los resultados del programa.

También se podría considerar alguna información sobre características de la población beneficiaria no relacionadas estrictamente con la evaluación del impacto, pero de interés para el análisis, como su nivel de pobreza o su opinión acerca del programa. Además, el evaluador también podría incluir medidas de costos para realizar análisis de eficacia en función de los costos u otras evaluaciones complementarias que no tenga relación estricta con la evaluación del impacto (Baker, 2000).

El tipo de diseño seleccionado para la evaluación del impacto también influirá en las necesidades de los datos. Éstas serán específicas para la metodología, población de interés, medidas de los efectos y otros elementos de la evaluación. Por ejemplo, si se usa un enfoque de “variable instrumental” (uno de los tipos de estrategias de comparación pareada), se deberá identificar e incluir en la recolección de datos las variables que servirán de instrumento para separar la participación en el programa de las medidas de resultados. Esto es lo que se hizo en la evaluación de impacto del Fondo de Inversión Social de Bolivia, donde se usó el conocimiento del fondo social y la presencia de ONG’s como variables instrumentales para evaluar el impacto de las intervenciones del fondo social (Baker, 2000).

2.3.7.8. Muestreo

El muestreo es una necesidad cuando se trabaja con poblaciones grandes y cuando los recursos asignados a la evaluación son limitados. Es recomendable que el especialista en muestreo sea incorporado al proceso de evaluación desde las primeras etapas. De esa forma, se revisa la información disponible necesaria para seleccionar la muestra y determinar si se requerirá algún trabajo de enumeración (marco poblacional), o algunos ajustes especiales para controlar sesgos.

Como con otras partes del trabajo de evaluación, es importante que exista coordinación entre el especialista en muestreo y el equipo de evaluación. Esto es decisivo al realizar comparaciones pareadas (*matching*), porque aquí el diseño del muestreo es la base del "pareo" y de la creación del escenario contrafactual.

Existen muchas relaciones inversas de compensación entre costos y exactitud del muestreo que deben aclararse al elaborar el marco muestral. Por ejemplo, realizar una muestra en dos o tres etapas reducirá los costos del muestreo y del trabajo en terreno, pero aumentarán los errores de muestreo y, por lo tanto, la precisión de las estimaciones.

Una vez que el equipo de evaluación ha determinado las variables de resultado y la población de interés, un primer paso para el especialista en muestreo sería determinar los cálculos de potencia estadística (Valadez & Bamberger, 1994). Puesto que se pueden realizar los cálculos de potencia usando sólo una medida de resultados y a menudo las evaluaciones consideran varias, se deberán tomar algunas decisiones estratégicas con respecto al indicador de resultado que se usará al diseñar la muestra.

Después de elaborar la estrategia y el marco muestral, el especialista en muestreo también debe participar en la selección de la muestra para el trabajo de campo y la prueba piloto. Una vez finalizado el trabajo de campo, el especialista en muestreo debe ayudar a determinar las ponderaciones de muestreo, para calcular los factores de expansión y corregir los errores y la falta de respuestas (Grosh & Muñoz, 1996; Kish, 1965). En este caso, el especialista puede ajustar los sesgos con métodos de postmuestreo como los *Bootstrap* o *Jackknife* (Davison & Hinckley, 1997; Efron, 1979, 1981, 1983; Efron & Tibshirani, 1993).

Por otro lado, dado que la recopilación de datos puede representar muchas veces más de la mitad del presupuesto de la evaluación (Baker, 2000), la reducción del tamaño de la muestra puede redundar en un ahorro considerable. Sin embargo, hay que tener cuidado, ya que cuando la muestra es más pequeña, disminuye la precisión estadística de las estimaciones y el nivel de desagregación del análisis. Entre los principales factores determinantes del tamaño de la muestra se encuentran los siguientes: el tamaño del efecto de tratamiento promedio estimado, la potencia deseada de la prueba estadística, la media y la varianza de las variables subyacentes, el nivel requerido de precisión estadística, si se usa o no un grupo de comparación, los tipos de análisis desagregado y si se precisan una o dos pruebas estadísticas unilaterales. Teniendo en cuenta estas consideraciones, se dispone de las siguientes opciones para reducir el número de entrevistas y ahorrar costos (Bamberger, 2006):

- Aceptar un nivel más bajo de precisión estadística (por ejemplo, un intervalo de confianza del 90% en lugar del 95%) o de potencia

estadística de la prueba (por ejemplo, un riesgo del 20% en lugar del 10% de rechazar un impacto real del proyecto). Por supuesto, esto aumenta la posibilidad de juzgar erróneamente si un proyecto ha tenido un impacto o no.

- Reducir el nivel de desagregación estadística del análisis (por ejemplo, obtener resultados sólo para la población beneficiaria total en lugar de comparar impactos en grupos diferentes o la efectividad de componentes diferentes del proyecto).
- Cuanto mayor sea la magnitud del efecto esperado, menor será la muestra necesaria para encontrar un impacto estadísticamente significativo.

Estos criterios se han aplicado en evaluaciones de impacto reales y se pueden combinar para abaratar costos sin sacrificar la validez del mismo.

2.3.7.9. Instrumentos de medición

Los instrumentos de medición pueden ser cuestionarios, test sociométricos, escalas de satisfacción, listas de verificación, fichas de recolección documental, guías de entrevistas, entre otros. Es una norma aceptada académicamente que los instrumentos de recolección de datos sean fiables y válidos. En un primer momento, es recomendable que en la elaboración del instrumento participe la mayor cantidad de agentes de evaluación (administrador, personal local, analistas que puedan aportar con conocimientos del país y del programa, etc.) (Grosh & Muñoz, 1996). Posteriormente, mediante el estudio piloto, se determina la validez inicial y la fiabilidad del instrumento.

A continuación se presenta una tabla resumen de los principales instrumentos empleados para la recolección de datos en los estudios de impacto.

Tabla N° 2.18. Algunas técnicas de recolección de datos para la evaluación de impacto (Fuente: Baker, 2000).

Técnica	Definición y uso	Fortalezas	Debilidades
Estudios de casos	Recopilación de información que genere un recuento que puede ser descriptivo o participativo y puede servir para responder a las preguntas: cómo y por qué.	Puede abordar una variedad completa de evidencias de documentos, entrevistas y observación. Pueden agregar poder explicativo cuando se centran en instituciones, procesos, programas, decisiones y sucesos.	Los buenos estudios de casos son difíciles de realizar. Requieren conocimientos especializados de investigación y redacción para que sean rigurosos. Los resultados no se pueden generalizar a toda la población. Tardan mucho tiempo. Son difíciles de repetir.

Técnica	Definición y uso	Fortalezas	Debilidades
Grupos representativos o grupos focales	Conversaciones focalizadas con miembros de la población beneficiaria que estén familiarizados con los temas pertinentes. El propósito es comparar las perspectivas de los beneficiarios con conceptos abstractos de los objetivos de la evaluación de impacto.	Ventajas similares a las entrevistas. Especialmente útiles cuando se desea una interacción entre los participantes. Una manera útil de identificar las influencias jerárquicas.	Pueden ser costosos y tardar mucho tiempo. Deben ser sensibles a la combinación de niveles jerárquicos. Los resultados no se pueden generalizar.
Entrevistas	El entrevistador plantea preguntas a una o más personas y registra las respuestas de los declarantes. Las entrevistas pueden ser formales o informales, directas o por teléfono, de interpretación cerrada o abierta.	Las personas e instituciones pueden explicar sus experiencias en sus propias palabras. Son flexibles. Muy útiles cuando prevén dificultades de lenguaje. Mayores posibilidades de obtener información de funcionarios superiores.	Tardan mucho tiempo. Pueden ser costosas Si no se realizan adecuadamente, el entrevistador puede influir en la respuesta del entrevistado.
Observación	Observación y registro de una situación. Esto incluye: quién participa, qué sucede, cuándo, dónde y cómo. La observación puede ser directa o participativa (el observador se hace parte del entorno durante un periodo).	Proporciona información descriptiva sobre el entorno y los cambios observados.	La calidad y utilidad de los datos dependen en gran medida de las capacidades de observación y redacción del encuestador. Los resultados pueden ser susceptibles a diversas interpretaciones. No se aplican fácilmente dentro de un plazo breve al cambio del proceso.
Cuestionarios	Elaboración de un conjunto de preguntas de encuestas cuyas respuestas se pueden codificar coherentemente.	Pueden llegar simultáneas a una muestra amplia. Dar tiempo a los declarantes para pensar antes de responder. Se pueden responder en forma anónima. Impone uniformidad al preguntar lo mismo a todos los declarantes. Facilita la recopilación y comparación de los datos.	La calidad de las respuestas depende en alto grado de la claridad de las preguntas. A veces es difícil convencer a las personas para que completen y devuelvan el cuestionario. Puede significar que hay que ajustar las actividades institucionales y las experiencias de las personas dentro de las categorías predeterminadas.
Análisis de documentos escritos	Revisión de documentos como: registros, bases de datos administrativos, materiales de capacitación, correspondencia, etc.	Permite identificar problemas e investigarlos en más detalles y proporciona evidencias de acciones, cambios y efectos para apoyar las percepciones de los declarantes. Pueden ser costosos.	Puede tardar mucho tiempo.

El diseño del cuestionario es importante para la validez de la información recopilada. Existen cuatro tipos de información necesarios para realizar una evaluación de impacto (Valadez & Bamberger, 1994). Estos incluyen:

- Clasificación de los datos nominales y clasificación de los declarantes según si son participantes en el proyecto o si pertenecen al grupo de comparación.
- Exposición a las variables de tratamiento en que se registra no sólo los servicios y beneficios recibidos, sino también la frecuencia, cantidad y calidad.
- Variables de resultados para medir los efectos de un proyecto, incluidos los productos inmediatos, resultados sostenidos o la entrega permanente de servicios durante un período prolongado, y los efectos del proyecto.
- Variables de intervención que afectan la participación en un proyecto o el tipo de efecto producido, como características de los individuos, de los hogares o de la comunidad; estas variables pueden ser importantes para examinar los sesgos.

La forma como se plantea la pregunta, así como el ordenamiento de las preguntas, también es muy importante al momento de generar información confiable. En general se recomienda utilizar indicadores objetivos más que subjetivos (como la percepción).

La literatura recomienda una serie de requisitos para un buen cuestionario: mantener su brevedad y concentración en preguntas importantes, asegurar que las instrucciones y las preguntas sean claras, limitar las preguntas a aquellas necesarias para la evaluación, incluida una opción “no opina” para las preguntas cerradas para garantizar datos confiables y usar procedimientos sólidos para aplicar el cuestionario, que desde luego pueden ser diferentes para las encuestas cuantitativas y las cualitativas (Baker, 2000).

El personal necesario para recolectar datos de campo será diferente de acuerdo con los objetivos y enfoque de la evaluación. Por ejemplo, una evaluación cuantitativa de los efectos de un programa de nutrición podría requerir la inclusión de un especialista en antropometría como parte de un equipo de encuestas, para que reúna medidas de altura por peso, mientras que la evaluación de impacto de una reforma educacional probablemente incluiría personal especialista en la aplicación de pruebas de rendimiento para medir el efecto de la reforma en el rendimiento académico. La mayoría de las encuestas cuantitativas requerirá al menos un administrador de encuestas, un administrador de datos, un administrador en terreno, supervisores en terreno, entrevistadores, operadores de ingreso de datos y conductores. Según el enfoque cualitativo que se use, el personal en terreno podría ser similar, con la excepción de los operadores de ingreso de datos. Sin embargo, los conocimientos de los entrevistadores serían bastante diferentes, ya que los entrevistadores cualitativos requieren una capacitación especializada, en particular para los grupos representativos, observación directa, etc.

2.3.7.10. Administración de datos, análisis de la información y publicación

Un buen sistema de administración de datos debe garantizar la oportunidad y calidad de los datos de la evaluación. Su oportunidad dependerá de la máxima integración posible entre la recopilación de datos y el procesamiento, de modo que se puedan verificar y corregir errores antes de completar el trabajo en terreno.

Se puede garantizar la calidad de los datos aplicando verificaciones de coherencia (control de calidad) para probar la validez interna de los datos recolectados durante y después del ingreso de los datos y asegurándose de que los analistas que usen los datos dispongan de la documentación adecuada. La documentación debe comprender dos tipos de información: (a) la información necesaria para interpretar los datos, incluidos libros de códigos, diccionarios de datos, guías para variables creadas y todas las traducciones necesarias; y (b) la información necesaria para realizar el análisis, la que con frecuencia se incluye en un documento de información básica que contiene una descripción del enfoque y objetivo de la evaluación, detalles sobre la metodología empleada, resúmenes o copias de los instrumentos de recolección de datos, información sobre la muestra, una descripción del trabajo en terreno y directrices para usar los datos.

Al igual que con otras etapas del proceso de evaluación, el análisis de los datos de la evaluación, ya sean cuantitativos o cualitativos, requiere la colaboración entre los analistas, generadores de datos y autoridades responsables para aclarar preguntas y garantizar resultados oportunos y de calidad. Seguramente surgirán problemas con la depuración e interpretación de los datos durante el análisis y se requerirán los aportes de diversos miembros del equipo.

El análisis de datos debe seguir procedimientos rigurosos y acorde a los objetivos planteados. Sea información cualitativa como cuantitativa, los análisis deben ser pertinentes y con suficiente poder estadístico para establecer conclusiones con la suficiente validez.

Por último, una etapa final es discutir los resultados encontrados y analizar las posibilidades explicativas del resultado obtenido. Antes de su publicación, el informe debe ser sometido a crítica de la comunidad académica y local. Finalmente, el informe final deberá difundirse mediante presentaciones para diversos públicos destinatarios, boletines de prensa, retroalimentación a los informantes y la publicación de la información en la Web. Esta estrategia de difusión se considera en las etapas iniciales del proceso de planificación, para garantizar que se incorpore al presupuesto y que los resultados lleguen al público previsto.

2.4. DEFINICIÓN DE CONCEPTOS

Evaluación ex ante

Determinación general de la pertinencia, la factibilidad y la sostenibilidad potencial de una intervención para el desarrollo antes de tomar la decisión de otorgar financiamiento. En los organismos, bancos, etc., de desarrollo, el propósito de la evaluación *ex ante* es presentar información durante el proceso de toma de

decisiones para determinar si la actividad representa el uso adecuado de los recursos de la institución. Término conexo: pre-evaluación.

Contrafactual

Situación o condición que podría presentarse hipotéticamente para las personas, las organizaciones o los grupos si no hubiera una intervención para el desarrollo como la evaluada.

Evaluación ex post

Evaluación de una intervención para el desarrollo que se efectúa una vez que ha concluido. Puede emprenderse inmediatamente o mucho después de su conclusión. La intención es identificar los factores de éxito o fracaso, evaluar la sostenibilidad de los resultados y las repercusiones, y extraer conclusiones que puedan ser útiles para otras intervenciones.

Impactos

Efectos de largo plazo, positivos y negativos, primarios y secundarios, producidos directa o indirectamente por una intervención para el desarrollo, intencionalmente o no.

Evaluación de programas

Evaluación de un conjunto de intervenciones estructuradas para alcanzar objetivos específicos de desarrollo a nivel global, regional, de país o sectorial (GAO, 1998). Un programa de desarrollo es una intervención de duración limitada que comprende múltiples actividades que pueden abarcar varios sectores, temas y/o zonas geográficas.

Monitoreo

El monitoreo comprende el seguimiento continuo de cómo se va implementando un proyecto en comparación a cómo se acordó en sus inicios que debería implementarse. Así, el monitoreo permite conocer si se van cumpliendo los cronogramas acordados y si están llegando a los beneficiarios del proyecto aquellos elementos que se planearon, ya sean insumos, infraestructura o servicios (OED, 1994).

Modelo de evaluación

Un modelo de Evaluación es un diseño (Ventosa Pérez, 1992) cuya función básica es proporcionar un marco de referencia, unas directrices que le orienten en la elaboración de una evaluación concreta. Un modelo proporciona al evaluador un marco conceptual y racional que le permite elaborar su diseño y poner en marcha el proyecto de evaluación. En el modelo aparecen implícita o explícitamente los presupuestos epistemológicos del autor, sus concepciones teóricas, criterios metodológicos y sus principios (González Soler, 1987).

Evaluación de impacto

La evaluación de impacto es una forma de evaluación de resultados que evalúa la serie de efectos de un programa en comparación con los resultados estimados que hubiesen ocurrido en ausencia del programa. (GAO, 1998). A este último se le conoce como estado contrafactual. Es un tipo de evaluación *ex post*.

Programas sociales

Conjunto específico de acciones humanas y recursos materiales, diseñados e implantados organizadamente en una determinada realidad social, con el propósito de resolver algún problema que atañe a un conjunto de personas. Todo programa social persigue mejorar algunas condiciones de vida que pueden ser de salud, infraestructura, educación, asistencia, justicia, entre otras.

III. DESARROLLO METODOLÓGICO ACTUAL: UNA REVISIÓN SISTEMÁTICA

El proceso de evaluación de los programas sociales es un fenómeno creciente. Según Mertens & Russon (2000) el número de sociedades de evaluación nacionales y regionales ha crecido de cinco, en 1995, a más de 30, en el año 2000. Con esta explosión de sociedades profesionales, también han crecido las especializaciones, redes de información y revistas internacionales sobre la materia (Mark, 2001).

Al respecto, hoy existen diversas fuentes directas de información especializada y de divulgación sobre evaluación de programas. Estas fuentes son revistas internacionales especializadas y han sido base primaria en la obtención de información metodológica que sustenta el Modelo que se propone en ésta obra.

En la siguiente tabla se presenta una lista de las revistas internacionales que contienen artículos especializados sobre evaluación social y, dentro de ellas, de evaluación de impacto.

Tabla N° 3.1. Revistas especializadas sobre evaluación de programas (Fuente: Elaboración propia).

Revista internacional	Ediciones
Evaluation and Program Planning	4 números por año. Desde 1978.
American Journal of Evaluation	4 números por año. Desde 1998.
Canadian Journal of Program Evaluation	3 números por año. Desde 1986.
Educational Evaluation and Policy Analysis	4 números por año. Desde 1979.
Evaluation and the Health Professions	4 números por año. Desde 1999.
Evaluation Review: A Journal of Applied Social Research	6 números por año. Desde 1999.
Evaluation: The International Journal of Theory, Research and Practice	4 números por año. Desde 1998.
Evidence & Policy: A Journal of Research, Debate and Practice	4 números por año. Desde 2005.
Zeitschrift für Evaluation (Journal for Evaluation)	2 números por año. Desde 2002.
LeGes (solo en Alemania)	No informa.
New Directions for Evaluations	4 números por año. Desde 1978.
Practical Assessment, Research and Evaluation	1 número por año. Desde 1999.
Studies in Educational Evaluation	4 números por año. Desde 1974.
The Evaluation Exchange	4 números por año. Desde 1995.
Japanese Journal of Evaluation Studies	3 números por año. Desde 2001.
Journal of Multi Disciplinary Evaluation	2 números por año. Desde 2004.
Performance Evaluation	12 números por año. Desde 1981.
Research Evaluation	3 números por año. Desde 1998.

Al respecto, para la elaboración del fundamento teórico-metodológico del Modelo ARYSIS (véase Capítulo IV) se han revisado los artículos contenidos en cada una de estas revistas, identificándose las investigaciones más significativas y con mayor aporte teórico-metodológico en la evaluación de programas referido al impacto.

A continuación se presentará, de forma muy sucinta, algunas experiencias en evaluación de impacto que han sido desarrolladas tanto en el ámbito nacional como internacional, así como un análisis detallado del desarrollo académico sobre el tema.

3.1. DESARROLLOS NACIONALES

En el ámbito nacional es difícil encontrar referencias teóricas-metodológicas serias sobre la evaluación de impacto de los programas sociales. En primer lugar, porque no existe aún mucha demanda de ello. En segundo lugar, porque la bibliografía en habla hispana sobre evaluación de impacto es casi nula o muy elemental. A pesar de lo dicho, existen algunas experiencias institucionales de evaluación de impacto –productos de consultoría externa- que rayan con conceptos y técnicas modernas de evaluación.

En efecto, dada la escasa tradición sobre evaluación de impacto de los programas sociales en el país, son muy pocos los trabajos sobre el tema que dan cuenta de la metodología empleada para tal evaluación y los resultados encontrados. Recientemente, Yamada & Pérez (2005) han publicado *“Evaluación de impacto de proyectos de desarrollo en el Perú”*, donde identifican las principales experiencias peruanas de evaluación de impacto hasta esa fecha, introduciendo de manera no técnica las metodologías utilizadas, y discutiendo las restricciones más importantes que se enfrentaron y cómo fueron resueltas. Otros importantes trabajos de evaluación de impacto de programas sociales en el país son el “Costo Efectividad del Programa de Desayunos Escolares de Foncodes y el Programa de Alimentación Escolar del Pronaa”, efectuado por Ravina, Paulini y Cancho (2002); “Una Metodología de Evaluación de Impacto de los Programas de Educación y/o Capacitación en el Mercado Laboral: el Caso de PROJoven” de Galdo (2000); y “Una Medición del Impacto del Programa de Capacitación Laboral Juvenil PROJoven” de Ñopo, Robles y Saavedra (2002), entre otros.

El estudio de **Ravina, Paulini y Cancho (2002)** analiza el costo-efectividad de dos programas de ayuda nutricional implementados al interior de escuelas (el PDE de Foncodes y el PAE del Pronaa). Emplea una técnica analítica (Análisis Costo-Efectividad) que permite comparar los costos de un proyecto con los beneficios resultantes, no expresados en la misma unidad de medida. En este caso, los costos de cada programa incluyen tanto los costos en alimentos como los costos operativos, mientras que la efectividad es evaluada en términos del impacto del programa sobre la asistencia de los niños a la escuela primaria. Para la estimación cuantitativa de dicho impacto, se utiliza la información de la Encuesta Nacional de Hogares (ENAHOG) correspondiente al segundo trimestre de 1999. En concreto, se aplica un modelo Probit binomial (para cada programa) donde la variable dependiente toma valor 1 si el niño asiste a la escuela y 0 si no lo hace; y

se consideran como variables explicativas a la variable de intervención (que toma valor 1 en caso el niño habite en un centro poblado donde se implemente el programa y 0 en caso contrario), así como algunas características de los niños, del hogar, de la vivienda y del contexto geográfico en el que habitan. De esta forma, se estima el efecto “puro” del programa sobre la asistencia escolar. Con el propósito de atenuar el problema del “sesgo de selección” por “variables observables” que estaría presentando la variable de intervención, se encontró previamente para cada programa un grupo de comparación a través de la metodología del *propensity score*. A partir de un modelo Probit binomial, se calculó la probabilidad de cada niño de ser potencial beneficiario del proyecto (pertenecer a un centro poblado donde se implementa el proyecto); seguidamente se seleccionó a aquellos potenciales beneficiarios para llevar a cabo un proceso de emparejamiento entre dicho grupo y el de no beneficiarios, minimizando las diferencias en las probabilidades estimadas mediante un proceso de apareamiento con reemplazo. De este modo, se aseguró que a cada niño, potencial beneficiario del programa, se le asigne un niño de un centro poblado donde no se implementa el mismo. Para dar solución al problema de “sesgo de selección” por variables “no observables”, se halló, a través del modelo Probit planteado para el cálculo del *propensity score*, una función monótona decreciente conocida como “ratio inverso de Mills” que luego se incorporó como un regresor adicional en la estimación del impacto de cada programa, es decir en la regresión de asistencia escolar⁴¹. De este estudio se concluye que el PDE de Foncodes ha tenido mejores resultados que el PAE del Pronaa (menores costos por unidad de efectividad).

Por otro lado, **Cueto y Chinen (2000)** han evaluado el impacto educativo de un programa de desayunos escolares en escuelas rurales de la zona alto andina del Perú. En la evaluación se comparan medidas de peso, talla, tasas de matrícula y deserción escolar, asistencia, aptitudes cognoscitivas y rendimiento en pruebas estandarizadas de matemática y comprensión de lectura en estudiantes de cuarto grado de primaria de un grupo de tratamiento que recibió desayuno por al menos tres años y un grupo de contraste, que no participó del programa. El aspecto más significativo del estudio es el diseño cuasi-experimental con control mediante Regresión Jerárquica.

Por otro lado, el estudio de **Galdo (2000)** evalúa el impacto de la primera convocatoria del programa de capacitación laboral juvenil PROJoven sobre la inserción laboral remunerada, los ingresos y la calidad de empleo de los jóvenes participantes de este programa, comparándolos con un grupo de control para aislar los posibles efectos del entorno macroeconómico o del proceso de maduración de los propios individuos. La metodología de evaluación empleada se basa en el modelo no experimental del *matched comparisons*. Inicialmente se estableció una equivalencia entre una muestra de beneficiarios y un grupo de control mediante un

⁴¹ Sin embargo, la metodología utilizada en este trabajo incurre en error al pretender combinar la metodología de emparejamiento o “matching comparisons” con los modelos econométricos de sesgo de selección que hacen uso del ratio inverso de Mills. En particular, al usarse una corrección por sesgo de selección del tipo ratio inverso de Mills se debió haber estimado una ecuación de asistencia escolar con sólo la muestra de “no beneficiarios” del programa, o alternativamente con sólo la muestra de los beneficiarios.

estricto apareamiento en dos etapas: se identificó a los potenciales integrantes del grupo de control mediante un mapeo de las zonas de origen de los beneficiarios y luego se procedió a aparearlos uno a uno (*cell matching*) con los integrantes del grupo de beneficiarios, sobre la base de la edad, sexo, nivel educativo, estrato socioeconómico y situación laboral previa al inicio del programa de los individuos. Seis meses después de concluido el programa (o 12 meses después del inicio del mismo), se hizo una comparación de los cambios mostrados por los beneficiarios y controles en las variables bajo estudio. En algunos casos incluso se distinguió el efecto por género, edad y nivel educativo de los individuos. La evaluación de impacto también incluye una estimación econométrica del diferencial de ingresos, antes y después del programa, entre el grupo de beneficiarios y controles. Para ello, se considera como variables independientes a una variable dicotómica que toma valor 1 para el grupo de beneficiarios y 0 para el grupo de control, y a un conjunto de variables de control postprograma (variables laborales y de capital humano). Para corregir el problema de “sesgo de selección” que pudiera estar presentando la variable dicotómica se utilizó, al igual que en el trabajo de Ravina et al. (2002), la técnica econométrica del “ratio inverso de Mills”: se estimó un modelo Probit binomial para identificar las variables que inciden mayormente en la participación de los jóvenes en PROJoven y con los residuos de esta relación se obtuvo el ratio en mención, que luego fue incorporado como regresor en una ecuación que estima el diferencial de ingresos después y antes de la participación en el programa, usando para dicha estimación la muestra tanto de beneficiarios como de controles⁴². Se concluye en este trabajo que PROJoven logra un impacto estadísticamente positivo sobre la situación laboral de los jóvenes beneficiarios.

Otro ejemplo significativo es la evaluación de impacto realizado por **Ñopo, Robles y Saavedra (2002)** en el programa de capacitación laboral juvenil PROJoven. Se utilizó información socio-laboral de una muestra de beneficiarios y un conjunto de individuos que conformaron el grupo de control. Esta información fue recogida antes de iniciados los cursos de capacitación de la sexta convocatoria (línea de base) y después de seis meses de concluido el programa por parte de los jóvenes beneficiarios (primera medición de salida). La muestra total de beneficiarios disponible para la evaluación fue de 1018, representando poco menos del 30% del total de jóvenes atendidos en la sexta convocatoria de PROJoven. Asimismo, se contó con información para 1,561 potenciales controles, es decir, individuos que poseen características comparables a las de los beneficiarios pero que no fueron parte del programa. El objetivo es comparar la nueva situación económica de los beneficiarios, como consecuencia de su participación en PROJoven, con la situación que hubiesen alcanzado en el caso de no haber asistido al programa. Como primer paso para la evaluación de impacto se realizó un proceso de emparejamiento (“*matching*”) entre beneficiarios y controles. Este proceso buscó asegurar que beneficiarios y controles sean comparables en las siguientes variables: sexo, edad, nivel de educación, situación laboral, nivel de pobreza, ingreso por hora pre-programa, horas trabajadas a la semana pre-programa y la propensión a participar en el programa (“*propensity score*”). El

⁴² En este trabajo metodológicamente se incurre en similar error que el comentado en el trabajo de Ravina et al. (2002).

método de emparejamiento utilizado fue el de minimización de distancias euclidianas. Se evaluaron cuatro tipos de impacto del programa sobre los jóvenes beneficiarios. En primer lugar, se consideró el efecto de ProJoven sobre la inserción laboral de los participantes. El estimador de “diferencia en diferencias” indica que el efecto del programa sobre la inserción laboral es de 6%, es decir un 6% adicional de los beneficiarios se encuentran ocupados en comparación a los que estuviesen ocupados de no haber participado en el programa. El valor de dicho efecto, calculado para el total de jóvenes participantes en la sexta convocatoria, se estima en US\$ 207 mil dólares anuales. Utilizando un análisis econométrico que controla por diferencias en las características de beneficiarios y controles que no se captan completamente en el proceso de emparejamiento, se obtiene una valoración de US\$ 166 mil dólares al año.

A continuación se presenta una tabla conteniendo un resumen de diversos estudios de evaluación de impacto realizados por las instituciones más reconocidas en el país durante la última década (Grade, Cuanto y Apoyo) y por instituciones internacionales como el Banco Mundial, USAID o el Banco Interamericano de Desarrollo (BID).

Tabla N° 3.2. Evaluaciones de impacto más robustas de los programas sociales en el Perú (Fuente: Elaboración propia).

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
Empleo	El Impacto del Programa "A Trabajar Urbano": Ganancias de ingreso y utilidad de las obras.	Centro de Estudios para el Desarrollo y la Participación (2003) Juan Chacaltana Institucional	Se realizó una selección aleatoria. Por razones de pérdida de parte de la muestra, se terminó con una muestra de 125 proyectos y 1500 participantes.	El ejercicio de evaluación de impacto se realiza con dos bases: La Encuesta Nacional de Hogares del cuarto trimestre del 2002 (ENAHO 2002-IV, INEI) y la Encuesta a participantes del Programa ATU. Se utilizó una Metodología de evaluación Cuasi Experimental. Para ello, se estimó el estado contrafactual a través de la comparación entre grupo de beneficiarios y no beneficiarios del programa, utilizando el modelo de emparejamiento de probabilidades de participación (Propensity score Matching).	La ganancia neta derivada del programa es de 61 soles, es decir, un 22% de la transferencia nominal. Sobre la base de la información y supuestos razonables sobre retornos de inversiones, se encontró que los beneficios adicionales para la comunidad de las obras representan un 93% adicional de la inversión inicial realizada por el programa en pago de mano de obra.
Empleo	Una medición del impacto del Programa de Capacitación Laboral Juvenil PROJoven	Grupo de Análisis para el Desarrollo (2002) Hugo Ñopo, Miguel Robles y Jaime Saavedra Institucional - GRADE	Se obtuvo una muestra aleatoria por muestreo estratificado de 842 beneficiarios apareados con sus respectivos controles, a partir de los cuales se cuantifica el impacto de PROJoven.	Se utilizó un método de evaluación Cuasi Experimental. Se estableció la línea base con una muestra de 3,586 beneficiarios a quienes se les aplicó una Encuesta sociolaboral. Se procedió a la comparación de grupos través de la comparación entre grupo de beneficiarios y no beneficiarios del programa, utilizando el modelo de emparejamiento por minimización de las distancias euclidianas.	La medición de impacto de la sexta convocatoria del programa PROJoven revela que los efectos del programa sobre los jóvenes beneficiarios son positivos.
Nutrición	Impacto del Programa del Vaso de Leche	CUANTO (1997 y 2000) Institucional	Hogares con niños menores de 6 años de edad ubicados en: Lima Metropolitana, Costa Urbana, Costa Rural, Sierra Urbana, Sierra Rural, Selva Urbana, Selva Rural.	Diseño de corte transversal, pre-test y post-test, utilizando como instrumento de valoración las Encuestas sobre Niveles de Vida (ENNIV) de 1997 y 2000. Los datos generados por ambas encuestas fueron trabajados mediante un análisis	Resalta el hecho de encontrarse en la ENNIV de 1997 una asociación directa de pequeña intensidad entre la pertenencia al Programa del VdeL y la Desnutrición Crónica (DNC).

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
				multivariado de Regresión Logística.	
Agricultura	Evaluación del Impacto del Programa de Reforzamiento al Desarrollo Alternativo-Sección 416(b)	Codina Giralt (2005) DEVIDA Institucional	Para el diseño de la muestra se formaron 2 estratos, beneficiarios de Aucayacu y beneficiarios de Aguaytía, ya que se deseaba analizar la información a nivel de cada uno de los valles. La muestra final estuvo conformada por 590 beneficiarios, 320 de Aguaytía y 290 de Aucayacu.	Se usó un diseño pretest-postest sin grupo control. No se pudo utilizar como línea base la Encuesta Anual del PDA del año 2002, porque las poblaciones bajo estudio eran diferentes. Por acuerdo con el Comité Técnico de DEVIDA se decidió que el Impacto de los proyectos productivos del Programa se medirían comparando los ingresos percibidos por los beneficiarios en la campaña anterior a la campaña en la que recibieron el apoyo al programa. Se usó la Encuesta de Hogares por Muestra. Los cuestionarios diseñados fueron sometidos a una prueba piloto en el campo y se efectuaron las modificaciones que esa experiencia aconsejó. Se utilizaron los siguientes cuestionarios: <ul style="list-style-type: none"> • Formulario: Beneficiarios del Proyecto Agropecuario, Apoyo Alimentario y Empleo Temporal. • Formulario: Censo a los Beneficiarios del Proyecto Palma Aceitera en Aguaytía 	No fue posible medir el impacto económico del proyecto; sin embargo los entrevistados perciben que el cultivo es rentable, por lo que intentarán ampliar sus áreas de bajo cultivo. La tasa de incremento en el valor bruto de la producción agrícola asistida por el Programa fue del 25% con relación a la campaña agrícola anterior al apoyo del Programa. No se puede determinar si hubo aumento o decremento de los conflictos sociales.
Calidad de vida	Sexta Evaluación Intermedia y Ex post de Foncodes FONCODES	APOYO (2000) Institucional	La muestra estuvo conformada por 800 proyectos y cerca de 5,000 hogares	Se comparó a hogares intervenidos con hogares no intervenidos (grupo de control) respecto a indicadores de salud, educación y condiciones de vida en general. La metodología utilizada en estos proyectos incluye cuestionarios a nivel nacional y entrevistas con autoridades y beneficiarios.	Se encontró impacto significativo.
Alimentación	Costo Efectividad del	APOYO	El universo muestral lo	Se decidió llevar a cabo el ejercicio	Se encontró que el Programa

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
	Programa de Desayunos Escolares de FONCODES y el Programa de Alimentación Escolar del PRONAA	(2000) Ravina Renato & Paulini Javier Institucional	conformaron los niños entre 5 y 14 años, que no hubieran acabado aún sus estudios primarios.	econométrico utilizando un modelo Probit. Se intento atenuar el problema del sesgo de selección por variables observables aplicando el método conocido como propensity score. Para la estimación de la efectividad de los programas de Foncodes y Pronaa se usó la información de la Encuesta Nacional de Hogares (ENAHOG) correspondiente al 2º trimestre de 1999.	de Desayunos Escolares presentó menores niveles de costos por beneficiario que el Programa de Alimentación Escolar.
Empleo	Evaluación de Impacto del Programa de Desarrollo y Capacitación a Microempresas de Mujeres	PROPYME (1997) Universidad del Pacifico	El número de clubes beneficiarios entrevistados alcanzó a 96.	Se aplicó una encuesta a los clubes de madres que habían recibido créditos otorgados por el proyecto.	El efecto del programa en el ingreso familiar puede ser comprobado por la percepción de las socias y la sus esposos respecto a que su participación en el club es importante en lo que se refiere a su aporte a la economía familiar.
Nutrición	Evaluación de impacto del Programa de Complementación Alimentaria para Grupos de Mayor Riesgo	Consortio Maximixe - Instituto CUANTO 2003	La muestra estuvo conformada por 432 niños beneficiarios del programa.	El diseño fue descriptivo y transversal. Se aplicó el método retrospectivo de recordatorio del consumo de alimentos de 24 horas.	Se evidenció que la papilla mejoró el aporte de energía, vitamina A y hierro en la dieta del niño, llegando a cubrir los requerimientos en la mayoría de los grupos etáreos
Salud	Evaluación de Impacto del Programa Alimentario Nutricional para el Paciente Ambulatorio con Tuberculosis y Familia PANTBC - 2002	CENAN-INS. (2003) Institucional	Se calculó una muestra aleatoria por conglomerados de 264 encuestas a ex beneficiarios PANTBC y de 135 encuestas a beneficiarios actuales. Todos debían tener entre 18 y 65 años.	La evaluación presenta un estudio descriptivo transversal prospectivo de fuente primaria de información, describiéndose Se diseñaron sistemas de ingreso de datos en el lenguaje de programación Visual Basic 6.0. El análisis y procesamiento estadístico de los datos se efectuó mediante el software SPSS (Versión 10.0)..	El diseño de la muestra permitió estimar una prevalencia de bajo peso entre los pacientes evaluados del 50%, con una precisión de +/- 7.5%, un nivel de significancia del 95%

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
Nutrición	Evaluación de Impacto del Programa Alimentario Nutricional para Familias en Alto Riesgo-PANFAR 2002	CENAN-INS. 2003 Institucional	Se estuvo una muestra aleatoria por conglomerado de 1435 niños beneficiarios del PANFAR pertenecientes al segundo semestre del año 2002.	Estudio de tipo descriptivo transversal El procesamiento estadístico se efectuó mediante el software SPSS (Versión 10.0). Las variables continuas se describieron utilizando medidas de tendencia central y dispersión. Mientras que las variables nominales se describieron mediante frecuencias absolutas y relativas.	No se encontró impacto.
Educación	Mejor Educación a través de más Tiempo en el Aula (META)	GRADE (2004) Torero Máximo & Santiago Cueto Institucional	La muestra está constituida por centros educativos rurales primarios de cinco departamentos seleccionados para participar del programa META.	El diseño es cuasi-experimental. El grupo de contraste estuvo conformado por los estudiantes y docentes del departamento de Ayacucho. Se registró la asistencia diaria para cada docente en el programa desde marzo hasta noviembre del 2004. Así mismo se cuenta con pruebas de rendimiento de los estudiantes a inicios y fines del 2004 en las áreas de matemática y comprensión de lectura. El método que se sigue en el presente trabajo es el de emparejamiento o "matching" que consiste en asignar a cada beneficiario un respectivo control, es decir un individuo que no participa en el programa pero que es "muy parecido" a él.	A pesar de haber encontrado un efecto positivo del programa en cuanto a la asistencia de los docentes, los resultados de las pruebas de rendimiento en cuarto y sexto de primaria no son concluyentes respecto a un impacto positivo del programa en el aprendizaje de los estudiantes.
Infraestructura	Estimando el Beneficio de los Caminos Rurales: Programa de Caminos Rurales (PCR)	GRADE (2001) Javier Escobal y Carmen Ponce Institucional	La selección de hogares y centros poblados beneficiados indirectamente o no beneficiados por el PCR se realizó a partir de la muestra aleatoria de caminos rehabilitados por el programa, buscando "parear" tramos de caminos testigos con tramos	Se aplicaron encuestas con el fin de servir de base para el análisis de los beneficios derivados de la inversión en rehabilitación y mantenimiento de caminos rurales. Se evitó el desbalance de la muestra, identificando los hogares que más se asemejaban a aquellos que han sido sujetos de intervención a partir del método del	Las estimaciones señalan un efecto positivo, que varía entre 6,8% y 7,5% para los caminos carrozables y entre 7,3% y 9,2% para los caminos de herradura. Aunque ciertamente no se trata de impactos dramáticos, es lo que se podría esperar en un

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
			rehabilitados por el PCR.	Propensity Score Matching. También se realizaron observaciones de hogares ficticios que maximicen la comparabilidad entre la muestra sujeta a intervención (asociada a caminos rehabilitados) y el grupo de control (asociado a caminos no rehabilitados).	contexto donde los impactos son usualmente de largo plazo.
Agricultura	Desarrollando mercados rurales: El rol de la incertidumbre y la restricción crediticia	GRADE (2005) Javier Escobal Institucional	La muestra empleada fue de 604 productores agropecuarios en distintas zonas de intervención del proyecto INCAGRO. El grupo de tratamiento lo integran aquellos que han hecho uso de servicios de asistencia técnica que les permiten acceder a mercados dinámicos (181) y el grupo control los productores que no han accedido a asistencia técnica (423).	Para medir la restricción crediticia, hemos seguido la propuesta elaborada por Diagne et al. (2000). Este método permite evaluar el impacto de relajar la restricción crediticia sobre cualquier variable de interés. Respecto de la medición del riesgo, las preguntas hipotéticas han sido diseñadas para calcular una medida relativa de aversión al riesgo.	Los resultados presentados hacen evidente que existe una “distancia” importante entre quienes ahora acceden al mercado de asistencia técnica y quienes estando en las mismas zonas aún no lo logran.
Ingresos Familiares	Institutions, Politics and Contracts: The Attempt to Privatize the Water and Sanitation Utility of Lima, Peru	Banco Mundial (2000) Alcazar, Lorena; Xu Lixin, Colin; Zuluaga, Ana María. Institucional	No específica	La evaluación utilizó una metodología de costo-beneficio para medir el beneficio de los efectos de la reforma. Se utilizó un método de evaluación de diferencias de medias y la simulación contrafactual. Período de evaluación 1994-2004	Se encontró que los beneficios alcanzados de las reformas parciales son mucho menos que los que se obtendrían si hubiera una concesión. Las reformas parciales son mejores para el gobierno pues mantienen las rentas de mayores precios, pero los consumidores pierden debido a una muy baja inversión.
Ingresos Familiares	The Social Impact of privatization and the Regulation of Utilities in Peru	Banco Mundial (2001) Torero, Maximo y Pasco-Font, Alberto	No específica	Se utilizaron tres metodologías complementarias: a) cálculo de las curvas de tendencias que muestra cómo los servicios se distribuyen en la población, b) medición de	El consumo y gasto en los servicios de teléfono se ha incrementado de 1991 a 1997 para todos los niveles de

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
		Institucional		<p>los cambios en el gasto familiar asociados con los cambios en la estructura y nivel de los precios, y c) medición de los efectos del incremento en el acceso.</p> <p>Método de evaluación Diferencia de medias y simulación contrafactual.</p> <p>Período de evaluación de 1991 a 1997</p>	<p>ingreso debido a las menores tarifas.</p> <p>El consumo y gasto en electricidad en el área urbana del Perú ha decaído debido al incremento de los precios del servicio eléctrico.</p> <p>Los gastos en el servicio de agua han decaído de 1991 a 1997.</p>
Microcrédito	Evaluating Social Funds: A Cross-Country Analysis of Community Investments	<p>Banco Mundial (2004)</p> <p>Rawlings, Laura, Lynne Sherburne-Benz, & Julie Van Domelen</p> <p>Institucional</p>	Áreas geográficas con altos niveles de fondos de inversión social (grupo de tratamiento) y áreas con bajos niveles de fondo de inversión social (grupo control)	<p>Se utilizó la técnica del emparejamiento para comparar áreas geográficas, y un diseño de estimación de diferencia en la diferencia, así como variables instrumentales.</p> <p>La línea fue establecida por la información de la <i>Living Standards Measurement Survey</i> (1994 y 1997) y el estudio de hogares conducido por el Instituto Nacional de Estadística e Informática.</p>	El efecto del programa sobre las tasas de empadronamiento escolar fue positivo y significativo.
Educación	The Allocation and Impacts of Social Funds: Spending on School Infrastructure in Peru	<p>Banco Mundial</p> <p>Paxson, Christina, and Norbert Schady. (2002)</p> <p>Institucional</p>	Niños de 6 a 11 años de edad	<p>La información de nivel distrital vino del Censo de Población y Vivienda de 1993 y la información a nivel de hogares vino de un estudio realizado por el Instituto Nacional de Estadística e Informática (1996) y de dos estudios sobre medición de los estándares de vida (1994 y 1997)</p> <p>Se utilizó un método de evaluación de Diferencia en la diferencia y de variables instrumentales.</p>	El efecto de los gastos escolares de FONCODES sobre las tasas de asistencia escolar fue positivo y significativo.
Pobreza	Sexta Evaluación Ex-Post del FONCODES: Evaluación de Impacto y Sostenibilidad	<p>Banco Mundial APOYO (2000)</p> <p>Institucional</p>	Se escogió 470 proyectos sobre la base de una selección sistemática con arranque aleatorio para conformar el grupo de intervención.	Se realizó un diseño metodológico con dos componentes: (i) un diseño probabilístico para la evaluación de cinco líneas (educación, agua, desagüe, letrinas y electrificación) basado en la definición de un	El fondo social de inversión en educación incrementó significativamente la probabilidad de ser admitido en la escuela por casi 2% de

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
			Se seleccionaron otros 470 proyectos de la lista de espera de proyectos de Foncodes para conformar el grupo de comparación.	grupo de comparación ad hoc y en el uso de instrumentos de cuantificación de respuestas por parte de beneficiarios, operadores y responsables de las obras; y (ii) un diseño no probabilístico (estudio de caso) para dos líneas de inversión en el área de infraestructura económica (riego y obras viales) con visitas de campo a proyectos preseleccionados.	aquellos que viven en extrema pobreza. El programa disminuyó significativamente la incidencia de la diarrea en los niños.
Pobreza	Modelo de medición de impacto sobre el bienestar objetivo y subjetivo: Un análisis de caso del Proyecto de Reducción y Alivio a la Pobreza (PRA)	Pedro Mateu y Jean Vilca (2004) Universidad del Pacífico – Agencia de los Estados Unidos para el Desarrollo (USAID) Institucional	424 (50% GC y 50% GE) madres de Cajamarca, sobre la base de un muestreo aleatorio. Incluye las provincias de La Esperanza, Huambocancha, Cajamarca y Ciudad.	Diseño cuasi-experimental pre-post con grupo control posterior obtenido mediante matching. Uso de encuesta estructurada sobre la base de las encuestas HOPE 1998, ENAHO 2001 IV Trimestre y ENNIV 1997. Análisis de regresión con mínimos cuadrados y logística.	Mide los efectos del programa PAR el bienestar subjetivo y objetivo de las madres. El PRA ha elevado los ingresos mensuales y el número de horas diarias laboradas de las participantes en 119.2% y 102.7% respectivamente.
Infraestructura	Programa de Caminos Rurales II y Programa de Transporte Rural Decentralizado	Torres, R., Quispe, E. & Serrie, N. (2006) Banco Mundial & BID. Institucional	El universo del grupo tratamiento se define sobre la base de las dos primeras prioridades de inversión (tramos) identificadas en los Planes Viales Provinciales (PVP): La muestra de tramos se determina de manera aleatoria del universo antes indicado (Alrededor de 20% del total de los caminos rehabilitados serán muestrales). En cada tramo se determina los centros poblados directamente beneficiados al inicio, al centro	Uso de grupo control y tratamiento. Una evaluación ex-post en 2000, que sirve también de línea de base para la siguiente operación. Esta evaluación se hizo de manera retrospectiva Una evaluación intermedia en 2004 Ambos evaluaciones fueron hecho por el método de diferencias de diferencias 2006: evaluación de las primeras fases y línea de base para el siguiente proyecto	En proceso de evaluación.

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
			<p>y al final (dependiendo del tamaño del tramo).</p> <p>En cada centro poblado se determina de manera aleatoria las familias a ser entrevistadas.</p> <p>Para cada uno de los tramos identificados y priorizados en los PVP se identifica otros tramos control con iguales características dentro de la provincia.</p> <p>Se identifica en “gabinete” tramos control, en base a información secundaria. Estos tramos luego se verifican en campo.</p>		
Salud	Midterm Impact Evaluation of the Reproductive health in the Community (reprosalud) Project	Ferrando D, Serrano N. & Pure C. (2002) Institucional (USAID)	<p>112 comunidades: 70 fueron beneficiarias directas del programa, 17 beneficiarias indirectas y 25 integraron el grupo control.</p> <p>La unidad de análisis estuvo integrada por mujeres en edad fértil y hombres entre los 15 a 59 años.</p>	<p>Metodología cuasi-experimental.</p> <p>Diseño de evaluación pre post con grupo control.</p> <p>Se utilizaron cuestionarios para las familias, mujeres, hombres y para las autoridades de salud.</p> <p>Análisis de comparación de grupos y Odds Ratio.</p>	Se encontró un efecto positivo del programa en las prácticas de salud de las mujeres en comparación con el grupo control. No obstante, el impacto es mediano respecto a los hombres.
Infraestructura - Agua y sanidad	Peru: CARE OPG Water Health Services Project	Haratani J, Viveros AM. & Becerra AM. (1981) Ancash – Perú Institucional - USAID	32 funcionarios y líderes entrevistados. Sin muestreo.	<p>Método cualitativo (entrevistas a profundidad).</p> <p>Se entrevistó a los funcionarios de AID y CARE quienes participaron del diseño e implementación del proyecto.</p> <p>Se desarrolló una entrevista semi</p>	La instalación de suministro de agua fue el componente principal de que produjo mayor impacto en la vida de los aldeanos.

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
				estructurada a los líderes de las aldeas, así como también a mujeres.	
Infraestructura	Evaluación de Impacto del Proyecto MSP Sub-Programa Cebolla Amarilla Supe	Centro de Servicios y Elaboración de Proyectos de Inversión - UNMSM (1996) USAID-ADEX-CESEPI Institucional	141 personas: productores y jornaleros agrícolas, consumidores y propietarios.	Aplicación de encuestas, entrevistas con grupos focales de productores.	Se halló un impacto positivo del programa en la población de Supe.
Educación, transporte, suministro de agua y sanidad	Perú: Improved Water and Land Use in the Sierra	Wilkinson JL, Mckean C, Meyer RE, Nunberg BS, Weil B. & Martinez H. (1984) USAID Institucional	Tres sub proyectos del Valle del Mantaro y tres subproyectos Cajamarca.	Revisión de documentos, visitas de campo, entrevista semi estructurada a agricultores. Se realizaron análisis costo beneficio para medir el impacto del proyecto.	El proyecto no obtuvo el impacto esperado
Infraestructura	Does Privatization Deliver? Access to Telephone Services and Household Income in Poor Rural Areas Using a Quasi-Natural Experiment in Peru.	Chong A, Galdo V. & Torero M. (2005) Perú BID Institucional	1000 familias rurales comprometidas en las actividades agrícolas y no agrícolas, beneficiarias o no con el servicio de telefonía privada en su localidad. Se determinó grupo control	Análisis de Regresión, Métodos de emparejamiento Propensity Scores y variables instrumentales.	Se encontró una relación positiva y significativa entre el acceso a los servicios de telefonía y las mediciones del ingreso familiar.
Empleo	Does the Quality of Training Programs Matter? Evidence from Bidding Processes Data	Chong A. & Galdo J. (2006) IZA Discussion Papers (Alemania)	Jóvenes de 16 a 24 años de edad. 1725 beneficiarios y 1742 no beneficiarios. Se determinó grupo control	Se utilizó el método de emparejamiento de diferencia en la diferencia de Kernel y análisis de regresión	Se encontró que los que los jóvenes que asisten a cursos de capacitación de alta calidad obtuvieron mayor promedio e impactos de tratamiento marginal.
Pobreza	Poverty, Health Infrastructure and the Nutrition of Peruvian Children	Valdivia M. (2004) Perú GRADE Economics and Human	368 distritos.	Se utilizó la base de datos de los censos de infraestructuras de salud de 1992, 1996 y 2000. Puntuaciones Z. Análisis econométricos	Se encontró que el efecto del programa sólo fue significativo en las áreas urbanas.

Área	Título del programa	Año / Institución evaluadora/autores	Muestra	Metodología (diseño, instrumentos y análisis)	Conclusiones
		Biology, 2004 Journal		Análisis de regresión	
Trabajo	Labor Market Reforms and Their Impact on Formal Labor Demand and Job Market Turnover: the case of Peru	Saavedra J. & Torero M. (2000) GRADE	400 establecimientos	Se utilizó un conjunto de información pseudo panel, de observaciones bimestrales entre 1987 y 1997. También se utilizó la información de Estudios de Hogares en para Lima Metropolitana. Diferencia en la diferencia y análisis de regresión.	Luego de la reforma se observó una disminución de la ocupación, siendo mayor y significativa para el caso de los trabajadores asalariados formales respecto a los informales.
Nutrición	The “Glass of Milk” Subsidy Program and Malnutrition in Peru.	Stifel D. & Alderman H. (2003) Banco Mundial Institucional	Utilizaron datos secundarios de tres fuentes: Encuesta Nacional de Enaho (1998-2000) - INEI Encuesta Nacional de Hogares sobre Medición de Niveles de Vida (1994-1997) – CUANTO Encuesta de Salud y Demografía (1996-2000) - USAID	Grupo de comparación sobre la base de población pobre y no pobre, elaborado a posteriori. Análisis de regresión de los valores porcentuales de los distintos estadísticos a través de los años.	No se encontró impacto del programa
Educación	Evaluación de impacto de la asistencia docente sobre el rendimiento de los estudiantes	Cueto S, León J, Torero S & Deustua, J (2003) GRADE Institucional	Población de 317 escuelas de 5 departamentos (Ayacucho, Cusco, Piura, San Martín y Puno) y 1092 docentes. Se utilizó una muestra de 209 escuelas y 619 docentes. Para la EI se utilizó solamente 178 escuelas, 354 docentes.	Diseño cuasi-experimental (no hubo asignación aleatoria a los grupos de tratamiento y contraste). Se utilizó un registro de monitoreo. Registros de asistencia diaria para cada docente en el programa. Asimismo, pruebas de rendimiento de los estudiantes a inicios y fines del 2004 en las áreas de matemática y comprensión de lectura. Se utilizó emparejamiento (matching). Estadísticas de comparación. Regresión.	Se encontró impacto del programa

En general, los estudios presentados en la tabla anterior pueden ser considerados como los más robustos y con mayor rigurosidad metodológica realizados en el país, pues han utilizado un diseño cuasiexperimental, de diseño pretest-posttest, muchos de ellos con grupo control y con manejo de las amenazas de selección empleando las técnicas de emparejamiento (*matching*). La mayoría han colectado sus propios datos mediante encuestas estructuradas, el uso de indicadores objetivos obtenidos de bases de datos de la Encuesta Nacional de Hogares (ENAH), las Encuestas sobre niveles de Vida (ENNIV), entre otros. Finalmente, algunas de ellas han utilizado técnicas estadísticas modernas como las de Regresión (lineal, logística, logit), para ajustar sus modelos econométricos o los sesgos iniciales de selección.

En el punto 3.2.2.1 se presentará un análisis estadístico de los estudios resumidos en la tabla anterior, en conjunto con el análisis de los estudios de Evaluación de Impacto (EI) internacionales y de forma comparativa.

3.2. DESARROLLOS INTERNACIONALES

Existen diversas experiencias internacionales relacionadas a la elaboración y fundamentación de propuestas de evaluación de impacto de los programas sociales; sin embargo, estas propuestas están insertadas dentro de esquemas reales de evaluación, no son propuestas propiamente metodológicas. La gran mayoría de ellas provienen de Organismos Internacionales de Desarrollo, mientras que otras, más modestas y académicas, provienen de investigadores de las ciencias sociales⁴³.

En este subcapítulo se presenta, en primer lugar, una breve revisión de las experiencias gubernamentales más exitosas en evaluación de impacto de América Latina, para luego revisar sistemáticamente las experiencias de investigación de impacto publicadas en los principales *journals* internacionales y, luego, sintetizar algunas novedades metodológicas.

⁴³ En el ámbito académico la aproximación más antigua de evaluación de impacto quizá provenga de Tyler (1942), quien elabora un modelo de evaluación centrada hacia los objetivos, valorando la coincidencia entre los objetivos del programa y los resultados reales. La evaluación se considera un proceso recurrente: se reordenan los objetivos en función de los resultados. No se explicita un diseño concreto. Posteriormente Scriven (1974) presenta modelo evaluativo “orientado hacia el consumidor libre de metas”. El evaluador es un sustituto informado del consumidor. El evaluador investiga todos los efectos del programa independientemente de sus objetivos. Propone diseños experimentales, pre-experimentales, análisis de costes y método libre de metas. Fue recién con Campbell y Stanley (1966), Cook y Campbell (1979) cuando se propone la aplicación del método científico a la evaluación. Plantean que no hay diferencias metodológicas entre evaluación y método científico. Utilizan exclusivamente el método cuasi-experimental. Adicionalmente, Lee Cronbach (1982) propone una evaluación asentada en una planificación previsor y flexible de las actividades evaluativas. Presenta el modelo de los UTOS (Unidades, Tratamiento, Observaciones, Situaciones), con esquemas de análisis muy rigurosos. Actualmente la visión más popular es la presentada por Campbell & Stanley, con sus diversas modificaciones.

3.2.1. Experiencias institucionales

En el ámbito del desarrollo social latinoamericano, la aplicación y difusión de las evaluaciones de impacto han estado encabezadas por instituciones multilaterales como el Banco Mundial (BM), el Fondo Monetario Internacional (FMI), el Banco Internacional para la Reconstrucción y Fomento (BIRF) y el Banco Interamericano de Desarrollo (BID). Estas instituciones han venido impulsando su aplicación durante la última década y sus resultados han sido incorporados en el diseño de las políticas sectoriales, las estrategias de asistencia de país, el diseño de programas y proyectos de desarrollo y, en general, en todas las operaciones de crédito que financian (CONPES, 2002; Choksi, 1995).

Dentro de los **desarrollos gubernamentales latinoamericanos** más exitosos se pueden citar al programa mexicano “Oportunidades” y la “Evaluación de Programas Gubernamentales (EPG)” de Chile.

En efecto, el *Programa social mexicano “Progresas”* (hoy conocido como “Oportunidades”), tiene una de las evaluaciones más completas realizadas a un programa social Latinoamericano. Es el único programa mexicano cuya evaluación fue prevista desde su diseño, partiendo de un diagnóstico de la situación antes de su inicio, y medida con la aplicación de encuestas a hogares con múltiples evaluaciones posteriores, generando volúmenes de documentos (Ejm. Skoufias, 2001; Skoufias, 2003).

Aunque se desconoce el monto de los recursos destinados por el gobierno para su evaluación, puede asegurarse sin temor a cometer grandes errores, que éste debió ser muchas veces superior a lo que se destinaba a la evaluación de los programas sociales antes referidos, a cargo de la Secretaría de Desarrollo Social (Sedesol), que fluctuó alrededor de los \$200,000 a \$550,000 para el año 2002. El otro aspecto que cabe resaltar es que el gobierno no le confió la responsabilidad de su evaluación a los propios con-nacionales, sino que recurrió al empleo de expertos de organismos internacionales⁴⁴.

De esta manera, mientras las evaluaciones peruanas (por sus limitados recursos) emplean sólo técnicas de encuesta, con mínima representatividad, aplicadas a una muestra de beneficiarios, entrevistas a funcionarios, observación directa y cálculo de algunos indicadores socioeconómicos, la evaluación del PROGRESA permitió un diseño experimental⁴⁵, mucho más costoso, que garantiza el mayor rigor en la identificación de relaciones causales entre el programa y los

⁴⁴ Si bien en la evaluación inicial (1998-1999) participaron organizaciones mexicanas como la Universidad Iberoamericana, el Instituto Nacional de Salud Pública y la Escuela Nacional de Antropología e Historia, la responsabilidad del proceso estuvo a cargo del International Food Policy Research Institute (IFPRI) de Washington, D. C., bajo la dirección del doctor Emmanuel Skoufias, quien forma parte del BID.

⁴⁵ El diseño experimental tiene la ventaja de permitir observar la diferencia entre las mediciones antes-después del programa, comparada con la evolución sufrida por grupos de similares características iniciales a los que el programa no se aplicó pero que, por otro tipo de razones ajenas al programa mismo, podrían también haber evolucionado, favorable o desfavorablemente, lo que a su vez permite identificar el verdadero impacto provocado por las acciones evaluadas.

efectos identificados, con mediciones antes-después y discusión en grupos focalizados, además de otras técnicas tradicionales⁴⁶.

El diseño de la evaluación de PROGRESA hubiera requerido idealmente una selección aleatoria de beneficiarios y no beneficiarios (grupo de control), a título individual, pero como la rigurosidad es costosa, se optó por una selección aleatoria en el nivel de las localidades. Éstas fueron 506, compuestas por 302 en que se aplicó el programa y 186 en que esto no se hizo (grupo de control), e incluyeron a 24,077 hogares, correspondientes a 78% de los beneficiarios⁴⁷.

Otra experiencia gubernamental interesante es la EPG de Chile. A partir del año 1997 el Gobierno de Chile incorporó la evaluación Expost de programas públicos a través de la *Evaluación de Programas Gubernamentales (EPG)*. Su objetivo es disponer de información que apoye la gestión de los programas públicos y el análisis de resultados en el proceso de asignación de recursos públicos. Esta iniciativa forma parte de un Protocolo de Acuerdo firmado entre el Congreso Nacional y el Ministerio de Hacienda con motivo de la aprobación de la Ley de Presupuestos, dejándose la responsabilidad de su ejecución en esta última institución⁴⁸. Es importante mencionar que desde su creación se han evaluado 145 programas públicos, a través de esta línea de evaluación (División de Control de Gestión de Chile, 2003).

La Evaluación de Programas Gubernamentales (EPG) de Chile, cuenta con un Comité Interministerial que tiene por objeto asegurar que el desarrollo de las evaluaciones sea consistente con las políticas gubernamentales, que las conclusiones que surjan de este proceso sean conocidas por las instituciones que conforman el Comité y que se disponga de los apoyos técnicos y coordinaciones necesarias para el buen desarrollo del mismo⁴⁹. Las evaluaciones son realizadas por paneles evaluadores constituidos por tres profesionales externos al sector público, quienes tienen la responsabilidad de efectuar las evaluaciones de acuerdo a los procedimientos definidos por el Ministerio de Hacienda. La selección de los evaluadores se realiza cada año mediante un riguroso análisis técnico a partir de una convocatoria pública la que se incorpora en el Sistema de Información para la

⁴⁶ En efecto, desde ya debe dejarse claro que los elementos clave que garantizan una evaluación exitosa y rigurosa son: que la evaluación sea un componente programado desde el diseño del programa, que tenga objetivos claros, que se cuente con el apoyo del ámbito institucional responsable del mismo, que el diseño de la evaluación ofrezca mediciones creíbles, que cuente con un adecuado marco lógico de evaluación, presupuesto suficiente y, de preferencia, un diagnóstico previo a la implantación del programa.

⁴⁷ Skoufias (2001) señala que este importante tamaño muestral podría ser incluso insuficiente para medir el impacto en algunas áreas, lo que de inmediato refuerza las dudas que dejan las evaluaciones de otros programas sociales, comúnmente realizadas en el país, donde es difícil que la muestra supere 20% de los beneficiarios, independientemente de los problemas enfrentados para garantizar una selección aleatoria.

⁴⁸ En el Perú, mediante la instalación del Sistema Nacional de Inversión Pública, se pretende controlar el diseño, ejecución y evaluación de los Proyectos de Inversión. La última etapa (expost) es la referida a la evaluación de impacto de los programas. Hasta la fecha (diciembre de 2006) solamente se está implementando las estrategias de formulación y ejecución.

⁴⁹ Este Comité está conformado por un representante del Ministerio Secretaría General de la Presidencia, del Ministerio de Planificación y Cooperación (MIDEPLAN) y del Ministerio de Hacienda, a través de la Dirección de Presupuestos, siendo presidido por esta última cartera.

Selección de Consultores del Programa. El proceso de selección considera básicamente las competencias de los profesionales en el área de evaluación y en el área de intervención del programa.

Adicionalmente, para la conformación de los paneles de evaluación se utiliza como criterio de selección la complementariedad de especializaciones entre los profesionales que postulan a integrar los distintos paneles. Por su parte, las instituciones responsables de los programas evaluados se incorporan al proceso de evaluación, a través de las siguientes actividades:

1. Proveer de un conjunto de información del programa para su entrega al panel evaluador, al inicio del proceso, y de cualquier otra información disponible y necesaria durante la evaluación.
2. Elaborar la matriz de marco lógico preliminar, para su entrega al panel evaluador como insumo base.
3. Participar en todas aquellas reuniones que solicite el panel evaluador y el Ministerio de Hacienda para analizar aspectos específicos del proceso de evaluación.
4. Analizar y emitir comentarios a los resultados intermedios y finales de la evaluación, los que son entregados al panel de evaluación a través del Ministerio de Hacienda.
5. Analizar y emitir comentarios a las recomendaciones, además de participar en el proceso de establecimiento de compromisos en torno al mejoramiento de los programas.
6. Informar del cumplimiento de los compromisos institucionales.

Dentro del EPG de Chile podemos apreciar un caso: El programa “Chile Joven”. Su evaluación de impacto arrojó conclusiones importantes, no sólo en términos de los efectos directos sobre la población beneficiaria, sino también de los procesos y la implementación del mismo. Esta experiencia ha servido como ejemplo para el diseño y la puesta en marcha del programa “Jóvenes en Acción” de la Red de Apoyo Social (RAS) en Colombia.

Por otro lado, en el **ámbito institucional**, el Banco Mundial es la organización internacional que más ha incidido en el desarrollo y la difusión de la evaluación de impacto en América Latina. El Banco Mundial ha publicado importantes documentos sobre la evaluación de impacto, como la obra de consulta de Baker (2000), del Departamento de Evaluación de Operaciones (2004), de Ravallion (2001, 2005) y, recientemente, de Bemberger (2006), de Bedi, Bhatti, Gine, Galasso, Goldstein & Legovini (2006) y White (2006). Además, de sus directivas metodológicas, es de resaltar que el Banco Mundial tiene casi una centena de estudios de evaluaciones de impacto divulgados a través del sitio web del Banco *Poverty Impact Analysis, Monitoring and Evaluation*. Otra institución que cuenta con una base de datos sobre artículos de investigación e informes institucionales sobre evaluación de impacto de programas sociales es el Banco Interamericano de Desarrollo (BID).

Un análisis individual de estos informes se presenta en la Tabla N° 3.3, conjuntamente con el análisis individual de otras fuentes institucionales y de los artículos de evaluación de impacto publicado en revistas internacionales.

3.2.2. Desarrollo académico-institucional

En el ámbito internacional se han desarrollado diversas investigaciones de impacto utilizando metodologías modernas y de base cuasiexperimental. Estas evaluaciones, publicadas en revistas internacionales especializadas en evaluación de programas (véase la Tabla 3.1) o ubicadas en base de datos de instituciones internacionales para el desarrollo (Ej. Banco Mundial) han aumentado significativamente en la última década, por lo que su revisión y análisis resulta necesario.

Por ello, en la siguiente tabla se resumen los principales estudios que introducen los métodos basados en diseños cuasi-experimentales que son aquellos en los que el avance en la rigurosidad ha sido mayor en los últimos años para la evaluación de programas.

Tabla N° 3.3. Desarrollo académico de la evaluación de impacto de programas sociales en el mundo según investigaciones publicadas en Journals internacionales o Instituciones de Desarrollo (Fuente: Elaboración propia).

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Suministro de agua y sanidad	A Transitory Regime: Water Supply in Conakry, Guinea	Menard, C.; George, C.R. (2000) Guinea Institucional	No disponible	Diferencia de medias, diferencia en la diferencias y simulación contrafactual.	El gobierno obtuvo una ganancia de 9.8 millones como resultado del régimen transitorio.
Salud, nutrición & población	Worms: Education and Health Externalities in Kenya	Miguel, Edward, and Michael Kremer. (2001) Kenya Institucional	75 escuelas primarias divididas aleatoriamente en tres grupos. De estos dos fueron de tratamiento y uno de control.	Se utilizó información secundaria obtenida por el equipo de evaluación sobre características de la población y de los estudiantes (1998 y 1999), y la investigación parasitológica conducida por el Ministerio de Salud de Kenya (1998) Los tres grupos cumplieron el rol de grupo de tratamiento y control de manera indistinta, según la etapa del programa.	No se encontró evidencia que la desparasitación intestinal influye en las puntuaciones académicas.
Educación	Teacher Incentives	Glewwe, Paul; Nauman, Ilias; Kremer, Michael (2003) Kenya Institucional	Muestra de 100 escuelas: 50 en el grupo de tratamiento y 50 en el grupo control.	Diseño de evaluación aleatoria. Metodología <i>Intention to Treat</i> Diferencia en la Diferencia Análisis de regresión de mínimos cuadrados.	Los estudiantes en las escuelas con un programa de incentivo al profesor tuvieron más probabilidad de tomar exámenes y obtener puntuaciones promedio a corto plazo. Sin embargo, Existe poca evidencia que los profesores respondan al programa adoptados medidas para reducir la deserción escolar o incrementar esfuerzos para estimular el aprendizaje a largo plazo.
Educación	Textbooks and Test Scores: Evidence from a	Glewwe, Paul; Kremer, Michael; Moulin,	25 escuelas primarias que fueron elegidas aleatoriamente	Diferencia en la diferencia.	No existe evidencia que el programa incremente la

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Prospective Evaluation in Kenya	Sylvie Kenya (2000) Institucional	de un grupo de 100 escuelas candidatas. Se estableció comparación de grupos según diversas modalidades de recibimiento del programa.	Análisis de regresión para la comparación de medias. Generalized Least Squares	puntuación promedio de las evaluaciones, o que reduzca el nivel de repitencia o tasa de deserción escolar.
Salud, nutrición y población	The Effects of a Fee-Waiver Program on Health Care Utilization Among the Poor: Evidence from Armenia	Chaudhury, Nazmul, Jeffrey Hammer, and Edmundo Murrugarra (2003) Armenia Institucional	La información se obtuvo de las muestras de dos estudios: The American Living Standard Survey (4,260 familias) y The Integrated Survey of Living Standards (3,600 familias). Diferenciándose grupo de tratamiento y grupo control	Diferencia en la diferencia y otros métodos de emparejamiento.	El programa fue inadecuado para disminuir la baja utilización de servicios en los beneficiarios.
Educación y empleo	The Growth of social work education programs, 1985-199: Its impact on economic and educational factors related to the profession of social work	Karger, H. & Stoesz, D. (2003) Estados Unidos Journal	Información de los 432 programas de bachillerado ofrecidos por Council on Social Work Education (CSWE)	Análisis de mercado laboral. Estudio con fuentes secundarias.	Los autores concluyen que un excedente de programas de educación en trabajo social ha sido perjudicial para el potencial ingreso de trabajadores sociales de nivel básico, para la calidad de la educación del trabajo social, y para las competencias de los graduados en trabajo social.
Pobreza	The Impact of an Integrated Family Planning Program in Russia	Sherwood-Fabre L, Goldberg H. & Bodrova V. (2002) Rusia Journal	6000 mujeres provenientes de tres ciudades rusas.	Evaluación cuasi experimental. Se tomaron dos ciudades para el grupo de tratamiento y una como grupo control.	Los resultados indican qué influyó en el conocimiento de las mujeres acerca de los métodos de planificación familiar, haciendo que ellas tengan una actitud más favorable hacia los anticonceptivos modernos. Asimismo, las tasas de aborto decrecieron
Pobreza	The Impact of Cash Transfers on Child Labor and School Attendance in Brazil	Cardoso, Eliana; Portela Souza, Andre (2004) Brasil	428740 familias Niños de 10 a 15 años	Se utilizó los datos del Censo 2000 Diferencia de medias Propensity Score Matching.	La transferencia de ingreso generó un incremento positivo y significativo de 3% en la asistencia escolar de los niños y niñas. Sin embargo, el

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional			programa no tuvo impacto en el trabajo infantil
Empleo	The Impact of Microcredit Programs on Self-employment Profits: Do Noncredit Program Aspects Matter?	McKernan, SM. (2002) Bangladesh Journal	1757 jefes de familia que participan o no participan en el Grameen Bank y de dos programas de micro-crédito similares	Se utilizó ecuaciones de regresión. El efecto total es medido al estimar una ecuación de perfil y el efecto de la falta de crédito es medido al estimar la ecuación de perfil condicional sobre el capital producido.	Se encontró grandes efectos positivos de la participación y de los aspectos de la falta de participación sobre los perfiles del auto-empleo.
Empleo	The Impact of Minimum Wages on Employment in a Low Income Country: An Evaluation using the Difference-in-Differences Approach	Alatas V. & Cameron L. (2003) Indonesia Institucional	1224 empresas con más de 20 empleados, que constituye el total de la población.	Se utilizó un enfoque de la diferencia en la diferencia. Diseño comparación con grupo control. Emparejamiento por Matching Los datos fueron tomados de la Annual Survey of Manufacturing Firms, acerca de la situación de las empresas durante el período de 1990 a 1996.	No se encontró evidencia de que el incremento en el salario mínimo reduce el empleo en las empresas grandes extranjeras o domésticas. Sin embargo, si se aprecia lo opuesto para el caso de las empresas pequeñas y domésticas. Estas estimaciones se realizaron de la información de un período comprendido en pleno crecimiento sostenido.
Salud	The impact of programs for high-risk drinker on population levels of alcohol problem	Smart RG. & Mann R.E. (2000) Canadá Journal	Estudios experimentales y cuasiexperimentales. No menciona tamaño muestral.	Estudio metanalítico sobre la evidencia correlacional del impacto de los programas de alto riesgo sobre las consecuencias finales del problema.	Los programas para los bebedores de alto riesgo pueden tener efectos de nivel agregado benéficos y son de esta manera un importante componente de los esfuerzos basados en la población para reducir los problemas de alcoholismo.
Pobreza	The Impact of R&D on the Singapore Economy: An Empirical Evaluation	Ho YP, Wong MH. & Toh MH. (2005) Singapur	Se utilizaron los gastos hechos en Singapur entre los años 1978 y 2001.	Análisis de impacto basado en Cobb-Douglas. Se utilizó el análisis de regresión.	Se encontró evidencia respecto al impacto significativo que tuvo la inversión R&D sobre el

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional			desempeño de su factor total de productividad en los últimos 20 años. Pero este impacto es débil si lo comparamos con otros países-
Educación	Back to School on Impact Evaluations: A second evaluation of farmer field schools (FFS) in Indonesia	Farchy, D. (2005) Indonesia Institucional	Se utilizó una muestra de 268 jefes de familia. Estos estuvieron divididos en tres grupos: aquellos que se graduaron en una FFS (112), los que estuvieron expuestos a la enseñanza de los graduados en una FFS (156) y los que conformaron el grupo control (52).	Se utilizó el estimador de diferencia en la diferencia para comparar la diferencia de resultados en los grupos.	Los análisis revelan no sólo los efectos directos sobre los graduados en el programa sino también hubo efectos secundarios sobre aquellos que estuvieron expuestos a la enseñanza del programa en comparación con el grupo control.
Educación	The impact of school mapping in the development of education in Tanzania: An assessment of the experiences of six districts	Galabawa J CJ, Obeleagu A, & Miyazawa I. (2002) Tanzania (Africa) Journal	El estudio fue desarrollado en seis distritos en los cuales los ejercicios de mapeo escolar fueron iniciados. Se evaluó a un total de 240 personas vinculadas al planeamiento educacional de los distritos fueron evaluadas (40 por distrito).	Para la recolección de los datos se utilizaron los siguientes instrumentos: Formulario de mapeo escolar, formato de entrevista, grupos de discusión focal y cuestionarios diseñados. La confiabilidad de los instrumentos se analizó a través de la fórmula Kuder Richardson.	Se encontró que el mapeo escolar impacta de manera positiva y en diferentes magnitudes sobre el desarrollo de la educación en términos de incremento de matrículas y asistencia., disminución del abandono escolar, promoción de la información para la toma de decisiones, y el incremento de las capacidades del campo de los actores para planificar y tomar la acción.
Alimentarios	The Impact of the Bolsa Alimentacao Program on Food Consumption	Olinto, Pedro; Flores, Rafael; Morris, Saul; Veiga, Alinne (2003) Brasil Institucional	696 beneficiarios y 309 no beneficiarios.	Diferencia de medias, variables instrumentales, otros métodos de emparejamiento.	Los resultados muestran que los hogares beneficiarios exhiben un mayor consumo estadísticamente significativo de diversos alimentos. Es más, el programa tiene un impacto positivo y estadísticamente significativo sobre la diversidad de las dietas.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Justicia	The Impact of the Social Security Program on Private Pension Contributions	Rejda G. & Schmidt J. (1979) Estados Unidos Journal	El período de tiempo bajo investigación es desde el año de 1950 a 1975.	Programa Old-Age, Survivors, Disability, and Health Insurance (OASDHI). Modelos de regresión para los dos tipos de planes. Se realizaron análisis de relación funcional entre las contribuciones de pensión respectiva.	El incremento de los impuestos OASDHI no parece disminuir las contribuciones hacia planes con ausencia de seguro. Los poderes de precio del monopolio permiten que las grandes compañías pasen tanto mayores impuestos OASDHI y costos de pensiones hacia el público a través de mayores costos.
Educación	The Oregon First Step to Success Replication Initiative: Statewide Results of an Evaluation of the Program's Impact.	Walker HM, Golly A, Zolna J, & Kimmich M. (2005) Estados Unidos Journal	La muestra está conformada por 181 niños (grupo experimental) que cursan el segundo grado de nivel inicial en 11 de los 36 condados de Oregon. Se pidió a los profesores de los alumnos que seleccionaran a aquellos alumnos que presentaban un significativo problema en el ajuste de su comportamiento social. Específicamente en las relaciones alumno-estudiante o estudiante-compañeros. El grupo control estuvo constituido por aquellos niños que calificarían para ser admitidos en el programa pero no los fueron (30 niños).	Se utilizó un diseño pre-post. Dos clases de mediciones son informadas. La primera, que incluye las tres escalas de reporte de los maestros, fue diseñada para medir los cambios de comportamiento pre-post relacionados con la intervención del programa. El segundo tipo de medición permitió conocer la satisfacción de los padres y maestros con el programa así como de la fidelidad de los procesos de implementación.	Tanto padres como maestros consideran efectos colaterales del programa en el contexto familiar y el salón de clase. Numerosas limitaciones en la evaluación fueron resultado de problemas logísticos, dificultades en la recolección de información, el uso de profesores como evaluadores e interventores.
Salud	The Reach and Impact of Social Marketing and Reproductive Health	Meekers D. & Van Rossem R. (2004)	6,782 mujeres y 1,928 varones.	Los análisis se basaron en la información de 2001 a 2002 del Zambia Demographic and Health	Los hallazgos indican que las campañas de salud reproductiva y mercado social

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Communication Campaigns in Zambia	Zambia Institucional		Survey. Para controlar la auto selección y endogeneidad, se utilizó un modelo de regresión de dos etapas para estimar el efecto de la exposición del programa sobre los comportamientos resultantes	en Zambia alcanzaron un gran porcentaje de la población y tuvo un significativo impacto sobre las discusiones de planificación familiar y uso del condón.
Educación	The social impact of Life Education: estimating drug use prevalence among Victorian primary school students and the statewide effect of the Life Education programme	Hawthorne, G. (1996) Australia Journal	1721 estudiantes en el programa <i>Life Education</i> y 1298 estudiantes que no participan en el programa.	Estudio exploratorio a través del re-análisis de la información de Hawthorne (1993 y 1995).	Los resultados indican que los programas de intervención deberían ser evaluados exhaustivamente antes de extender su aplicación, y dichas evaluaciones deberían considerarse en el estudio del impacto que tienen los programas a lo largo de toda la comunidad.
Educación	The tribes process TLC: A preliminary evaluation of classroom implementation & Impact on student achievement	Kiger, D. (2000) (Estados Unidos - Canadá) Journal	Se utilizaron tres salones de estudiantes: Un salón donde el proceso Tribes está completamente implementado (N=37), otro donde el proceso está parcialmente implementado (n=29) y un salón donde el proceso no está implementado (n = 66)	Diseño pretest-postest Para el análisis de los datos se utilizó ANCOVA	La comparación por grupo mostró que los alumnos que integran el proceso Tribes (M= 54.26) demostraron un desempeño significativamente mayor que los alumnos del proceso triba Parcial (M=48.91) y que no están en el proceso (M= 50.47)
Suministro de agua y sanidad	The Welfare Effects of Private Sector Participation in Guinea's Urban Water Supply	Menard, C.; George, G.; Zuluaga, A. (2000) Guinea Institucional	Grupo control y experimental.	Diferencia de medias Simulación contrafactual	El gobierno obtuvo ganancias de 9.3 millones de dolares.
Educación	Updating the Economic Impacts of the High/Scope Perry Preschool Program.	Nores M, Belfield, CR, Steven W, Schweinhart (2005)	Se realizó una selección aleatoria de los niños para integrar el programa (119) y el grupo control.	La información sobre las diferencias de ingreso de los participantes está disponible ahora que ellos tienen una edad de 40	Los datos indican sólidas ventajas para el grupo de tratamiento en términos de mayores ingresos de vida y

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Estados Unidos Journal		años. Estos ingresos son traducidos en términos monetarios y se los compara con el costo de ejecutar el programa para calcular el valor neto del programa tanto para los participantes como para la sociedad.	menor actividad criminal.
Pobreza (Trabajo infantil)	Use of Survey Design for the Evaluation of Social Programs: The PNAD and the Program for the Eradication of Child Labor in Brazil	Pianto, Donald M., and Sergei Soares. (2004) Brasil Institucional	800 municipalidades	Datos del Brazilian Nacional Household Sample Survey (1997-1999) y del Ministry of Social Welfare. Diferencia en la diferencia y Propensity Score Matching	Para las municipalidades que ingresaron en 1998 los resultados muestran un significativo incremento per cápita. De igual manera, el impacto es mayor en los municipios menores.
Educación	Bangladesh Food for Education Program: An Evaluation of its Impact on Educational Attainment and Food Security	Ahmed, Akhter; Del Ninno, Carlo (2002) Bangladesh Institucional	930 niños con edad escolar primaria.	VARIABLES instrumentales de dos etapas y otros métodos de emparejamiento.	El programa tuvo un efecto positivo y significativo sobre la matrícula escolar; asimismo, las mujeres tienen una mayor probabilidad de matricularse que los varones.
Educación	Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment	Angrist, Joshua D.; Bettinger, Eric; Bloom, Erik; King, Elizabeth; Kremer, Michael (2002) Colombia Journal	Selección aleatoria del grupo de tratamiento y grupo control (por sorteo).	Entrevista a través del teléfono. Diferencia de medias	El programa tuvo un efecto positivo y significativo en la probabilidad de recibir una educación privada y en el número de años completos de escolaridad y sobre el gasto escolar.
Suministro de Agua y sanidad	Water For Life: The Impact of the Privatization of Water Services on Child Mortality	Galiani, Sebastian; Gertler, Paul; Schargrodsky, Ernesto (2002) Argentina Institucional	Muestra aleatoria estratificada de 40000 hogares de los municipios urbanos con más de 5000 habitantes.	Información del Censo de 1994 y de la Encuesta de Desarrollo Social (1997) Método de evaluación Diferencia en la Diferencia y la Propensity Score Matching	Existe una relación estadística negativa entre la privatización y mortalidad infantil.
Calidad de vida	A cluster evaluation of	Schwerina MJ,	Las selección de la muestra fue	Diseño de evaluación solo postest.	Se encontró que los

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Navy quality of life programs	Michaelb PG, Glaserc DN. & Farrar KL. (2002) Estados Unidos Journal	intencional de 10765 beneficiarios de programas	Se utilizaron mediciones por auto reporte Se utilizó análisis de correlaciones y la prueba Chi cuadrada.	programas tienen impacto en los miembros de servicio militar y sus familias.
Empleo	WorkFirst Longitudinal Study	Klawitter, MM. (2002) Estados Unidos Institucional	Se utilizó una muestra de 1330 personas	Estudio de cohorte de un período de un año. Se utilizó información los archivos estatales Unemployment Insurance. Se entrevistó a los sujetos a través de la línea telefónica. Se midió el efecto de: Taller para Buscar Trabajo, Capacitación en Habilidades Laborales, Capacitación Pre-empleo y Trabajos Comunitarios, a través de análisis multivariados.	Se encontró que el Taller para buscar trabajo y la capacitación pre-empleo incrementarían las ganancias y posibilidades de empleo en los participantes
Calidad de vida	A multimethod approach to evaluate transition into community life	Gliner JA. & Sample P. (1996) Journal Estados Unidos	35 personas adultas con retardo en el desarrollo participaron del programa.	Se utilizó la escala Índice sobre la Calidad de Vida. El estudio de caso de 3 participantes; dos varones y una mujer Se realizaron análisis factorial de la varianza	El Índice de Calidad de Vida falló en mostrar un efecto estadísticamente significativo del programa para los beneficiarios. Aunque la evaluación subjetiva y los métodos de estudio de caso indicaron lo contrario.
Sistema de pensiones - justicia	A State without ownership: the welfare impact of British privatizations 1979-1997	Florio, M (2002) Reino Unido Institucional	Se analizó el impacto en las empresas, empleados, beneficiarios, consumidores y contribuyentes.	Se realizó una evaluación a través del enfoque costo-beneficio.	Las privatizaciones británicas tuvieron efectos modestos sobre la eficiencia de producción y consumo. Sin embargo, tuvieron importantes efectos en la distribución del

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
					ingreso y salud.
Violencia -Salud	Achieving social change on gender-based violence: A report on the impact evaluation of Soul City's fourth series	Usdin S, Scheepers E, Goldstein S. & Japhet G. (2005) Sudafrica Journal	No se especifica.	Con el fin de evaluar el impacto en tres niveles se siguió un diseño multifacético, que comprende a seis estudios interrelacionados: - Un estudio nacional. - Una evaluación nacional del impacto. - Evaluación de la relación entre el programa SC4 y la Nacional Network on Violence Against Women. Se utilizó un diseño cualitativo y cuantitativo.	La evaluación mostró una asociación consistente entre el Programa Soul City Fourth y el cambio positivo acerca de la violencia de género.
Pobreza - migración	Aid and Migration: An Analysis of the Impact of Progresa on the Timing and Size of Labour Migration	Angelucci M (2004) México Institucional	10000 Familias rurales pobres compuestas por 27000 individuos. De estos 17000 integraron el grupo de tratamiento y 10000 el grupo control	La información se obtuvo de dos estudios de migración (1997 y 1999) Diseño post test. Se utilizó la técnica de aleatorización. Los datos fueron se sometieron al análisis de regresión.	Se encontró que el programa esta asociado con un incremento en el promedio de migración internacional
Salud	An impact evaluation of a falls prevention program among older people	Deery HA, Day LM. & Fildes BN. (2000) Autralia Journal	Se seleccionó a de manera aleatoria a 223 personas mayores de 60 años. 107 beneficiarios del programa y 116 participantes como grupo control.	Diseño de evaluación no aleatorio pre- post. Se utilizó la prueba Chi Cuadrada, la regresión logística y el análisis de covarianza.	Se encontró que el programa tuvo un impacto positivo en el conocimiento de las caídas y su prevención.
Educación - discriminación	Attitude change amongst nursing students towards Australian Aborigines	Hayes L, Quine S. & Bush J. (1994) Australia	La muestra fue de 246 estudiantes.	Diseño de evaluación de impacto pre test y post test. Se utilizó un cuestionario de	Se encontró que el programa fomentó las actitudes favorables hacia los niños australianos aborígenes.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Journal		actitudes auto-administrado. Los datos se analizaron con la prueba de rangos de Wilcoxon, la prueba T de Student para muestras dependientes y los análisis de regresión múltiple.	
Educación, nutrición & población	Baseline Report of the Evaluation of Familias en Accion	Attanasio, Orazio (2004) Colombia Institucional	Se seleccionó de manera intencional dos conjuntos de municipalidades para el grupo de tratamiento y dos conjuntos para el grupo control. Cada conjunto de municipalidades estuvo integrado por 25 municipalidades. Se realizó una selección aleatoria por estratos, de las familias elegibles para el programa. De estas, 10660 integraron el grupo de tratamiento y 8347 el grupo control.	Se siguió un método cuasi-experimental. La información es de tipo secundaria, obtenida de un estudio de familias, estudios de centros de escuela y salud, y un estudio de comunidades (2002) Diferencia de medias Propensity Score Matching para emparejar el grupo de tratamiento y de control.	El programa tuvo un efecto positivo significativo en la matrícula escolar de los niños, existe diferencias significativas entre las municipales en tratamiento y las de control respecto al número de días que los niños de 2 a 6 años consumen huevo. Asimismo, el programa tuvo un efecto positivo y significativo en el peso y altura de los niños de las áreas rurales.
Salud	Benefits and costs of prevention and early intervention programs for youth	Aos S, Lieb R, Mayfield J, Miller M & Pennucci A. (2004) Estados Unidos Institucional	Se seleccionó aquellos programas que presentaron al menos una evaluación rigurosa.	Se realizó una búsqueda electrónica sobre las evaluaciones de programas. Se construyó un modelo costo beneficio para medir el valor monetario de los resultados obtenidos en los diferentes programas	Se encontró que algunos programas de prevención e intervención temprana pueden lograr significativamente más beneficios que costos.
Microcrédito	Clients in context: The impacts of microfinance in three countries.	Snodgrass, DR. & Sebstad, J. (2002) Multinacional. Peru, India & Zimbabwe	Se estudiaron a tres instituciones: SEWA Bank en India; Accion Comunitaria del Peru/ Mibanco en Peru; y Zambuko Trust en Zimbabwe.	Análisis de casos. Estudio cualitativo.	Se confirmó que el impacto de las microfinanzas en un nivel familiar es condicional y heterogéneo. Existe impacto positivo en

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional (USAID)	Sin muestreo. Se analizaron 3 estudios individuales, uno en cada institución. Datos secundarios.		algunas variables entre algunos grupos e impactos neutros en otras.
Infraestructura	Cost-Benefit Analysis of the High-Speed Train in Spain	De Rus, G. & Inglada, V. (1997) España Journal	No específica.	Evaluación ex post con un análisis de costo-beneficio.	Se encontró una tasa de rebaja social de 6% en términos reales. La evaluación económica del proyecto de tren rápido muestra que este no debería haber sido realizado en 1987 en el corredor Madrid-Sevilla.
Educación	Cost-Effectiveness of Education Policies in Latin America: A Survey of Expert Opinion	Schiefelbein, Wolff L. & Schiefelbein, P. (1998) Latinoamérica. Institucional	10 expertos internacionales y 30 planificadores o profesionales.	Se hizo una evolución de Sondeo. Método Delphi Se aplicó una escala de 40 ítems.	Existe un razonable consenso entre los expertos internacionales respecto al relativo mérito de las intervenciones particulares, pero existe menos consenso acerca de la medida de su impacto. Mientras que los profesionales tienen un mayor optimismo.
Justicia - empleo	Detecting Effects of Living Wage Laws	David Neumark & Scott Adams (2003) Estados Unidos Journal	13000 familias, se determinó grupo de tratamiento y grupo control.	La información utilizada proviene de los archivos CPS Outgoing Rotation Group de 1996 a 2000. Se utilizó la estrategia de Diferencia en la Diferencia para comparar las ciudades con tratamiento de las ciudades control.	Se encontró considerables efectos salariales positivos en ciudades con extensas normas salariales que cubren la asistencia de negocios recibida por los empleados de la ciudad. También se encontró evidencia de sacrificio entre sueldo y empleo como efecto de las normas de salario vital.
Salud	Do Health Sector Reforms Have Their Intended Impacts? The World	Wastaff, A. & Shengchao, Y. (2005)	Se estudio a tres provincias beneficiarias y 17 no beneficiarias del proyecto.	Diseño de evaluación pre post. Se utilizó la información de	Los resultados indican que el proyecto Health VIII ha sido exitoso en casi todas sus

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Bank's Health VIII Project in Gansu Province, China	China Institucional	La muestra fue de 1116 personas que viven en las tres provincias beneficiarias y 6465 personas que viven en las provincias no beneficiarias. Se determinó grupo de tratamiento y grupo control.	Gansu Survey of Children and Familias (GSCF) Se utilizaron los siguientes estimadores de comparación: diferencia simple, diferencia en la diferencia, ajuste de heterogeneidad a través de los métodos de regresión y emparejamiento.	metas.
Pobreza	Do we know what works? A Systematic Review of Impact Evaluations of Social Programs in Latin America and the Caribbean	Bouillon CP. & Tejerina L. (2006) Multinacional en América Latina: (Argentina, Bolivia, Chile, Brasil, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, México, Nicaragua y Perú). Institucional (BID)	51 estudios de 47 programas sociales. Recolección intencional de informes.	Meta-análisis. Analiza programas que incrementa la capacidad de los pobres y el incremento de las oportunidades económicas de los pobres. Análisis descriptivo a nivel de frecuencia.	Se ha encontrado un impacto promedio positivo en la región. Se concluye que la EI de los programas muestra que estos pueden ser herramientas muy efectivas para reducir la pobreza y la desigualdad a largo plazo y para aliviar la pobreza a corto plazo.
Empleo	Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification	Lechner, M. (1999) Alemania Journal	Personas que estuvieron trabajando antes de la unificación alemana.	Se analizó la información proveniente del estudio Socio-Economic Panel (GSOEP, 1990-1994) Se utilizó como metodología de evaluación empírica el enfoque de resultados potenciales para la causalidad.	Se halló que al menos en el corto plazo no hay efectos positivos del programa.
Agricultura	Economic Analysis of Crossbreeding Programmes in Sub-Saharan Africa: A Conceptual Framework	Karugia JT, Mwai OA, Kaitho R, Drucker AG, Wollny CBA. & Rege JEO, (2001)	No se menciona.	Se utilizaron dos modelos de evaluación: el modelo del sector agrícola para las diversas mediciones de impacto de la cría de ganado y el modelo Nivel de	Los programas tuvieron un impacto positivo en el bienestar de la sociedad.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	and Kenyan Case Study	Kenya Institucional		Ingreso Familiar y Simulación Política.	
Justicia	Economic evaluation of drug court: methodology, results, and policy implications	Logan TK, Hoyt W., McCollister K, French M, Leukefeld C, & Minton L. (2004) Estados Unidos Journal	La muestra estuvo constituida por los graduados (n= 222), los finalistas (n= 731) o los candidatos a los Programas de Juzgado de Drogas de los años 1996 a 1998.	Se utilizó información proveniente de la base de datos sobre encarcelación Ofender Records Information and Operations Network, Kentucky Department of Mental Health and Mental Retardation, Department of Motor Vehilces, y otros. Se realizó un análisis de costo económico	El estudio encontró que, particularmente para el caso de los graduados, la inclusión al Juzgado de Drogas estuvo asociada con la disminución en la encarcelación, los servicios de salud mental, y los costos legales, asimismo con el incremento en el ingreso y pagos de apoyo infantil.
Suministro de Agua y sanidad	Buenos Aires Water Concession	Alcazar, Lorena; Manuel Abdala, and Mary Shirley (2000) Argentina Institucional	Se comparó los beneficios brindados a los consumidores, trabajadores, directivos y compradores con un grupo contrafactual (1987-1998)	Método de evaluación de diferencia de medias y simulación contrafactual.	No se registró niveles de ganancias significativos.
Pobreza	Economic Motors for Poverty Reduction in Madagascar	Dorosh P, Haggblade S, Lungren C, Razafimanantena T. & Randrianmiarana Z.(2003) Madagascar Institucional	No se menciona.	Se utilizó una Matriz de Evaluación Social para evaluar el impacto de los cuatro motores para el bienestar de las familias pobres (Agricultura, construcción de carreteras, inversión privada para la exportación y turismo). Se utilizó modelos de simulación.	Los cuatro motores evaluados generan resultados diferentes. De esta manera, cada motor juega un rol diferente en la batalla nacional contra la pobreza.
Salud	Effects of mental health reform on client characteristics, continuity of care and community tenure.	Brown L. et al. (1994) Estados Unidos Journal	Se consideró a toda la población de beneficiarios.	Se utilizaron los datos de los Sistemas de Información de Mental Health Division. Todos los análisis de los datos fueron de tipo descriptivo.	En general no hubo indicadores de deterioro del servicio, más bien se encontró que algunos servicios la atención se favoreció.
Educación	Electronic Course Evaluation; Does an on-	Avery RJ, Bryant WK, Mathios A, Kang H,	La muestra fue de 31 cursos, cuatro de estos formaron parte	Se utilizó información histórica específica sobre la evaluación de	Los métodos de evaluación por web llevarían a una menor

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	line delivery system bias student evaluation?	Bell B; (2006) Estados Unidos Journal	de un experimento de muestra emparejada La información individual corresponde a 972 estudiantes.	curso de un programa de políticas públicas (1999-2001) Se utilizó una escala en dos modalidades: escrita y en línea web	tasa de respuesta, pero esto no afectaría el promedio de las puntuaciones de la evaluación.
Empleo - justicia	Employment of People with Disabilities Following the ADA	Kruse, D. & Schur, L. (2003) Estados Unidos Journal	Personas discapacitadas de 21 a 58 años de edad.	Se utilizó la información proveniente del Survey of Income and Program Participation (SIPP) Las tendencias relativas de empleo se evaluaron a través de un enfoque de Diferencias en las Diferencias.	Hay razones para ser cuidadosos respecto a los hallazgos ya sea de los efectos positivos o negativos debido a las limitaciones de mediciones para determinar a quiénes se extiende el de American Disability Act.
Educación - salud	Estimating causal effects of public health education campaigns using propensity score methodology	Yanovitzkya I, Zanuttob E. & Hornik R (2005) Estados Unidos Journal	3,184 de individuos expuestos a la campaña y 807 no expuestos a la campaña	Se utilizó el modelo de comparación Propensity Store. Asimismo, se utilizó el Análisis de Varianza	Se encontró un significativo aunque pequeño efecto de la exposición a la campaña sobre la conversación de los padres con los hijos sobre asuntos de drogas.
Justicia - empleo	Evaluating the economic effects of a new state-funded school building program: the prevailing wage issue	Greenberg M, et al. (2005) Estados Unidos. Journal	No se mencionó.	Se utilizó modelos de ingreso y egreso así como econométrico para cuantificar los impactos económicos de las leyes de salarios dominantes en New Jersey. Para tal fin se siguió un modelo de simulación.	Se halló impacto del programa.
Empleo	Evaluating the Employment Impact of a Mandatory Job Search programme	Blundell R, Costa Dias M, Meghir C, Van Reenen J. (2003) Reino Unido Institucional	No disponible	Se realizó un piloteo basado en áreas y reglas de selección relacionados con la edad para identificar el efecto del programa de mercado laboral.	Se encontró que el programa incrementó significativamente la transición hacia el empleo en cerca de cinco puntos porcentuales. El impacto es robusto para una amplia variedad de estimadores no experimentales.
Educación - salud	Evaluation of HIV/AIDS	Hughes-d'Aeth, A.	Jóvenes de ambos sexos de a	Se utilizó un modelo de	Los proyectos tienen un efecto

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	peer education projects in Zambia	(2002) Zambia Journal	los 15 años.	evaluación de estudio de casos. La información fue recopilada a través de: revisión de documentos, entrevista a informantes clave y visitas de observación. Se realizaron análisis individuales y comparativos de los estudios de caso.	favorable para la educación en SIDA.
Salud	Evaluation of Life Skills Training and Infused-Life Skills Training in a rural setting: Outcomes at two years	Smith EA., Swisher JD, & Vicary JR. (2004) Estados Unidos Journal	Se eligieron 9 centros educativos y se distribuyeron de manera aleatoria en los siguientes grupos: LST, I-LST y grupo control. La muestra fue de 732 estudiantes.	Se utilizó una evaluación prospectiva con diseño de grupos aleatorios- Se realizó 5 observaciones de la muestra. Los datos fueron sometidos a un análisis de correlación intraclase, modelos de regresión de covarianzas, y análisis de tendencias.	El programa LST reduce significativamente el consumo de alcohol, las borracheras, consumo de marihuana, y el uso de inhalantes luego de un año para el caso de la mujeres, y el programa I-LST disminuye significativamente el fumar, las borracheras y el uso de marihuana en las mujeres. No se observó efecto para el caso de los varones en cualquiera de las situaciones.
	Evaluation of the Benefits of Transnational Transportation Projects	Jenkins GP. & Kuo CY. (2006) Argentina, Uruguay y Brasil Journal	No se menciona	Se usó la información disponible sobre la producción doméstica y las estadísticas de comercio. Se simularon los beneficios acumulados para los países importadores.	Se halló un incremento de beneficios tanto desde la perspectiva internacional como regional.
Educación	Can Cultural Barriers Be Overcome in Girls' Schooling?: The Community Support	Kim, Jooseop; Alderman, Harold; Orazem, Meter (1998)	El grupo de tratamiento incluyó a 355 niños, 175 fueron mujeres y 80 varones. El grupo de comparación incluyó a 1023	Método de evaluación de Diferencia en la Diferencia Se compararon aldeas a través de	El programa generó un impacto significativo en la matrícula primaria de las mujeres.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Program in Rural Balochistan	Pakistán Institucional	niños, 595 mujeres y 428 varones.	la Prueba de Equidad de Medias. La decisión de los padres acerca de la educación de sus hijos fue evaluada a través de la Prueba de la Igualdad de Coeficientes Conductuales en el Modelo de Elección de Matrícula.	
Salud	Evaluation of the Effect of a Consumer Driven Health Plan on Medical Care Expenditures and Utilization	Parente S, Feldman S. & Christianson JB.(2004) Estados Unidos Journal	Se utilizó a 3 grupos de cohorte. 3,636 contratos	Diseño cuasi-experimental pre-post. Se utilizó el modelo de regresión de diferencia en la diferencia.	Una evaluación temprana del gasto y utilización del programa revela que este plan es una alternativa viable para los diseños de salud existentes.
Salud	Evaluation of the effects of outreach to women with multiple vulnerabilities on entry into substance abuse treatment	Melchiora LA, Hubaa GJ, Brown VB. & Slaughterb R. (1999) Estados Unidos Journal	665 mujeres	Se aplicó una encuesta sobre antecedentes clínicos, una escala para medir su asumir cambios en la vida. Se utilizó análisis de regresión logística,	El ingreso al tratamiento del consumo de sustancias estuvo en función no sólo de su preparación para reducir su consumo de alcohol sino también con su búsqueda de consejería.
Salud	Evaluation of the orphans reunification project in Eritrea	Morah E, Mebrathu S. & Sebhatu K. (1998) Eritrea (África) Journal	200 familias fueron las beneficiarias y 28 el grupo control	Evaluación de tipo cuasi-experimental. Se utilizó el auto reporte, informe de otros, observaciones directas, medición de resultados y grupos focales.	Se halló que existe un claro impacto en la integración psicológica de los niños que estuvieron en los orfanatos en comparación
Salud	Family Planning Program Effects on Contraceptive use in Morocco, 1992-1995	Hotchkiss DR, Magnani RJ, Brown LF. & Florence CS. (1999) Marruecos	1680 mujeres	Se utilizó la información proveniente de Morocco Demographic and Health Surveys (1992 y 1995). Se utilizó un modelo de evaluación de efectos fijos.	El efecto del programa para ampliar el uso de los métodos anticonceptivos es bajo.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional			
Salud	Fertility, child work and schooling consequences of family planning programs: evidence from an experiment in rural Bangladesh	Sinha, N. (2003) Bangladesh Institucional	La muestra fue de 4364 familias residentes en 141 aldeas. Se estudió a dos submuestras: 4892 mujeres casadas y 2520 niños. Se determinó grupo de tratamiento y grupo control.	Se utilizó información del estudio Matlab Health and Socioeconomic Survey (1996). Estimador de diferencias de medias.	Se encontró que mientras el programa fue efectivo en reducir la infertilidad, no tuvo impacto significativo en la matrícula escolar de los hijos. Sin embargo, parece que el programa incrementó significativamente la participación de los niños en la fuerza laboral.
Justicia - pensiones	Fiscal implications of Pension Reforms in Italy	Brugiavini, A. & Peracchi, F. (2005) Italia Institucional	Se obtuvo una muestra aleatoria de los registros de los archivos del National Institute for Social Security. La muestra es de 200 000 trabajadores ingresados en los archivos (desde 1973-1997)	Se analizó la reforma Dini. Simulación de tres reformas hipotéticas. Se utilizó un modelo econométrico semi estructural para predecir las probabilidades de retiro bajo diferentes escenarios políticos.	Se encontró que las reformas, particularmente la reforma Dini, tienen un impacto sustancial en la decisión de retiro de las personas y en sus redes de seguridad social, así como ganancias sustanciales para las finanzas gubernamentales.
Educación	Foundations for Learning: Safe and Civil Schools Project	Washington State Institute for Public Policy (2001) Estados Unidos Institucional	La muestra estuvo conformada por 183 centros educativos: 105 beneficiarios del proyecto y 78 no beneficiarios (grupo control). La unidad de análisis estuvo conformada por los profesores de los centros educativos.	Diseño de comparación de diferencias en diferencias.	La evaluación preliminar del Proyecto de las Fundaciones es esperanzador desde que el estudio de los profesores muestra influencias positivas luego del primer año del esfuerzo de implementación planificado.
Salud	Health Policy. Lessons for Health Strategies in Europe: The Evaluation of a National Health Strategy in England	Fulop N, Elston J, Hensher M, Mckee M. & Walters R. (2000) Reino Unido	Selección aleatoria de ocho autoridades de salud.	Diseño de evaluación de estudios de caso. Entrevista semiestructurada a los actores clave sobre una gama de organizaciones (n=133), análisis	Las percepciones y documentos evidencia el impacto de la estrategia gubernamental de salud sobre la política local y cambios en el gasto.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Journal		de documentos (n= 189) y análisis de gasto de los periodos 1991/92 y 1996/97.	
Salud	Impact evaluation of a Dutch community intervention to improve health-related behavior in deprived neighborhoods	Kloeka GC, Van Lenthea FJ, Van Nieropb PWM, Koelenc MA. & Mackenbach JP. (2006) Holanda Journal	La muestra fue de 1926 adultos entre 18 y 65 años que viven en distritos pobres. Se determinó grupo de tratamiento y grupo control.	Diseño de evaluación cuasi-experimental de una cohorte longitudinal. Se utilizaron cuestionarios vía postal. Se realizó análisis de regresión logística, análisis de covarianza,	Se halló un pequeño impacto del programa sobre el consumo de vegetales por parte de la población.
Empleo	Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings	Abadie A, Angrist J. & Imbens G. (2002) Estados Unidos Journal	20 000 participantes del Programa	Se utiliza un método de variables instrumentales que mide los impactos del programa sobre los cuantiles de las variables resultantes. Los datos provienen del estudio de evaluación del programa Job Training Partnership Act.. Se hizo una simulación del método por cuantiles.	La estimación de cuantiles del efecto del programa sobre los cuantiles de de la distribución de ganancias sugieren interesantes e importantes diferencias en los efectos del programa a nivel cuantil, y diferentes en el impacto distribucional para los hombres y mujeres,
Educación	Can Private School Subsidies Increase Schooling for the Poor?: The Quetta Urban Fellowship Program	Kim, Jooseop; Alderman, Harold; Orazem, Peter (1998) Pakistán Institucional	Se realizó una selección aleatoria por estratos de los barrios de tratamiento y control.	Se utilizó un diseño de evaluación aleatorio. Los autores utilizaron tres métodos para evaluar el impacto del proyecto: Comparaciones Reflexivas, Diferencia de Medias y Diferencia en la Diferencia.	Todos los tres métodos de evaluación estimaron que el programa incrementó significativamente el ingreso de las niñas a la escuela primaria en 33%. La tasa de ingreso escolar de los niños también se incrementó en 27.5%.
Salud	Intervention against loneliness in a group of elderly women: an impact	Anderson, L. (1985) Estocolmo (Suecia)	108 mujeres adultas mayores, se determinó grupo de tratamiento y grupo control	Diseño de evaluación experimental.	Se encontró que el programa disminuye la soledad e incrementa el bienestar

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	evaluation	Journal		<p>La selección para la asignación de grupos fue aleatoria.</p> <p>Se utilizó escalas y cuestionarios psicométricos.</p> <p>Se utilizó la Prueba T de Student para muestras independientes y análisis de regresión.</p>	general entre las participantes del programa.
Nutrición	Focusing on Women Works: Research on Improving Micronutrient Status through Food-Based Intervention	Jonson-Welch C. (1999) Multinacional (Etiopía, Tanzania, Kenya, Perú y Tailandia) Institucional	5 casos de intervención en nutrición social	<p>Estudio de tipo cualitativo y cuantitativo</p> <p>Diseño de evaluación cuasi-experimental pre post.</p> <p>Grupos de discusión focal e instrumentos de evaluación cuantitativa.</p>	Se encontró que los programas evaluados obtuvieron un impacto positivo en la efectividad de las intervenciones de micronutrientes al aumentar incrementar en las mujeres su acceso a los recursos indispensables.
Empleo	Is Labour Market Training a Curse for the Unemployed? Evidence from a Social Experiment	Rosholm M & Skipper L.(2003) Dinamarca Institucional	<p>La muestra experimental fue de 812 personas</p> <p>Se determinó grupo de tratamiento (n = 425) y grupo control (n = 387).</p>	<p>Diseño experimental.</p> <p>Selección aleatoria del grupo de tratamiento y grupo control.</p> <p>Se utilizó estimadores de variables endógenas y de emparejamiento.</p>	Se encontró que las clases de capacitación incrementaron significativamente las tasas de desempleo individual.
Empleo	Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation	Paserman, D. (2004) Israel Institucional	<p>Muestra de 1008 Trabajadores varones.</p> <p>Se distinguió tres grupos: con ganancias bajas, medias y altas.</p>	<p>Se utilizó información del Estudio Longitudinal Nacional sobre Juventud (NLSY, 1978-1996)</p> <p>El efecto de las políticas se estimó a través de nivel de discontinuidad hiperbólica.</p> <p>Se estimaron parámetros estructurales</p>	El impacto de las diferentes políticas varían sustancialmente dependiendo si el modelo está calibrado con preferencias hiperbólicas o exponenciales.
Salud	Learning, private	Chan TY. & Hamilton	2467 adultos infectados con	La información proviene de	El enfoque de nuestra

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	information and the economic evaluation of randomized experiments	BH. (2003) Estados Unidos Journal	VIH. Se realizó una selección aleatoria para los cuatro grupos de tratamiento.	AIDS Clinical Trial Group Study 175 Se realizó análisis de regresión para la comparación de grupos.	evaluación halló que el bienestar del paciente podría ser incrementado al ofrecer un meno de terapias, asimismo el tratamiento mixto es preferido por la mayoría de los pacientes.
Justicia	Legal Reform, Externalities and Economic Development: Measuring the Impact of Legal Aid on Poor Women in Ecuador	Owne, B. & Portillo, J.(2003) Ecuador Institucional	Se evaluó a 362 madres. 181 madre beneficiarias y 181 madres que conformaron el grupo control.	Se utilizó información de la Encuesta de Condiciones de Vida (1988-1989). La selección de las madres beneficiarias fue aleatoria mientras que la selección del grupo de madres del grupo control fue intencional. Se realizaron análisis econométrico y estudios de grupos focales. Se utilizó el Modelo Probit Bivariado, el Modelo de Selección Heckman, el Modelo Tobit,	La evidencia empírica es consistente con la clínicas de ayuda legal que benefician a sus clientes, y existe algunos indicios de un efecto excedente.
Salud	Lessons from the Broward County Mental Health Court Evaluation	McGaha A, et al. (2002) Estados Unidos Journal	24 informantes clave. 100 acusados para el estudio longitudinal, considerando los casos del condado de Hillsborough como grupo control.	Se desarrolló un diseño de estudio de metodología cualitativa y cuantitativa. Se utilizó la siguiente información: entrevista a informantes clave, descripción del proceso de corte, estudio longitudinal de los resultados de los clientes de MHC y análisis de datos secundarios de justicia criminal y los servicios de salud	

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				mental contenidos.	
Salud	Measuring County Performance on Health: Selected for 115 Countries.	Wang J, Jamison DT, Bos E, Prever A. & Peabody J. (1999) Multinacional Institucional	115 países	Los datos fueron obtenidos de archivos mundiales sobre los temas de mortalidad, fertilidad, educación, ingreso per cápita, todos cubren el período de 1960 a 1990. Se utilizó el análisis de regresión de series de tiempo.	Los países varían sustancialmente en sus niveles de desempeño sobre varios indicadores de salud, dado los mismos niveles de ingreso y educación.
Educación	New Evidence on Classroom Computers and Pupil Learning	Angrist J. & Lavy V. (2001) Israel Institucional	Se estudió a tres muestras: 4779 estudiantes de cuarto grado, 3271 estudiantes que solicitaron por el programa Tomorrow Finds y una muestra de 2891 estudiantes evaluados en 1991.	Se utilizó la información de la evaluación realizada por el National Institute for Testing and Evaluation (1996) del nivel primario y elemental. Se hizo estimaciones a través del modelo de regresión.	No se encontró relación entre el programa de enseñanza por computador y el rendimiento en los exámenes, excepto respecto a su efecto negativo en puntuaciones de matemáticas del 8vo grado.
	New Jersey's Transitional Housing Demonstration Program: the relationship of service delivery structure and process to the attainment of more permanent forms of housing	Michael J. Camasso MJ, Jagannathanb R. & Walker CC.(2004) Estados Unidos Journal	172 participantes Se determinó grupo de tratamiento y grupo control.	Análisis de varianza para la comparación de grupos. Prueba Ji para determinar homogeneidad de muestras. El análisis de regresión proporcional de harzard para medir los efectos del programa	A pesar de las importantes limitaciones en el diseño de evaluación, se encontró que el programa ayuda a las familias pobres a encontrar y mantener una casa permanente. Sin embargo, los efectos del programa requieren de sustanciosos costos monetarios y humanos.
Educación	Can Student Loans Improve Accessibility to Higher Education and Student Performance? An Impact Study of the Case of SOFES, México	Canton, Erik, and Andreas Blom. (2004) México Institucional	La muestra incluyó a personas de 18 a 24 años que concluyeron sus estudios de secundaria.	La información de análisis proviene de la base de datos del programa, de estudios a gran escala entre los estudiantes y universidades privadas (2003) y del estudio de familias sobre logro e ingreso. Un modelo probit es desarrollado.	Los resultados empíricos sugieren que los estudiantes beneficiarios del Programa de Prestamos estudiantil (SOFES) muestran mejor desempeño académico que los no beneficiados.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				Diseño de Regresión de Discontinuidad.	
Educación	Organizational learning: A cross-national pilot-test of effectiveness in children's services	Orthner DK, Cook P, Sabah Y. & Rosenfeld J. (2006) Israel y Estados Unidos. Journal	153 niños de Israel y 153 niños de Estados Unidos.	Evaluación de tipo cuasi experimental con diseño longitudinal. Se utilizó la Escala de Evaluación de Aprendizaje Organizacional. Se utilizó la Prueba T para muestras independientes, correlación de Spearman no paramétrica y el Análisis Factorial	El Programa tuvo efecto para el caso de los niños estadounidenses pero no para los de Israel.
	Partial benefit-cost in the evaluation of the Canadian Networks of Centres of Excellence	Rank D. & Williams D (1999) Canadá Journal	No se menciona	Se evaluó el impacto a través del Análisis Costo-Beneficio.	Se encontró un gran impacto del programa en la investigación, capacitación y logro socio-económico.
Educación	Paying for Performance: The Effect of Teachers' Financial Incentives on Students' Scholastic Outcomes	Lavy, V. (2003) Israel Institucional	Profesores de escuela Se determinó grupo de tratamiento (17 escuelas) y grupo control (12 escuelas).	Diseño de evaluación pre post con diseño cuasi-experimental. El efecto del programa se identificó a través de tres estrategias: estimación del efecto experimental de la medición de error en la variable asignada, el método de regresión de discontinuidad y Propensity matching store.	Se encontró que los incentivos económicos al desempeño de los profesores tuvieron un efecto significativo en el aprendizaje de los estudiantes.
Educación	Peer-Led Asthma Education for Adolescents: Impact Evaluation	Gibson PG, Shah S. & Mamoon HA. (1998) Sydney, Australia Journal	Estudiantes mujeres del nivel secundario en 62 escuelas. Se determinó a escuelas del estatal como grupo de tratamiento (n = 62) y escuelas	Evaluación no aleatoria con diseño pre post. Se aplicó el Cuestionario de Conocimientos sobre Asma en ambos grupos de escuelas.	Se encontró que el programa educativo fue bien recibido en el ambiente escolar y permitió desarrollar los conocimientos sobre Asma en las estudiantes

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
			del sistema educativo católico como grupo control (n = 30).	La información fue analizada utilizando SAS (Cary, NC) y la prueba T de Student.	
Educación	Programa Nacional de Becas Estudiantiles Impact Evaluation Findings	Heinrich CJ. & Cabrol M. (2005) Argentina Institucional (BID)	Estudiantes provenientes de 8 provincias de Argentina, que incluyen 24 escuelas. La muestra depurada fue de 2586 estudiantes. Grupo de tratamiento y grupo control, obtenido con Matching, distinguiendo entre beneficiarios de beca por año (1-5 años) y no beneficiarios (33.37%).	La información fue obtenida de tres fuentes: Encuesta de los Aspirantes de Becas, los Registros sobre evaluación escolar de 1999-2003, y la encuesta administrada a los solicitantes de becas beneficiados y no beneficiados. 2003 Se utilizó Propensity Score Matching y análisis de regresión.	Se encontró un gran impacto del programa en la asistencia escolar en aquellos estudiantes que recibieron becas por 3 o más años. Sin embargo, no se encontró un impacto significativo para el caso de estudiantes que recibieron becas por menos de dos años.
Pobreza	Pro-growth, pro-poor: Is there a tradeoff?	López, JH. (2004) multinacional Institucional	134 observaciones de 41 países	Se obtuvo información de dos fuentes de datos: La base de datos sobre la inequidad de Dollar y Kraay's (2002) y la base de datos de Loayza, Fajnzylber, and Calderon (2002) Se estimaron modelos de panel dinámicos para diferenciar entre los impactos a corto plazo y los de largo plazo.	Se encontró que todas las políticas pro-crecimiento evaluadas conduce a menores niveles de pobreza a largo plazo. Aunque se halló, también, que algunas políticas conducen a una mayor desigualdad y mayores niveles de pobreza en el corto plazo.
Educación	Results of a School Voucher Experiment: The Case of Washington, D.C. After Two Years	Wolf PJ, Peterson PE, West MR. (2001) Estados Unidos Institucional	Se evaluaron dos muestras: 2,023 estudiantes y 1928 padres. Se determinó grupo control para ambas muestras.	La evaluación fue diseñada como experimento aleatorio debido a la selección aleatoria de los beneficiarios. Se utilizó cuestionarios para padres y estudiantes.	El programa influye en una mayor satisfacción de los padres con la escuela de sus niños, en la comunicación entre los padres y la escuelas, en el número de tareas de casa para los niños., Los niños en escuelas privadas presentan una mayor

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
					tolerancia política.
Empleo	Revisiting the Employability Effects of Training Programs for the Unemployed in Developing Countries.	Calderón-Madrid, A. (2006) México. Institucional (BID)	1786 beneficiarios del programa y un grupo control de 437 personas. Datos secundarios iniciales (Ministerios de Educación o Trabajo). Datos posteriores obtenidos de una encuesta de 1994 (Datos secundarios).	Metodología cuasi-experimental. Diseño: Emparejamiento a través de Propensity Score Matching. El impacto fue calculado a través ecuaciones diferenciales lineales discontinuas (Regresión). Modelo Mixto Proporcional de Hazard (MPH).	El impacto del programa en la dinámica del reempleo es positiva porque los participantes mantienen su trabajo por un largo período, aun si el programa no es efectivo en ayudarles a encontrar un trabajo más fácilmente.
Salud	Savings/credit group formation and change in contraception	Steele, F, Amin S. & Naved RT. (2001) Bangladesh Journal	La muestra inicial fue de 6,456 mujeres. Se determinó dos grupos de tratamiento y un grupo control	Se utilizó la información brindada por un estudio panel a cargo de Save the Children USA.(1993) Evaluación pre-post con diseño cuasi-experimental Se realizó selección aleatoria sólo para el caso de un grupo de tratamiento. Se utilizó análisis de regresión logística.	Se encontró un efecto positivo del programa de crédito sobre el uso de anticonceptivos modernos
Educación	School Choice in Dayton, Ohio After Two Years: An Evaluation of the Parents Advancing Choice in Education Scholarship Program	West MR, Peterson PE. & Campbell DE. (2001) Estados Unidos Institucional	Los beneficiados del programa suma un total de 765 estudiantes que provienen de escuelas privadas (250) y de escuelas estatales (515)	La selección de las familias para el programa fue aleatoria. Se utilizó cuestionarios para padres y estudiantes.	Se encontró que el programa tuvo impacto en la satisfacción de los padres, clima escolar, habilidades escolares, consideraciones étnicas y religiosas, tareas, así como en el compromiso y comunicación de los padres.
Pobreza	Capacity for effectiveness: the relationship between coalition structure and community impact	Hays C, Hays S, DeVille J, & Mulhall P. (2000)	28 programas en alianza para la prevención contra el consumo de drogas.	Evaluación de tipo correlacional Se aplicó una escala para medir la percepción que tienen los	Los análisis indican que las características organizacionales de una coalición podrían facilitar el

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Estados Unidos Journal		miembros acerca de la participación, inversión y cohesión de los participantes en la coalición	desarrollo de capacidades de manera diferente y, de esta manera, podría conducir a una variedad de impactos en la comunidad.
Educación	Sending Farmers Back to School: The Impact of Farmer Field Schools in Indonesia	Feder G, Murgai R. & Quinzon J. (2004) Indonesia Journal	Estudiantes y sus vecinos	Se empleo el estimador de Diferencia en las Diferencias para evaluar el impacto en el desempeño y la disminución del uso de pesticidas	Se halló que el programa no impactó significativamente en el desempeño de los estudiantes y sus vecinos.
Pobreza	Sowing and Reaping: Institutional Quality and Project Outcomes in Developing Countries	Dollar D. & Levin V. (2005) Multinacional Institucional	Proyectos financiados por el Banco Mundial en cerca 90 países los años noventas.	Estimación a través de variables instrumentales. Se utilizó análisis de regresión	Se encontró que existe una muy fuerte relación positiva entre la calidad institucional y el éxito de los proyectos.
Justicia	Testing for Structural Breaks in the Evaluation of Programs	Piehl, AM, Cooper, SJ., Braga, AA. & Kennedy, DM., (1999) Estados Unidos Institucional	No disponible	Evaluación pre post, a través del modelo de serie de tiempos, Se utilizó una prueba para la inestabilidad de parámetros.	La intervención de programa Boston Gun Project estuvo asociada con cerca del 60% de la disminución del homicidio juvenil.
Justicia	The 1997 revisions to Washington's Juvenile Offender Sentencing Laws: an evaluation of the effect of local detention on crime rates	Washington State Institute for Public Policy (2002) Estados Unidos Institucional	Arrestos de personas entre los 10 a 17 años de edad entre los años 1989 y 2000	La información fue obtenida de los reportes de las oficinas de policía. Se utilizó el análisis de regresión.	Se encontró que la tasa de admisión para las instalaciones de detención juvenil parece determinar muchos tipos de tasas de arresto juvenil.
Justicia	The comparative costs and benefits of programs to reduce crime	Aos, S. Phipps, P. Barnoski, R. & Lieb, R. (2001) Estados Unidos y Canadá	Se revisó una muestra de 400 estudios de los últimos 25 años que presentan métodos de investigación que fueron realizados en Estados Unidos y Canadá	Diferencia de medias. El estudio utilizó la comparación "apples-to-apples" de los programas.	

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional			
Empleo - justicia	The Consequences of 'In-Work' Benefit Reform in Britain: New Evidence from Panel Data	Francesconi, M. & Van der KLaauw, W. (2004) Reino Unido Institucional	La muestra comprende 3 333 mujeres sin matrimonio o convivencia, de estas 1507 son madres solteras y 1826 son mujeres sin hijos.	Se utilizó información longitudinal proveniente de 11 estudios panel sobre familias británicas (1991-2001) Se utilizaron estimadores de diferencia en la diferencia y la información panel.	Se halló que la reforma condujo a un sustancial incremento den las tasas de empleo.
Educación y trabajo infantil	The Effect of Conditional Transfers on School Performance and Child Labor: Evidence from an Ex-Post Impact Evaluation in Costa Rica	Duryea, S. & Morrison, A. (2004) Costa Rica Institucional (Banco Mundial)	Se evaluó a 746 familias beneficiarias del programa y 1042 familias no beneficiarias. Se utilizó encuestas. Muestreo intencional.	Se utilizó tres diferentes técnicas estadísticas simples: comparación de medias, análisis de regresión y emparejamiento por Propensity Score Matching.	Se encontró que el programa logra su meta fomentar la asistencia escolar y con menor magnitud respecto al desempeño escolar. El programa no reduce la probabilidad de que el niño trabaje.
Empleo – justicia	The Effect of Disability on Labour Market Outcomes in Germany: Evidence from Matching	Lechner, M. & Vazquez-Alvarez, R. (2003) Alemania Institucional	1642 personas con discapacidad. Se determinó grupo de tratamiento y grupo control.	Se utilizó la información proveniente de los encuestados alemanes occidentales al Panel Socioeconómico Alemán (GSOEP. 1984-2001). Se hizo comparación de grupos, asimismo se emparejó grupos a través del Matching Propensity Scores.	El impacto de las políticas de discapacidad para el discapacitado no fue efectivo al reducir su costo de participación dentro las actividades del mercado laboral competitivo.
Salud	The effect of the Tarrant County drug court project on recidivism	Bavon, A. (2001) Estados Unidos Journal	Se determinó grupo de tratamiento y grupo control. La muestra fue de 264 personas.	La información obtenida provino de tres fuentes primarias: DIRECT Proyect Closure List, DIRECT client files y Criminal Justice Crime Información System. Se utilizó un diseño de evaluación	Mientras la medida de los efectos sustantivos del programa es pequeña, el estudio encontró que no existe diferencia significativa en la reincidencia entre los beneficiarios del programa y los no beneficiarios.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				pre- post y de comparación de grupos no equivalentes. La diferencia de grupos fue evaluada por medio de la prueba T para medias.	
Salud	The Effects of Education and Family Planning Programs on Fertility in Indonesia	Angeles G, Guilkey DK. & Mroz TA. (2003) Indonesia Institucional	5, 025 mujeres de 13 a 51 años.	Se utilizó la información de Indonesian Family Life Survey (1993) Para la evaluación del impacto se utilizó la estrategia de modelamiento longitudinal descrita por Mroz y Weir (2003) y la simulación del ciclo de vida.	Se halló que los programas de planificación familiar tuvieron mayor efecto para reducir la fertilidad que las generadas como consecuencia del desarrollo de la calidad escolar.
Ayuda alimentaria	Child Growth, Shocks, and Food Aid in Rural Ethiopia	Yamano, Takashi; Alderman, Harold; Christiaensen, Luc. (2003) Ethiopia Institucional	Niños que viven en áreas donde existe la ayuda alimenticia (grupo de tratamiento) y niños que viven un área donde no se recibe ayuda alimenticia (grupo control)	Se utilizó la información brindada por tres estudios de familia nacionales en los años 1995-1996: a) Welfare Monitoring Survey (WMS) (b) Food Security Survey (FSS) y (c) un estudio de una muestra agrícola. Variables instrumentales y otros métodos de emparejamiento	El programa tuvo un efecto positivo y significativo sobre el crecimiento de los niños a los 6 a 24 meses de edad.
Justicia - discriminación	The Effects of State and Local Antidiscrimination Policies on Earnings for Gays and Lesbians	Klawitter, MM. & Flatt, V. (1998) Estados Unidos Journal	13000 parejas heterosexuales casadas, 14500 parejas heterosexuales sin casar y 6800 parejas homosexuales.	El estudio utiliza la información del censo de 1990. Se utilizó el análisis de regresión multivariado	Se encontró que las políticas han sido adoptadas en lugares con altas ganancias y que las parejas homosexuales viven con mayor probabilidad en áreas que han adoptado las políticas. Sin embargo, los resultados no muestran evidencia de un efecto directo de las políticas de anti-discriminación en el promedio de ganancia o ingreso para miembros de las

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
					parejas homosexuales.
Salud	The Impact of Child Support Enforcement Policy on Nonmarital Childbearing	Plotnick, RD.; Garfinkel, I.; McLanahan, S. Ku, I. (2006) Estados Unidos Institucional	Muestra de 15201 mujeres	Se utilizó la información disponible en los Archivos de Historias Matrimoniales (19885-1993) y los Archivos de historias adoptivas (1985-1993) Se utilizaron los modelos de riesgo multinivel.	Se encontró que las políticas que aumentan las pensiones para el cuidado de los niños en los hombres podría reducir la paternidad no marital.
Salud	The Impact of Family Planning Service Provision on Contraceptive-Use Dynamics in Morocco.	Steele F, Curtis SL. & Choe M. (1999) Marruecos. Journal	4,753 mujeres	Se utilizó la información proveniente de Morocco Demographic and Health Surveys -DHS (1992 y 1995). Se aplicó un cuestionario modificado de DHS. Se utilizó modelos de historia de eventos para evaluar el impacto de los indicadores a nivel comunitario de la provisión de servicio de planificación familiar. Se utilizó información cuantitativa y cualitativa.	Se encontró impacto entre la provisión de servicios de planificación familiar y la adopción de métodos modernos y la baja tasa en el fracaso de métodos anticonceptivos.
Salud	The impact of Family Planning Supply Environment on Contraceptive Intentions and Use in Morocco.	Magnani RJ, Hotchkiss DR, Florence CS. & Leigh LA. (1999) Marruecos Journal	3,168 mujeres participaron en ambos estudios.	Se utilizó información del estudio Demography and Health de Morocco (1992) y del Estudio Panel de Morocco (1995). Se utilizó una versión modificada del cuestionario DHSII "Modelo A"	Se encontró que los factores del programa de planificación familiar influyen en la conducta anticonceptiva.
Empleo	The Impact of Public and Private Job Training in Colombia.	Medina, C. & Núñez, J. (2005) Colombia	Se obtuvo una muestra de 10,000 hogares Se determinó grupo de control	Se utilizó información proporcionada por la Encuesta Nacional de Calidad de Vida (1997).	Los cursos cortos y largos proporcionados por el programa SENA no tienen un impacto significativo en el ingreso.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional		La comparación de grupos fue a través del Propensity Score Matching. También se utilizaron los estimadores de emparejamiento: semejanza promedio simple, regresión kernel y regresión lineal local.	
Pobreza	The impact of the National Minimum Wage in low-wage sectors: does the Earnings Top-up Evaluation study add to our understanding?	Urwin P, Jack G. & Lissenburgh S. (2006) Reino Unido Journal	Se estudio a una muestra de albergues	Se utilizó la información del estudio Earning Top-up Evaluation (1996, 1997 y 1999)	Se encontró un bajo impacto del programa.
Empleo	The Manager's Role in Enhancing the Transfer of Training: A Turkish Case Study	Gumuseli AI. & Ergin B. (2002) Turquía. Journal	Representantes de ventas que participan del programa. Se determino grupo control.	Diseño de evaluación experimental. El estudio se basa en el modelo de evaluación de cuatro etapas de Kirkpatrick's.	
Salud	The MEMA kwa Vijana Project: Design of a community randomised trial of an innovative adolescent sexual health intervention in rural Tanzania	Hayes, RJ. et al. (2005) Tanzania (Africa) Journal	9645 adolescentes.	La evaluación de impacto se realizó a través de una comunidad de ensayo aleatoria, en la cual una cohorte de 9645 adolescentes fueron evaluados por tres años.	Se encontró que la intervención mostró ser efectiva.
Salud	The Reach and Impact of Radio Communication Campaigns on Reproductive Health in Malawi	Meekers D, Van Rossem R, Silva M. & Koleros A. (2004) Malawi. Institucional	10,465 mujeres y 4,486 varones que fueron sexualmente activos en el año transcurrido.	Se utilizó la información del estudio Demographic and Health (2000) El efecto del programa se estimó a través de un modelo de regresión de dos colas. Se utilizó la prueba Chi cuadrada,	La exposición a los programas radiales tuvo un significativo impacto sobre las discusiones de reproducción en las familias. Sin embargo, el impacto en el uso del condón ha sido muy limitado

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				prueba T para muestras independientes y técnicas de regresión de etapa única estandarizadas.	
Salud	The relative impact of treatment program `robustness' and `dosage' on client outcomes.	Jerrell JM, & Ridgely MS. (1999) Estados Unidos Journal	132 personas con ambos desordenes: mentales y de consumo de sustancias. 35 sujetos para la evaluación cualitativa	Se utilizó un método de evaluación cuantitativa y cualitativa. Se desarrolló una entrevista semi estructurada a las 35 personas destinadas. Análisis de Regresión. Análisis de Varianza.	No se encontró diferencias entre los impactos de los tres tipos de intervención.
Empleo	Advantage through Training? A Microeconomic Evaluation of the Employment Effects of Active Labour Market Programmes in Poland	Puhani, PA (1998) Polonia Institucional	El grupo de tratamiento estuvo integrado por 938 sujetos y el grupo control por 7853	Evaluación cuasi- experimental por emparejamiento del grupo de tratamiento y grupo control, de tipo retrospectivo. Se utilizó el modelamiento econométrico en la forma de modelos de duración con la heterogeneidad individual no observada.	Se encontró que la capacitación favorece las oportunidades de empleo tanto para hombres como para mujeres, mientras que la intervención y los trabajos estatales no: los trabajos de intervención prolongan el desempleo, al igual de los trabajos estatales en el caso de los hombres.
Educación, transporte, suministro de agua y sanidad	Combining Longitudinal Household and Community Surveys for Evaluation of Social Transfers: Infrastructure Rehabilitation Projects in Rural Georgia	Lokshin, Michael; Yemtsov, Ruslan (2004) Georgia Journal	Selección de la muestra no fue aleatoria	Se tuvo a disposición a información secundaria de algunos estudios, así: La información a nivel de familias proviene del estudio oficial de familias en Georgia (SGHH, 1996), la información de nivel comunitario proviene del Rural Community Infrastructure Survey (RCIS, 2002). Diferencia en la diferencia y	El número de estudiantes se ha incrementado de manera significativa como resultado del programa. Hubo un desarrollo significativo en la percepción de cambios en el acceso a la educación de las aldeas beneficiarias en comparación con las no beneficiarias.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				Propensity Score Matching	
Salud	The Role of User Charges and Structural Attributes of Quality on the Use of Maternal Health Services in Morocco	Hotchkiss DR, et al. (2003) Marruecos (África) Institucional	Mujeres de hogares pobres y mujeres de hogares acomodados.	Se utilizó un modelo de simulación de políticas. Se realizó una comparación de grupos	La simulación de políticas mostró que el incremento en los costos de los gastos por servicios públicos puede tener muy poco efecto en las mujeres que viven en familias mejores, pero tendría un efecto substancial y perjudicial sobre la pobreza.
Salud	The social environment of transitional work and residence programs: Influences on health and functioning	Schutt RK, Rosenheck RE, Penk WE, Drebing CE. & Seibyl CL (2005) Estados Unidos Journal	1525 veteranos	Diseño de evaluación pre post. Se aplicó la Community Oriented Program Environment Scale, en su forma corta, la escala WorK Environment Scale y la escala Addiction Severity Index. Se utilizó la prueba T para muestras dependientes, el modelo lineal jerárquico	Se encontró que el trabajo y la situación domiciliaria habían sido significativamente alcanzados a los 3 meses de dejar el programa.
	The Social Impact of Social Funds in Jamaica: A Mixed-Methods Analysis of Participation, Targeting and Collective Action in Community Driven Development	Rao, V. & Ibáñez, AM. (2003) Jamaica Institucional	500 familias con 684 individuos. Se determinó grupo control según el estado de beneficiario o no del fondo.	La información cualitativa se obtuvo a través de entrevistas semi-estructuradas a participantes clave en la comunidad. Se aplicó un cuestionario. Se utilizó el modelo Propensity Score y la regresión logística	Se encontró un efecto causal de los fondos recibidos sobre el desarrollo de la unión y la capacidad para la participación colectiva.
Pobreza	The Socio-Economic Impact of Favela-Bairro: What do the Data Say?	Soares F. & Soares. Y. (2005) Brasil Institucional	54 comunidades: 34 como grupo de tratamiento y 17 como grupo control.	Se utilizó la información proveniente del General Population Census, de los Certificados de Defunción del Departamento de Salud Municipal Se midió el impacto a través del estimador de diferencia en la	Se halló que el programa tuvo un gran impacto en la expansión de los servicios públicos, sin embargo, los medios que permitieron dicha evaluación son muy limitados.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				diferencia. Los grupos se definieron por medio de Propensity Score Matching.	
Salud	The utility of case-control methods for health policy and planning analysis	Mock NB, Magnani RJ, Dikassa L, Rice JC, Abdoh AA, Bertrand WE. & Mercer DM (1993) Zaire Journal	113 casos obtenidos de dos hospitales pediátricos. Se estableció grupo de tratamiento (72) y grupo control (41).	Diseño de evaluación a través de una metodología de caso control. Se utilizó entrevistas individuales estructuradas y de observación. Se utilizaron cuestionarios de contenido antropológico, socioeconómico y de salud. Se utilizó la prueba Chi Cuadrada para muestras apareadas, método de escalas aditivas simples y de regresión.	La evolución demostró la importancia de la política de higiene relacionada al conocimiento y las prácticas como factores de riesgo para la enfermedad diarreica severa.
Salud	Think globally, act locally: assessing the impact of community-based substance abuse prevention	Saxe, et al. (1997) Estados Unidos Journal	500 personas beneficiarias y 500 para el grupo control.	Evaluación con diseño cuasi-experimental El estudio se caracteriza por ser mixto: cuantitativo y cualitativo. Se aplicó amplias encuestas tipo escala, encuestas telefónicas y se consideró indicadores sociales para el estudio de comunidades	
Empleo	Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects	Lee, D. (2005) Estados Unidos Institucional	No disponible.	Se evalúa el impacto a través de un procedimiento intuitivo para la selección de la muestra.	Se encontró un efecto positivo del programa sobre el incremento salarial por medio del fomento del capital humano.
Salud	Transfer Payments, Mother's Income, and Child Health in Ecuador	León, M. & Younger, SD. (2004) Ecuador	Madres de niños infantes	Se utilizaron estimadores de forma cuasi-reducida para medir el impacto del programa Bono Solidario	El programa ha tenido un significativo pero modesto efecto en el estado nutricional de los niños.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional			
Salud	Using time-series analysis to evaluate the impact of policy initiatives in child welfare	Albert, VN.(2001) Estados Unidos Journal	La muestra estuvo conformada por siete condados.	Se utilizó el enfoque de serie de tiempos para analizar el impacto del programa.	
Agua y Sanidad	Views of selected government officials on the impact of a rural water supply program in Ghana.	Akuoko-Asibey, A. (1997) Ghana (África) Journal	9 funcionario públicos de agencias gubernamentales	Estudio de naturaleza cualitativa. La información fue obtenida a través de la entrevista individual a cada funcionario público, dicha entrevista fue registrada en audio.	Se encontró que los funcionarios públicos en el nivel ministerial no tienen o tienen poco conocimiento acerca del impacto del programa sobre los correspondientes beneficiarios.
Educación	Compensatory Education for Disadvantaged Mexican Students: An Impact Evaluation Using Propensity Score Matching	Shapiro J & Moreno J (2004) México Institucional	Se seleccionó a los alumnos de manera no aleatoria. La muestra de estudio está representada por 45000 estudiantes de 3000 escuelas primarias. (EN) Se determinó grupo de tratamiento y grupo control.	Se utilizó la metodología Propensity Score Matching La información sobre las calificaciones de los estudiantes en sus exámenes proviene de <i>Estimadores Nacionales</i> (EN: 1998-2002).	- El programa CONAFE es más efectivo en promover el aprendizaje de las matemáticas y el español. La Educación Telesecundaria y bilingüe para los estudiantes indígenas mostraron promover el logro de los estudiantes.
Justicia - salud	Welfare Reforms, Family Resources, and Child Maltreatment	Paxon, C. & Waldfogel, J. (2001) Estados Unidos Institucional	La muestra final fue de 415 niños	Los datos sobre maltrato son obtenidos del National Center for Child Abuse and Neglect (NCCAN, 1990-1998) Se utilizaron análisis de regresión, mínimos cuadrados	Los resultados obtenidos señalan claramente algunas situaciones de alerta tal como los potenciales efectos adversos de las recientes reformas de bienestar sobre el maltrato infantil.
Educación	Wilderness challenge programs for delinquent youth: a meta-analysis of outcome evaluations	Wilson SJ. & Lipsey MW. (2000) Estados Unidos	La muestra fue de 28 estudios de investigación elegibles, que incluyó una muestra de 3000 individuos.	Evaluación meta-analítica del impacto del programa Wilderness. Se eligieron estudios que presenten diseños meta analíticos	Se encontró que todos los programas reportados por los estudios fueron efectivos en reducir la conducta antisocial y delictiva.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Journal		orientados a la diferenciación de grupos respecto la participación del programa Wilderness.	
Salud	Women's health and pregnancy outcomes: do services make a difference?	Frankenberg E. & Duncan T. (2001) Indonesia Journal	La muestra fue de 8824 individuos que intervinieron en las evaluaciones de 1993 y 1997. Se estudio a cuatro grupos: mujeres en edad reproductiva, hombres en la misma edad, mujeres mayores y hombres mayores. Se determinó grupo control	Se utilizó información del estudio Indonesia <i>Family Life Survey</i> (1993 y 1997). Se utilizó un diseño cuasi-experimental. Se utilizó el estimador de Diferencia en las diferencias. Utilizó el análisis de regresión para evaluar el cambio en el Índices de Masa Corporal.	El impacto del programa Village Midwifery se evidencia en que la adición de una partera a la aldea está asociado con el incremento del índice de masa corporal en las mujeres de edad reproductiva, asimismo, la presencia de una partera durante el embarazo se asocia con el incremento del peso al nacer.
Educación y apoyo social	Education Maintenance Allowance: The First Year. A Quantitative Evaluation	Ashworth K, Hardman J, Liu WCh, Maquire S & Middleton S. (2001) Reino Unido Institucional	7,559 jóvenes de 16 a 19 años- Se determinó grupo de tratamiento y grupo control.	Estudio de cohorte, jóvenes que concluyeron estudios de secundaria en verano de 1999. Se utilizó el estimador de diferencia en las diferencias. Análisis de regresión, Propensity Score y Técnicas de densidad de Bernal.	Se encontró que el programa tuvo un gran impacto en lo que va de su primer año de ejecución.
Educación y apoyo social	Child Education and Work Choices in the Presence of a Conditional Cash Transfer Programme in Rural Colombia	Attanasio O, Fitzsimons, Gómez A, López D, Meghir C, Mesnard A. (2006) Colombia Institucional	11,500 menores de 17 años distribuidos en 122 municipalidades: 57 conformaron el grupo de tratamiento y 65 el grupo control.	Diseño de evaluación cuasi-experimental pre post. Se seleccionó una muestra representativa por estratos. Se estimó un modelo Probit para la matrícula escolar y la participación escolar antes de iniciar el programa.	Se encontró que el programa tuvo un gran efecto en el incremento de la matrícula escolar.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				Diferencia en las diferencias. Propensity Score Matching no paramétrico.	
Educación y Nutrición	Early <u>Evaluation</u> of a New Nutrition and Education Programme in Colombia	Attanasio O, Syed M. & Vera-Hernández M. (2004) Colombia Institucional	25 comunidades beneficiarias. Se determinó grupo de tratamiento y grupo control.	Se realizaron entrevistas por agentes preparados Puntuaciones Z. Propensity Score Matching	La evaluación preliminar del programa muestra que su efecto es importante y prometedor.
Educación y apoyo social	Education Choices in México: Using a Structural Model And a Randomized Experiment to Evaluate Progresa	Attanasio O, Meghir C. & Santiago A. (2005) México Institucional	25,000 familias pertenecientes a 506 localidades beneficiarias. Se determinó grupo de tratamiento y grupo control.	Se hizo una selección aleatoria de la muestra de estudio. Aplicación de cuestionarios. Modelo estructural propuesto por el autor.	Se encontró que el programa tiene un efecto muy positivo en la matrícula escolar, especialmente al final de la escuela primaria
Nutrición y apoyo social	Medium and Long Run Effects of Nutrition and Child Care: Evaluation of a Community Nursery Programme in Rural Colombia	Attanasio OP. & Vera-Hernández M. (2004) Colombia Institucional	11,500 menores de 17 años distribuidos en 122 distritos: 57 conformaron el grupo de tratamiento y 65 el grupo control.	La selección de los distritos de estudio fue aleatoria. Aplicación de cuestionario. Puntuaciones Z. Análisis de regresión. Variables instrumentales.	Se encontró que el programa tiene un impacto duradero y positivo.
Apoyo social	The impact of a conditional cash transfer programme on consumption in Colombia	Atanasio O. & Mesnard A. (2005) Colombia Institucional	11,500 hogares distribuidos en 122 distritos, 57 de estos conformaron el grupo de tratamiento y 65 el grupo control.	Diseño de evaluación pre-post. Diferencia en las diferencias. Control de las diferencias observadas a nivel individual y comunitario.	Se encontró que el programa tiene un efecto positivo en el gasto total de consumo.
Educación y apoyo social	The Impact of a Conditional Education Subsidy on School	Attanasio O, Fitzsimons E. & Gómez A. (2005)	Familias beneficiarias con pago y familias beneficiarias sin pago.	Regresión paramétrica lineal	El programa ha sido efectivo en incrementar la matrícula escolar, especialmente entre

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Enrolment In Colombia	Colombia Institucional			los 14 a 17 años, tanto en zonas urbanas como en rurales, siendo más beneficioso para los varones.
Pobreza	Coping with the Coffee Crisis in Central America: The Role of Social Safety Nets in Honduras.	Coady, David; Olinto, Pedro; Caldes, Natalia (2003) Honduras Institucional	5,600 hogares seleccionados de manera aleatoria estratificada. Se determinaron tres grupos con diferentes modalidades de intervención del programa y un grupo control. Se seleccionaron aquellos hogares con niños menores de 13 años o madres gestantes.	Evaluación pre-post. Se utilizó el análisis de regresión para la comparación de medias de los grupos cafeteros y no cafeteros. Diferencia de medias, diferencia en la diferencia, variables instrumentales y selección aleatoria.	Se incrementó de modo significativo el consumo en los hogares.
Salud, Nutrición y apoyo social	The short-term impact of a conditional cash subsidy on child health and nutrition in Colombia	Attanasio O, Gómez LC, Heredia P. & Vera-Hernández M. (2005) Colombia Institucional	Familias de los distritos beneficiados o no por el programa. No se menciona el tamaño Se determinó grupo control	Selección aleatoria de las municipalidades beneficiarias y su emparejamiento con el grupo control. Se utilizó diferencia en las diferencias, análisis de regresión y puntuaciones Z.	El programa fomentó es estado nutricional y la morbilidad de los niños menores.
Empleo y apoyo social	Evaluating the Labor Market impacto of Working Families` Tax Credit using difference in differences	Blundell R, Brewer M. & Shephard A. (2005) Reino Unido Institucional	Padres y madres de familia	Se utilizó información del <i>Estudio sobre la Fuerza Laboral</i> y el <i>Estudio de Recursos Familiares</i> , Diferencia en las diferencias. Análisis de los efectos de tiempo y respuestas heterogéneas	La reforma tuvo un impacto positivo en el incremento de la tasa de empleo de los padres de familia, siendo mayor para el caso de las madres solteras.
Empleo y capacitación laboral	The Determinants and Effects of Work-Related Training in Britain	Blundell R, Dearden L. & Meghir C. (1996) Reino Unido Institucional	1,735 varones y 1,661 mujeres.	Estudio de cohorte. Se utilizó información del <i>Estudio Nacional Sobre Desarrollo Infantil</i> .	Se encontró que las personas que se beneficiaron de los cursos de capacitación tuvieron mejores ventajas salariales. Siendo la mayoría de estos varones.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Empleo y apoyo social	The Labour Market Impact of the Working Families' Tax Credit	Blundell R, Duncan A, McCrae J. & Meghir C. (2000) Reino Unido Institucional	1,807 padres solteros y 4,694 hogares con dos adultos.	Se utilizó la información del <i>Family Resources Survey</i> .(1994-95 y 1995-96) Se utilizó el modelo de microsimulación TAXBEN	Se encontró que el sistema tiene mayor impacto en la participación laboral de las madres solteras.
Empleo	Long Term Effects of a Mandatory Multistage Program: The New Deal for Young People in the UK	De Giorgi G. (2005) Reino Unido Institucional	Jóvenes desempleados de 18 a 24 años	Se utilizó la información de dos bases de datos: <i>New Deal Evaluation Database</i> y de <i>JUVOS data</i> . Estudio de evaluación pre post. Diseño de e regresión simple discontinua. Regresión lineal local simple no paramétrica.	Se encontró que el programa incrementó el empleo en un 6-7%.
Educación y apoyo social	Education subsidies and School Drop-Out Rates	Dearden L, Emerson C, Frayne C. & Meghir C. (2005) Reino Unido Institucional	8,414 individuos. Se determinó grupo de tratamiento y grupo piloto.	Propensity Score Matching para uniformizar a la muestra de evaluación. Entrevista directa a padres e hijos Seguimiento de los hijos a través de la entrevista telefónica. Análisis de sensibilidad a través de diferencia en las diferencias.	Se encontró que el programa reduce significativamente la tasa de abandono escolar de los jóvenes que concluyen sus estudios básicos.
Pobreza	Income Support and Staying in School: What Can We Learn from Australia's AUSTUDY Experiment?	Dearden L. & Heath A. (1996) Australia Institucional	13,868 individuos de 16 a 18 años de edad que cursan los últimos años de escuela. Donde de ellos 4000 estudiaron antes de la reforma y 9,868, luego de la reforma.	Evaluación pre post. Se utilizó la información disponible del <i>Australian Longitudinal Survey (1984-1986)</i> y del <i>australia Youth Survey (1989-1993)</i>	El programa tiene una significativa influencia para que los estudiantes con antecedentes de relativa pobreza permanezcan en la escuela pasada la edad mínima.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
			Se determinó grupo control.	Método de las variables instrumentales en dos etapas y el enfoque experimental.	
Educación	Evaluation of Aimhigher: Excellence Challenge The Early Impact of Aimhigher: Excellence Challenge on Pre-16 Outcomes: An Economic Evaluation	Emmerson C, Frayne C., McNally S. & Silva O. (2005) Reino Unido Institucional	Alumnos que completaron los 9 ó 11 años de estudios.	Evaluación a través de un estudio piloto y comparación de áreas. Se utilizó información de The National Pupil Database. Se aplicó un cuestionario para el nivel de los alumnos. Regresión lineal Propensity Score Matching. Diferencia en las diferencias.	Se encontró que la política tiene un impacto positivo para la participación de los alumnos en la educación superior.
Pobreza	The Distributional Impact of Social Transfers in the European Union: Evidence from the ECHP	Heady C, Mitrakos T. & Tsakloglou P. (2001) Multinacional (Europa) Institucional	13 países europeos	La información se obtuvo del <i>European Community Household Panel</i> . Para medir el impacto de la transferencia sobre la pobreza se utilizaron los índices Foster.	Se encontró que la transferencia ayudó a reducir la inequidad y pobreza en todos los países evaluados, sin embargo, con diferencias significativas entre los países.
Educación	Evaluation of Education Maintenance Allowance Pilots: Young People Aged 16 to 19 Years. Final Report of the Quantitative Evaluation	Middleton S, Perren K, Maguari S, Rennison J, Battistin E, Emerson C. & Fitzsimons E. (2005) Reino Unido Institucional	6.364 jóvenes mayores de 16 años	Estudio de evaluación de cohorte longitudinal Se empleó dos cohortes (1999 y 2000) Propensity Score Matching,	Se encontró que la iniciativa gubernamental incrementó la participación en estudios a tiempo completo en los jóvenes mayores de 16 años, siendo este incremento mayor para el caso de los varones.
Justicia	Data Integration in the Evaluation of Juvenile Justice Education	Parson K, Li S, & McEntire R. (2002) Estados Unidos Journal	-200 jóvenes comprometidos en los programas durante los años fiscales de 2000-2001. - 24 programa de justicia juvenil seleccionados según su representatividad en niveles de seguridad, tipo de programa,	Se hizo una evaluación pre-post para los jóvenes. El diseño de evaluación longitudinal.	Es posible sobrellevar los obstáculos políticos y burocráticos en la integración de las fuentes de información, pero tales impedimentos permanecen en el modo de política informada e

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
			perfil demográfico de la juventud que atiende.		implementación práctica.
Pobreza	Child Support Reform: Some Analysis of the 1999 White Paper	Paull G, Walker I. & Zhu Y. (2000) Reino Unido Institucional	1,904 madres cuidadoras	Se utilizó información de Family Resources Survey (1997) Se evaluó el impacto a través de un modelo de simulación.	Se encontró un efecto benéfico de la reforma respecto a la pobreza infantil, sin embargo no se logra el alcance esperado.
Empleo	An Evaluation of the Swedish System of Active Labour Market Programs in the 1990s	Sianesi, B. (2003) Suecia Institucional	116, 130 individuos Se determinó grupo control	La información se obtuvo de dos fuentes principales: <i>Händel</i> y <i>Askat</i> . Análisis de Series de Tiempo Propensity Score Matching	Se encontró que el programa incrementó la tasa de empleo en los participantes. Sin embargo, también prolongó su permanencia en el programa.
	Differential Effects of Swedish Active Labour Market Programmes for Unemployed Adults During the 1990s	Sianesi, B. (2003) Suecia Institucional	30.800 individuos adultos, beneficiarios de seis programa de promoción del empleo Se determinó grupo control	La información se obtuvo de dos fuentes principales: <i>Händel</i> y <i>Askat</i> . Propensity store Análisis de serie de tiempos	Se encontró que el efecto de los programas evaluados sobre el empleo y desempleo fueron desalentadores para todos ellos excepto para los de subsidio laboral
Educación	Ability, Parental Background and Education Policy: Empirical Evidence from a Social Experiment	Meghir C. & Palme M. (2003) Suecia Institucional	La muestra final fue de 11,950	Estudio de tipo prospectiva. Se evaluó la cohorte de 1948. Cuestionarios Entrevistas personales y telefónicas. Propensity Score Matching Análisis de Componentes principales.	La reforma incrementó el logro educativo de aquellos individuos de padres obreros. Asimismo, incrementó significativamente el ingreso de aquellos padres obreros.
Educación	Education Maintenance Allowance Transport Pilots – Quantitative Findings from Year 1 and 2 (2000-2001/2001-2002)	Perren K, Middleton S. & Emerson C. (2003) Reino Unido Institucional	Jóvenes que salen de la escuela entre fines de los años 1999 y 2002 4,432 (2002), 2,266 (2001) y 4,427 (2000)	El diseño de evaluación estadística fue un estudio de cohorte longitudinal. La selección de la muestra de estudio fue aleatoria. Propensity Score Marching	No se encontró un significativo impacto del programa

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
			Selección no aleatoria de cinco áreas beneficiarias del programa y 11 actuaron como grupo control		
Empleo	The Impact of Training Policies in Latin America and the Caribbean: The Case of Programa Joven	Aedo, C. & Núñez, S. (2004) Argentina Institucional	3, 340 beneficiarios y no beneficiarios del programa, es estableció grupo control	Método de evaluación no experimental pre post. Se utilizó información de los registros del programa. (1996) La selección del grupo fue aleatoria. Método de emparejamiento a través del estimador Propensity Score. Análisis costo beneficio de los efectos del programa.	El impacto tuvo un impacto estadísticamente significativo en el ingreso de los varones y las mujeres adultas.
Pobreza	Costs and Benefits of Privatization: Evidence from Brazil.	Anuatti-NEto F, Barossi-Filho M, Gledson de Carvalho A. & Macedo R. (2003) Brasil Institucional	264 compañías, Se utilizó grupo control	Se utilizó información de los informes financieros anuales de empresas de propiedad estatal y de algunas compañías privadas. (1987-1990). Se utilizó la prueba de media y mediana así como el análisis de datos panel.	El principal efecto del programa de privatización de las empresas es el incremento en su eficiencia y desempeño.
Pobreza y desigualdad Social	Economic Reform and Wage Differentials in Latin America	Behrman JR, Birdsall N. & Székely M. (2000) Latino américa Institucional	18 países de Latinoamérica 79 estudios de familias (1980-19998)	Se construyó un panel de información sobre el salario por nivel de instrucción. La estimación del impacto se realizó a través del coeficiente de vectores y el análisis de regresión.	Las reformas económicas de los años 1980-1998 tuvieron como efecto ampliar las desigualdades de los sueldos en el corto plazo. Asimismo, tuvo como efecto positivo el incremento de los niveles del salario, pero afectando negativamente en el salario de los menos instruidos.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Pobreza y desigualdad social	Pobreza, desigualdad, y liberalización comercial y financiera en América Latina	Behrman JR, Birdsall N. & Székely M. (2001) Latinoamérica Institucional	93 encuestas de hogares para 17 países latinoamericanos.	Se obtuvo información de las encuestas de hogares correspondientes a 1977-2000. La evaluación económica se realizó a través de un análisis econométrico. Se utilizó la ecuación semi-logarítmica de Mincer	Se encontró que la liberalización comercial no tuvo efectos distinguibles sobre la desigualdad y pobreza en la región, mientras que la liberalización financiera sí lo tuvo.
Educación	The Impact of Experimental Nutritional Interventions on Education into Adulthood in Rural Guatemala: Preliminary Longitudinal Analysis	Behrman JR, Martorell R. & Stein AD. (2003) Guatemala Institucional	Estudio de una cohorte Niños nacidos entre 1969 y 1977.	Estudio de evaluación longitudinal. La estimación del impacto se realizó a través del modelo de forma reducida. Los datos se analizaron a través de estimaciones Probit.	Existe un efecto significativamente positivo y del suministro de <i>Atole</i> , recibido a la edad de 6 a 24 meses, sobre los resultados educativos evaluados 27-32 años después.
Educación y protección social	Did Social Safety Net Scholarships Reduce Drop-Out Rates during the Indonesian Economic Crisis?	Cameron, Lisa A (2002) Indonesia Institucional	Se seleccionó a 120 familias por cada una de las 100 aldeas de Indonesia que fueron consideradas para la evaluación. Esta selección no fue aleatoria. Se determinó grupo de tratamiento y grupo control.	Diferencia de medias, variables instrumentales, Propensity Score Matching, Two Stage Least Squares y Análisis de regresión. Dos métodos de emparejamiento fueron utilizados: el 'five nearest neighbors' y el método basado en Kernel.	Se halló que el programa tiene un impacto significativo negativo en los casos de deserción escolar.
Ingreso Familiar	What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments	Bitler MP, Gelbach JB. & Hoynes HW. (2004) Estados Unidos Institucional Institucional	4803 casos. Se determinó grupo de tratamiento y grupo control	Se utilizó información de March CPS data, Diseño de evaluación experimental pre-post. Selección aleatoria de grupos. Se determinó el impacto a través del estimador Propensity Score.	Los efectos de las reformas de bienestar son más variados y extensos de lo que se había esperado.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Empleo	Evaluating the Working Families Tax Credit	Blundell R, Duncan A, & Meghir C. (2002) Reino Unido Institucional	Se utilizó dos muestras: Padres de familia solteros (1807) y parejas casadas o convivientes (4694).	Se analizó la información de los Estudios de Gasto Familiar 1994-1995 y 1995-1996. Se utilizó el estimador de diferencia en la diferencia. y un modelo de simulación estructural.	Se encontró un efecto moderadamente positivo del programa en las repuestas de empleo.
Pobreza y desigualdad social	Labor Market Regulations and Income Inequality: Evidence for a Panel of Countries	Calderón C, Chong A. & Valdés R. (2004) Multinacional Institucional	Se obtuvo información de 121 países perteneciente a los años 1970-2000.	Se utilizó la información proveniente de dos bases de datos sobre regulaciones laborales: Rama y Artecona (2002) y Botero, Djankov, La Porta, López-de-Silanes y Shleifer (2003). Análisis de regresión su dimensión transversal y de información panel.	Se halló que las regulaciones de derecho no favorecen la distribución del ingreso, la conformidad con las regulaciones existentes parecen favorecer la distribución del ingreso asimismo, un mayor salario mínimo tiende a empeorar la distribución del ingreso.
Pobreza y crecimiento económico	Privatization discontent and its determinants: evidence from Latin América	Carrera J, Checchi D, & Florio M. (2005) Multinacional Institucional	Muestra de 18522 personas en 17 países. Estimador de orden Probit	Se utilizó los resultados de Latinobarómetro (2002) Se utilizó suma de cuadrados, análisis probit y análisis de regresión.	Se halló un descontento generalizado de los ciudadanos con la privatización respecto a los temas de distribución.
Empleo	Impact evaluation of Empleo en Acción: Lessons and relevante for Colombia	Castro MF. (2004) Colombia Institucional	160 Familias beneficiadas y no beneficiadas del programa	Diseño de evaluación cuasi-experiimental pre post. Estudios de evaluación cara a cara, entrevistas cara a cara.	El impacto del programa para el corto plazo fue positivo.
Pobreza y desigualdad social	An evaluation of the Distributional Power of PROGRESA'S Cash Transfers in México	Coady DP. (2001) México Institucional	14,856 familias beneficiarias del programa y 9221 familias no beneficiarias.	Se utilizó la información de los censos tanto de la muestra nacional como a nivel del programa.	Se encontró que el programa ha tenido un relativamente alto poder distributivo del poder comparado con una diversidad de alternativas consideradas.
Pobreza y educación	Does Child Care Pay?: Labor Force Participation and Earnings Effects of Access to Child Care in	Deutsch, R. (1998) Brasil	1720 familias con hjos menores de seis años de edad.	Entrevistas y instrumento de sondeo. Se utilizó la comparación de	Se halló que el financiamiento público de los servicios de cuidado infantil no sólo ofrece esperanzas de largo plazo para

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	the Favelas of Rio de Janeiro	Institucional		medias.	la reducción de la pobreza, sino también una inmediata retribución en la forma de empleo y oportunidades de empleo para las madres beneficiarias.
Empleo	Labor Market Regulation and Employment In the Caribbean	Downes A, Mamingi M. & Belle Antoine RB (2000) Barbados, Jamaica y Trinidad y Tobago Institucional	No se especifica.	Se utilizó datos anuales sobre empleo. La investigación de las propiedades temporales de las series se hizo a través de la prueba de raíz de Dickey-Fuller aumentada en una unidad. También se utilizó el Test de cointegración de Johansen y el modelo de corrección no lineal de Phillips –Loretan.	Las regulaciones laborales en los tres países estudiados no han causado un impacto significativo en comparación con los otros países de la región.
Empleo	An Econometric Cost-Benefit Analysis of Argentina's Youth Training Program	Elías VJ, Ruiz F, Cossa R. & Bravo B. (2004) Argentina Institucional	1670 jóvenes beneficiarios y 1670 no beneficiarios.	Se realizó un análisis econométrico costo-beneficio. Se utilizaron técnicas de estimación de no emparejamiento (pre post, diferencia en la diferencia) y de emparejamiento basado en la estimador Propensity Score (Por aproximación, Kernel, Regresión lineal local) para compensar el sesgo de selección.	En comparación con otros países se considera que el programa tiene un impacto positivo en la búsqueda laboral de los jóvenes.
Salud	Evaluation of the Family Allowance Program (PRAF) in Honduras: Health and Nutrition Impacts	Flores R, Morris S, Olinto P, Medina J. & Neidecker O. (2003) Honduras Institucional	4197 beneficiarios y 7849 pertenecientes al grupo control	Se utilizó un diseño de evaluación de prueba aleatoria de grupos, aleatorización	El programa tuvo impacto en el cuidado prenatal y la examen del niño, pero no en el examen perinatal.
Educación	Do Community-Managed Schools Work? An	Jimenez, Emmanuel, and Yasuyuki Sawada.	897 estudiantes en escuelas puras o mixtas con EDUCO y	Los datos fueron obtenidos por el Ministerio de Educación en un	El programa promueve de manera significativa la

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Evaluation of El Salvador's EDUCO Program	(1998) El Salvador Journal	154 en escuelas tradicionales puras.	estudio de 311 escuelas. El estudio aplicó cinco cuestionarios: para estudiantes, padres, directores de escuela, profesores y asociación de padres. Variables instrumentales para controlar el sesgo de selección. Se utilizó un proceso Heckman de dos etapas.	asistencia estudiantil.
Pobreza	Decentralized Targeting of an Anti-Poverty Program	Galasso M. & Ravallion M. (2005) Bangladesh Institucional	3,625 familias con niños de 5 a 16 años de edad.	Los análisis empíricos se basan el Estudio de Gasto Familiar (1995-1996). Regresión no paramétrica, regresión Tobit. Prueba T.	El programa está orientado hacia la pobreza, una mayor proporción de los pobres recibe ayuda que los no pobres.
Empleo y crecimiento económico	The Costs and Benefits of Privatization in Argentina: A Microeconomic Analysis	Galiano S, Gertler P, Schargrodsy E. & Sturzenegger F. (2003) Argentina Institucional	21 empresas no financieras de propiedad estatal, actualmente privatizadas o no.	Logaritmos de costos de unidad. Modelo de regresión cuantílica. Estimador de la desviación absoluta mínima. Estimador diferencia en la diferencia.	El programa de privatización tuvo un impacto positivo en el incremento de la rentabilidad pero muy negativo en el empleo.
Salud y bienestar social	An Experiment in Incentive-Based Welfare: The Impact of PROGESA on Health in Mexico	Gertler PJ. & Boyce S. (2001) México Institucional	506 comunidades (89,293 personas): 320 asignadas al grupo de tratamiento y 185 al grupo control. Estimador de diferencia en la diferencia. Análisis de regresión.	Asignación aleatoria de los grupo de tratamiento y control. Se utilizó información administrativa de las clínicas públicas y del estudio panel a gran escala de Progesa.	Se halló que el programa incrementó de modo significativo la utilización de las clínicas de salud pública para el cuidado preventivo. El programa también disminuyó el número de hospitalizaciones y las visitas a proveedores privados. La promoción de la salud fue significativa.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Salud	Efectos de la reforma en salud sobre las conductas irregulares en los hospitales públicos: El caso de Bogotá, Distrito Capital, Colombia	Giedion U, Gonzáles L. & Acosta OL. (2001) Colombia Institucional	22 hospitales	Entrevista a informantes clave, encuesta a personal hospitalario mediante selección de muestra aleatoria, censo a enfermeras de todos los hospitales de Bogotá, presupuestos de hospitales. Se construyó una matriz de conductas irregulares y con variables institucionales e individuales.	Los hospitales del Distrito, contrario a lo buscado en la reforma, están lejos de ser autónomos tanto en términos de manejo de su personal como su independencia financiera
Salud y Nutrición	On the Bias in Estimating the Impact of a Nutrition Intervention Using Clinic Based Data	Handa S. & Huerta MC. (2003) México Institucional	505 comunidades, 323 beneficiarias del programa y 181 como grupo control. Esto comprende un total de 125,000 niños con 12 a 23 meses de edad	Se utilizó información antropométrica de los niños beneficiarios obtenidos de los centros de salud. Se utilizó la regresión logística.	Los resultados de impacto obtenidos en el ámbito clínico son menores a los mostrados por los experimentos sociales.
Empleo y seguridad social	The Cost of Job Security Regulation: Evidence from Latin American Labor Markets	Heckman J. & Pagés C. (2000) Multinacional Institucional	43 países: 28 europeos y 15 latinoamericanos.	Se utilizó una metodología de evaluación comparativa y de serie de tiempos.	Se encontró que las regulaciones de seguridad laboral tuvieron un substancial impacto sobre el empleo y la tasa de despidos en los países de Latinoamérica y Europa.
Empleo y seguridad social	Labor Market Policies and Employment Duration: The Effects of Labor Market Reform in Argentina	Hopenhayn HA. (2001) Argentina Institucional	64,000 individuos uniformemente distribuidos por años.	Se utilizó la información panel del estudio de hogares para el área del Distrito Federal de Buenos Aires. Análisis de regresión lineal	Se encontró que la reforma tuvo un fuerte impacto en el despido laboral, incrementando la tasa de riesgo durante el período de prueba por 40%.
Empleo y seguridad social	The Incidence of Job Security Regulations on Labor Market Flexibility and Compliance in Colombia: Evidence from the 1990	Kluger A. (2000) Colombia Institucional	No se especifica.	Diseño de evaluación pre post. Se utilizó información de los Estudios de Hogares en Colombia (1988, 199 y 1996) Se estimó las tasas de riesgo. Se utilizó el estimador de diferencia en las diferencias y los modelos de riesgo exponencial.	Se encontró que la reforma laboral ha generado un gran dinamismo del mercado laboral así como la disminución de la tasa de desempleo.
Empleo	Macro Policy and employment problems in	Lora E. (1998) Multinacional	18 países de Latinoamérica	Se utilizó la información de los estudios de hogares y otros	Las reformas de estabilización y estructurales han generado

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Latin América.	Maluccio J. Institucional		estudios de evaluación de impacto. Modelo de simulación macro laboral y análisis de regresión.	un cambio del empleo hacia el sector de servicios y el incremento del mercado laboral informal.
Salud y bienestar	Impact Evaluation of the Pilot Phase of Social Safety Net (RPS) in Nicaragua: Health and Nutrition Impacts	Maluccio J. & Flores R. (2003) Nicaragua Institucional	706 familias beneficiarias y 690 familias como grupo control	Entrevista Diferencia de medias	El programa tuvo impacto en el incremento de consultas para el cuidado preventivo de los niños menores de tres años, la disminución de la tasa de deserción escolar a nivel primario y el incremento del gasto alimenticio en la familia.
Transporte (Infraestructura)	Do Donors Get What They Paid For? Micro Evidence on the Fungibility of Development Project Aid	Van De Walle, Dominique and Dorothyjean Cratty. (2005) Vietnam Institucional	La muestra final fue de 38 distritos de los cuáles se seleccionaron aleatoriamente 100 comunidades beneficiarias del proyecto y 100 no beneficiarias.	Estudio de evaluación en serie de tiempos. Evaluación a través del diseño de estimación Diferencia en la Diferencia. Se utilizó la información del Estudio de Impacto de los Vías Rurales en Vietnam. La comparación de las comunidades beneficiarias del proyecto con las no beneficiarias fue a través del Propensity Score Matching.	El programa tuvo un efecto positivo y significativo en la rehabilitación de vías de transporte terrestre.
Empleo	Trade and Employment: Evidence from Latin America and the Caribbean	Márquez G. & Pagés C. (1997) Multinacional Institucional	Información panel anual de 18 países de Latino América y el Caribe.	Análisis econométricos.	Se encontró que las reformas de comercio tuvieron un efecto negativo, aunque pequeño, sobre el incremento del empleo.
Empleo	The Effects of Labor Market Regulations on	Mondito G. & Montoya S (2000)	1.398 empresas manufactureras	Estudio de evaluación de tipo descriptivo.	Se encontró que las reformas tuvieron un gran y

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Employment Decisions by Firms: Empirical Evidence for Argentina	Argentina Institucional		Se utilizó la información panel sobre horas trabajadas, horas extras, salarios y producción física (1990-1996) Análisis de regresión.	significativo efecto en las decisiones de empleo,
Empleo	The Impact of Regulations on Brazilian Labor Market Performance	Paes R. & Corseuil CH. (2001) Brasil Institucional	5000 establecimientos manufactureros que emplean más de cinco trabajadores.	Se utilizó la información de Pesquisa Industrial Mensal (1985-2001) para la información longitudinal mensual de la demanda laboral. Método de diferencia en las diferencias. Análisis de regresión.	Se encontró que los despidos disminuyeron luego de la reforma constitucional pero sólo en el caso de empleos por períodos largos, lo opuesto fue para los empleos por períodos largos.
Salud y Seguridad social	How Do Social Security and Income Affect the Living Arrangements of the Elderly? Evidence from Reforms in Mexico and Uruguay	Shinkai N. (2000) México y Uruguay Institucional	2697 ancianos en mexicanos y 9806 ancianos uruguayos.	Se utilizó información de los Estudios Nacionales tanto en México como Uruguay.	Se encontró que el sistema de ayuda del gobierno no tiene influencia en el desarrollo de los adultos mayores de las áreas rurales en México. En el caso de Uruguay, el sistema de ayuda gubernamental explicaría la probabilidad de las mujeres ancianas vivan solas.
Nutrición	ADRA/Ghana's Food Security Program (PL 480 Title II)	Adu-Gyamfi JJ, Twum-Ampofo K, Akuamoah J. & Kwaku A. (2006) Ghana Institucional (USAID)	Beneficiarios: 30,000 granjeros de 468 comunidades de Ghana. Se usó una muestra aleatoria de 47 comunidades donde opera ADRA.	Revisión de documentos e informes. Entrevista al equipo de profesionales y beneficiarios del proyecto. Grupos Focales. Comparación del estado nutricional de los niños. Grupo control y grupo experimental. Se usó comparaciones Z.	La intervención de ADRA ha sido positiva en la vida y bienestar general de los granjeros y sus familiar. Asimismo contribuyó al logro de metas de la Misión USAID/Ghana.
Apoyo social	Impact evaluation of the peace corps small project	Alexieva A, Alexieva K. & Vasev I.	Se obtuvo una muestra intencional de 16 proyectos de	Método cualitativo.	Mientras que el Programa SPA parece tener mayor

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	assistance (SPA) project.	(2003) Bulgaria Institucional (USAID)	un total de 141.	Análisis de documentos: Se revisó el informe final de los 16 proyectos. Se realizó entrevista a informantes clave a través de cuestionarios guía.	impacto en las ciudades más pequeñas o de mediana dimensión, los efectos son bastante contingentes sobre la participación y cooperación entre varios de los actores comunitarios.
Infraestructura	Rural Roads Evaluation Summary Report	Anderson GW & Vandervoort ChG. (1982) Multinacional (Honduras, Jamaica, Colombia, Sierra Leone, Liberia, Kenya, Filipinas, Tailandia) Institucional (USAID)	Revisión de 8 informes de proyectos en diversos países.	Método cualitativo. Diseño de Estudio de casos. Revisión de documentos (informes de evaluación).	Según el informe de todos los estudios de caso, la producción agrícola se incrementó como resultado de la ejecución del proyecto.
Agricultura y apoyo social	Agricultural credit in the Dominican Republic	Araujo F, Viveros AM. & Murphrey J (1985) República Dominicana Institucional (USAID)	250 agricultores.	Método cualitativo. El estudio tiene un enfoque de análisis sociocultural. Se entrevistó a miembros de los familiares que participaron.	Se informa con certeza la relación entre los efectos del programa y el incremento de la producción agrícola.
Educación	Impact evaluation of the NIS Farmer-to-Farmer Program	Askin P, Childress M, Delmare R, Estes V, Smith M. & Sullivan G. (1996) Multinacional ((Russia, Kazakhstan, Kyrgyzstan, Ukraine, Moldova, Armenia y Georgia). Institucional (USAID)	Se evaluó a 77 organizaciones anfitrionas y a 100 personas integrantes de estas asociaciones. Se realizó muestreo aleatorio.	Método cualitativo – cuantitativo (mixto). Diseño de Estudio de casos que evalúa el impacto en tres niveles: cambios en la psicología de los beneficiarios, cambios en otros además de los beneficiarios y cambios en los sistemas subyacentes de producción, comercio, política, etc.	Se evaluaron tres niveles. Se encontró que los mayores efectos del programa fueron evidentes en el nivel uno: en la psicología de los beneficiarios (conocimientos, actitudes, competencias). En los otros niveles, los efectos fueron significativamente menores.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				Entrevistas profundas con las organizaciones beneficiarias, cuestionarios a muestra aleatoria, entrevistas telefónicas. Estadísticas descriptivas.	
Electrificación rural	Economic and social impact evaluation of the Rural Electrification Program in Bangladesh.	Barkat A, Khan SH, Rahman M, Zaman S, Poddar A, Halim S, Ratna NN, Majid M, Maksud AKM, Karim A. & Islam K. (2002) Bangladesh Institucional (USAID, NRECA, REB)	Se evaluó a 23 unidades Palli Bidyut Samities (PBSs), los cuales son modelos únicos de gobierno en Bangladesh. Se determinó grupo control. 2278 familias beneficiarias del programa y 1440 familias sin electricidad. Constituyendo un total de 3718 familias. Las muestras se distribuyeron según cuatro unidades de observación.	Diseño experimental de estudio sólo postest. Ausencia de línea base (pretest) Selección aleatoria de grupo de tratamiento y grupo control. Se usaron 4 tipos de cuestionarios; 4 listas de chequeo; 5 formatos de recolección de datos secundarios y 2 guías de grupos de discusión. 27 Grupos focales con la muestra y 9 grupos focales con las autoridades. Regresión OLS y TOBIT. Análisis descriptivos.	El programa de electrificación fue exitoso.
Pobreza	Do Family Caps on Welfare Affect Births Among Welfare Recipients? Reconciling Efficacy and Effectiveness Estimates of Impact through a Blended Design Strategy.	Jagannathan R, Camasso M. & Killingsworthusa M. (2004) Estados Unidos Journal	9000 casos: 6000 en el grupo experimental y 3000 en el grupo control.	Se utilizó un diseño investigación experimental y cuasi-experimental (diseño ciego). Se realizó una asignación aleatoria de los grupos.	Las estimaciones encontradas sobre el programa pueden diferir dramáticamente, reflejando la sensibilidad hacia una heterogeneidad de tiempo inducido que se le ha denominado efecto Heráclito.
Educación	The Monrovia consolidated school system: an impact evaluation	Pewu G. & Belleh W. (1982) Liberia (África) Institucional (USAID)	Muestra aleatoria de 20 informantes clave, entre directores, especialistas, y participantes.	La evaluación del impacto del proyecto fue realizada a través del enfoque del Informante Clave (Método Cualitativo). Se revisó documentos.	El impacto del programa fue parcial.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Educación	Promoting education girls in Nepal.	Benoiel S, Ilon L, Sutton M, Karmacharya DM, Lamichhane S, Rajbhandry P, Der Kafle B. & Giri S. (1998) Nepal Institucional (USAID)	95 madres de niños en escuela primaria, provenientes de cuatro villas. Se determinó grupo de tratamiento y grupo control, sin asignación aleatoria.	Diseño pre-experimental. Comparación sólo postest. Entrevistas estructuradas mediante cuestionario. Comparación de porcentajes.	Se halló que las mujeres que completaron los programas de alfabetización se comprometen mucho más en la educación de sus niños, en varios modos.
Educación	Promoting primary education for girls in Malawi.	Bernbaum M, Fair K, Miske S, Moreau T, Nyirenda D, Sikes J, Wolf J, Harber RB, Hartwell A. & Schwartz B. (1999) Malawi Institucional (USAID)	130 documentos y 200 personas dentro y fuera del programa educativo.	Método cualitativo. Se hicieron observaciones y entrevistas a profesores, padres y alumnos. Se revisaron documentos. Se utilizó técnicas cualitativas (triangulación) y análisis de datos cuantitativos.	El programa tuvo significativos logros.
Educación en ciudadanía y derechos	The CEE/SEE Program: Impact and Potential	Blue, RN, Galaty M. & Green A. (2006) Estonia, Hungría, Croacia, Bulgaria y Eslovenia Institucional (USAID)	Cinco países: cuatro países donde el programa estuvo activo y un país control, donde el programa no ha sido ejecutado. La unidad de análisis fueron los líderes de ONGs. Selección no aleatoria.	Método mixto. El diseño de evaluación básico utilizó un método de comparación de países. Se utilizó entrevistas cualitativas/cuantitativas a informantes clave. Aplicación de Cuestionario y Grupos Focales.	Se encontró una respuesta positiva de los cuatro países beneficiarios hacia el programa.
<i>Proyecto creado para la provisión de servicios de salud y desarrollo comunitario</i>	Haiti: Hacho Rural Community Development.	Brinkerhoff DW, Fotzo PT, Ormod BJ. (1983) Haití Institucional (USAID)	77 personas clave contactadas y entrevistadas.	Método cualitativo (triangulación). Se empleó una mezcla de técnicas de reconocimiento rápido: examen de documentos e informes, entrevistas a	El programa fue exitoso.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				informantes clave, visitas de campo, y observación directa.	
Calidad de vida y desarrollo	Small Project Assistance (SPA)	Creative Associates & Benchmarks, INC. (1985) Multinacional (Ecuador, Senegal Gambia, Honduras, Tailandia, entre otros 34 países) Institucional (USAID)	44, 23 y 87 miembros de Peace Corps, AID y Peace Corps voluntarios. 62 proyectos evaluados.	Método mixto (Encuesta de opinión y estudio de casos). Examen documentario de 62 proyectos evaluados. Cuestionarios, entrevistas semi estructuradas.	El programa brinda una herramienta esencial para el desarrollo de la comunidad y contribuye enormemente a la satisfacción laboral de los voluntarios.
Educación	Management Education in Modern Tunisia: L'Institut Supérieur De Gestion, Túnez.	Casstevens TW, Jonson NC, King GJ. & Willens L.(1985) Túnez Institucional (USAID)	Se incluye a personas de negocio y académicos. No menciona el número. Muestreo intencional.	Metodología cualitativa Entrevistas personales y revisión documentaria.	El proyecto es bastante exitoso, aunque su desarrollo difirió de los planes originales.
Resolución de conflictos sociales	Centre for Conflict Resolution: Impact Evaluation Report.	Centre for Conflict Resolution (2004) Kenya Institucional (USAID)	Se realizó un muestreo aleatorio estratificado. De una población beneficiada de 6000 personas, se seleccionaron 276 beneficiarios del programa y nueve informantes clave de Samburu, Nakuru & Laikipia. 39 grupos focales.	Método Mixto. Se utilizó: cuestionarios, discusiones de grupo focal, entrevistas a informantes claves, entrevistas individuales y grupales, así como la revisión documentaria. Los datos fueron sometidos a análisis cuantitativos (estadísticas descriptivas).	La intervención del programa en la comunidad generó un ambiente de paz.
Infraestructura de Agua potable	Korean Potable Water System Project: Lessons from Experience.	Chewynd E. & Dworkin DM. (1981) Korea	Selección aleatoria estratificada de las familias beneficiarias. 181 Familias beneficiadas por el programa.	Método mixto. Entrevista a hogares beneficiados, mediante cuestionario.	Se encontró que el programa tuvo mayor impacto en la tasa de suscritos para la instalación a la red de agua potable en la zona rural que la urbana.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional (USAID)	Informantes clave.	Entrevista a informantes clave.	
Educación	An Impact Evaluation of the Educational Support Services Trust's Programme.	Ebrahim R. (1996) Sudafrica) Institucional (USAID)	200 estudiantes del cuarto grado de seis escuelas primaria de Cabo Oeste, Khayelitsha y Cabo Este. De estos, 96 participaron en el grupo de tratamiento y 104 en el grupo control.	Método cuasi-experimental. Diseño de evaluación pre post y con grupo control no equivalente. La obtención de datos fue a través de la aplicación de una Ficha de Ejercicios a los estudiantes. Se realizó análisis de varianza con medidas repetidas y estadísticas descriptivas.	No se encontró efectos significativos del programa.
Empleo	Do Workfare Participants Recover Quickly from Retrenchment? Programa Trabajar II	Ravallion, Martin; Galasso, Emanuela; Lazo, Teodoro; Philipp, Ernesto (2001) Argentina Institucional	1500 de los participantes del programa fueron entrevistadas, quienes fueron seleccionadas de manera aleatoria. Para la selección del grupo control se utilizó el Permanent Household Survey (1999)	Método cuasi-experimental. Propensity Score Matching y otros métodos de emparejamiento de grupos.	El programa tiene un efecto positivo y significativo sobre el ingreso.
Educación	Youth Reintegration Training and Education For peace (YRTEP) Program	Fauth G. & Daniels B. (2001) Sierra Leona (África) Institucional (USAID)	Se realizó un muestreo aleatorio estratificado de 56 lugares de un total de 1300. En cada sitio seleccionado se escogió entre 10 y 20 participantes. Muestra total: 482 participantes.	Método pre-experimental (Diseño solamente postest). Se aplicó un Cuestionario de evaluación, diseñado y adaptado a partir de un estudio piloto de 20 participantes del programa. Se utilizó estadísticas descriptivas (porcentajes).	Se encontró un significativo efecto del programa en los participantes.
Infraestructura – Irrigación Agricultura	Bangladesh Small-Scale Irrigation	Friedkin T, Lester R, Blank H. & Ahmed N. (1983) Bangladesh	400 agricultores participaron del programa y 200 integraron el grupo control.	Método pre-experimental. Diseño sólo postest con grupo control.	Un impacto obvio del programa ha sido el incremento del ingreso de los agricultores. Asimismo, se observan otros efectos

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional (USAID)		Se revisó informes y documentos. Se entrevistó al personal encargado del proyecto Se aplicó un cuestionario a los agricultores. Estadísticas descriptivas.	positivos no planificados.
Infraestructura eléctrica y agricultura	The Product is Progress: Rural Electrification in Costa Rica	Goddard PO, Gómez G, Harrison P. & Hoover G (1981) Costa Rica Institucional (USAID)	Se realizó una selección intencional de la muestra a través de expertos en ciencias sociales. 96 familias distribuidas en 17 comunidades. Se estableció grupo control.	Método pre-experimental. Diseño solo postest con grupo control. Se aplicó un cuestionario de 61 ítems. Estadísticas descriptivas.	No se encontró un efecto importante del programa y el desarrollo agrícola de las comunidades.
Nutrición y educación	Growth Monitoring, Nutrition and Education: Impact evaluation of an effective applied nutrition program in the Dominican Republic.	Mora JO, Pielmeier N, Ávila P. & Ramírez M. (1988) Santo Domingo Institucional (USAID)	663 Madres de niños menores de 3 años, distribuidas en una muestra aleatoria de 18 comunidades beneficiarias del programa y 18 comunidades no beneficiarias (grupo control)	Diseño Cuasi- Experimental con grupo control y pre-post. (Matching). Medidas repetidas. Se utilizó información de la evaluación interna del programa realizado por CARITAS y se revisó documentos e informes. Se entrevistó a informantes clave. Se aplicó un cuestionario de actitudes para madres. Se usó Chi Cuadrada y Análisis de Varianza.	Se encontró que el programa redujo significativamente la prevalencia de niños malnutridos en República dominicana
Comunicación y transporte	Honduras Rural Roads: Old Directions and New	Maxwell, Chapin, Demetre & Fletcher (1981) Honduras Institucional (USAID)	Dos rutas de carretera. Informantes clave y agricultores (no menciona número).	Método cualitativo (estudio de casos). Evaluación expost facto. Revisión de archivos del Proyecto.	El proyecto favoreció la producción agrícola y el ingreso económicos de los agricultores.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				Entrevista a informantes clave y a agricultores.	
Titulación de tierras	Evaluación del impacto en los beneficiarios del programa de titulación masiva de tierras.	Hevia, H. (2001) Honduras Institucional (USAID)	473 productores titulados seleccionados aleatoriamente.	Diseño pre-experimental sin grupo control ni línea base (diseño encuestas). Se aplicó una encuesta de campo. Entrevistas estructuradas con los responsables de las agencias intermedias financieras.	Según percepción de los beneficiarios el programa ha incrementado su seguridad en la legalidad de sus tierras y el uso que hacen de ellas.
Infraestructura - Agricultura	Sederhana: Indonesia Small-Scale Irrigation	Holloran S, Corey GL & Mahoney T. (1982) Indonesia Institucional (USAID)	29 subproyectos apoyados por AID. Agricultores. Sin muestreo ni especificación de número.	Método cualitativo (observación). Entrevista a agricultores y observación de campo.	Se encontró que el programa ha incrementado y desarrollado el bienestar de los pequeños agricultores propietarios en Indonesia.
Agricultura	The Wheat Development Program	Jonson WF, Ferguson CE. & Fikry M. (1983) Túnez Institucional (USAID)	98 funcionarios, autoridades y granjeros.	Método cualitativo (entrevistas a profundidad y análisis documental). Se utilizó información de documentos y archivos institucionales. Entrevista a informante clave.	Se encontró un impacto positivo del programa.
Electrificación y apoyo social	Ecuador: Rural electrification	Kessler JL. Ballantyne J, Maushammer R. & Romero N.(1981) Ecuador Institucional (USAID)	3200 personas comprendidas en 600 familias.	Diseño pre-experimental (diseño encuestas sin grupo control ni línea base). Se aplicó un cuestionario con 38 preguntas cerradas. Estadísticas descriptivas.	Se encontró impacto del programa en el desarrollo del comercio y servicios en las poblaciones beneficiadas.
Comunicación y transporte	Philippines: Rural Roads I and II.	Levy I, Zuvekas C. & Stevens C. (1981).	8 caminos de una muestra de 43. Muestreo intencional.	Diseño pre-experimental (entrevistas sin grupo control ni línea base).	El programa ha tenido un gran impacto en el desarrollo agrícola pero no de manera

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Filipinas Institucional (USAID)	181 agricultores y pescadores.	Aplicación de formularios y encuestas mediante entrevistas individuales y grupales. Estadísticas descriptivas.	equitativa.
Educación	Does Child Labor Displace Schooling? Evidence on Behavioral Responses to Enrollment Subsidy	Ravallion, Martin, and Quentin Wodon. (1999) Bangladesh Institucional	Se utilizó la muestra rural del Household Expenditure Survey conformada por niños de 5 a 16 años.	Información obtenida en el Bangladesh Bureau Statistics (1995 - 1996). Diferencia de medias, variables instrumentales y otros métodos de emparejamiento.	El programa tuvo un efecto significativo sobre la participación laboral de los niños. Asimismo, el programa tuvo un efecto significativo y positivo sobre la probabilidad de estar en la escuela.
Agricultura y transporte	Agricultural Transportation Assistance Program (ATP)	Lucas K, Davis T. & Rinkard K. (1996) Tanzania Institucional (USAID)	Revisión de documentos de 20 regiones. 64 informantes clave.	Método cualitativo (observación de campo). Revisión de documentos escritos. Entrevista a informantes clave. Observación de campo. Discusión con los encargados del programa y con otros implicados en el desarrollo del sector transporte.	Se halló que el impacto del programa ha sido positivo y creciente desde sus inicios.
Agua potable	Evaluation of the Provincial Water Project in the Philippines	Magnani R, Tourkin S. & Hartz S. (1984) Filipinas Institucional	La selección de la muestra no fue aleatoria. 2500 Familias Se determinó grupo de tratamiento y grupo control no equivalente.	Diseño de evaluación cuasi-experimental. Se utilizó información proveniente del estudio de hogares mantenidos por los distritos de Bacolod y Cagayan de Oro. El impacto del programa se determinó a través de la comparación de medias y Regresión logística.	Se encontró beneficios positivos del programa
Comercio	Impact Evaluation on	Makanza M. &	21 / 50 Agentes de aduanas.	Estudio de encuestas	Se encontró que la

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Implementation of the WTO evaluation agreement for Mozambique	Munyaradzi R. (2004) Mozambique Institucional (USAID)		(descriptivo). Aplicación de cuestionarios a agentes de aduanas. Entrevista a informantes clave.	Intervención es beneficiosa para la administración de los clientes.
Electrificación y apoyo social	The Philippines: Rural Electrification	Mandel DH, Allgeir PF, Wasserman G, Hickey G, Salazar R. & Alviar J. (1980) Filipinas Institucional	7 cooperativas.	Método cualitativo (observación) Entrevista no estructurada a pobladores y a informantes clave. Análisis de información en registros de compañías eléctricas	Se encontró un impacto limitado del programa
Salud reproductiva	Evaluation of AID Family Planning Programs: Ghana Case Study.	McClelland DG, Johnston AG, Pedersen BR, Schwartz JB. & Schmeding R.(1993) Ghana Institucional (USAID)	4 regiones. 110 pobladores de Ghana. 49 visitas institucionales.	Método mixto. Estudio de casos. Revisión bibliográfica, entrevista a informantes clave, entrevista a pobladores de cuatro regiones de Ghana, aplicación de escala de 30 ítems y técnica de grupos focales (25).	Los programas tuvieron un impacto positivo en el incremento de los comportamientos de planificación familiar de los pobladores.
Educación	Liberian Primary Education Program (PEP):Final Impact Evaluation of the PEP Pilot Radio Broadcast.	McDowell J. & Moulton J. (1990) Liberia (África) Institucional	305 Profesores y directores de escuela. Se utilizó grupo control	Diseño de evaluación pre post. Se controló la maduración, Aplicación de una escala de desempeño.	Se encontró efectos positivos del programa en la promoción de habilidades y la supervisión de los profesores.
Agua potable	Panamá: Rural Water	Meehan R. & Viveros-long A. (1982) Panamá Institucional	70 Madres con hijos menores de 5 años. Líderes o responsables gremios o comunidades.	Selección aleatoria de 26 comunidades beneficiarias del programa. Aplicación de cuestionario a las madres y otro a los líderes o responsables sociales.	Se encontró que el programa tuvo impacto en las comunidades donde se ejecutó.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Agricultura y apoyo social	Impact Evaluation: Resettlement Grant Activity. Emergency Recovery: Agricultura and Comercial Trade (ER:ACT)	Miller, J. (2002) Mozambique Institucional	630 beneficiados	Revisión de documentos Entrevista a informantes clave Encuestas a beneficiarios del programa. La selección de la muestra fue sistemática.	Se encontró un impacto favorable del programa en la economía de los beneficiarios.
Educación	U. S. Aid to Education in Paraguay: The Rural Education Development Project	Nicholson E, Miles HL, Johnson DN. (1983) Paraguay Institucional	No especifica	Revisión de documentos, discusión de procedimientos, entrevistas e informes escritos. Aplicación de cuestionarios a directores, profesores, alumnos y padres de familia.	Se encontró un impacto favorable del proyecto en la educación rural.
Energía y agua potable	The On-Farm Water Management Project In Pakistan	Painter JE, Balwin E, Ahmed AS, Siddiqui MA.& Khan MH. (1982) Pakistán Institucional	55 pobladores vinculados a la agricultura	Como criterio de comparación se utilizó el Estudio de Productividad Agrícola del Plan Regional de la Organización Sind. Entrevista y aplicación de cuestionario.	Se encontró impacto positivo del proyecto en el ahorro de agua y disponibilidad de energía.
Salud	An education Program in Alcoholism for Social Workers: Its Impact on Attitudes and Treatment-Orientede Behavior	Manohar V, Des Roches E, & Ferneau E. (1976) Canadá Journal	42 trabajadores sociales	Se aplicaron tres cuestionarios: A, B y C. El grupo C recibió el tratamiento.	El programa educacional provocó un plan de referencia que facilitaría la circulación de los pacientes de la sala de hospital hacia la clínica de Alcoholismo de Boston City Hospital.
Suministro de agua y alcantarillado	Does Piped Water Reduce Diarrhea for Children in Rural India?	Jalan, Jyotsna, and Martin Ravallion (2003) India	Población de 33,000 hogares. Se determinó grupo de tratamiento y grupo control.	Se consideró la información proveniente del National Council of Applied Economic Research de la India (1993-1994).	El acceso a agua de cañería disminuye significativamente la incidencia de diarrea, al igual que la duración de las enfermedades.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Journal		Se utilizó el estimador de Diferencia de medias, así mismo, se comparó los resultados utilizando el Propensity Score Matching individual frente al de la aldea.	
Educación	Impact Evaluation of Community-Based Coastal Resource Management Projects in The Philippines	Pomeroy RS, Pollnac RB, Predo CD. & Kanton BM. (1996) Filipinas Institucional	314 beneficiarios	Entrevista a beneficiarios Entrevista a informantes clave. Se utilizó las pruebas de hipótesis Chi cuadrada y T de Student para muestras apareadas e independientes, el análisis de componentes principales, análisis de correlación, regresión de mínimos cuadrados ordinaria, modelo de análisis de elección dicotómica (logit analysis)	El impacto del proyecto fue parcial
Salud Reproductiva	Evaluation of A.I.D. Family Planning Programs Tunisia Case Study	Rea S, Martin R, Young M, & Kress D. (1993) Túnez Institucional	Estudio de caso	Entrevistas a Informantes clave, observación en clínicas y otros establecimientos de salud, discusión en grupos focales.	El programa ha tenido impacto en la calidad de servicios brindados, en las competencias de los proveedores de servicios y del mecanismo encargado de la continuidad del uso de IUD.
Comunicación y transporte	Kenya: Rural Roads	Roberts JE, Clapp-Wincek C, Brokensha DW. (1982) Kenya Institucional	31 carreteras individuales	Visitas de campo. Entrevista a informantes clave.	Desde una perspectiva global, existe impacto del proyecto.
Agricultura	Philippines: BICOL Integrated Area Development	Sommer Jg, Aquino R, Fernández C, Gelay FH, Simmons E, Castillo GH. & Roco	Selección aleatoria de 70 agricultores	Entrevistas individuales a los agricultores	Se encontró un impacto parcial del programa.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Ch. (1982) Filipinas Institucional			
Agricultura	Korean Irrigation	Steinberg DI, Morrow RB. & Dong-il K.(1980) Korea Institucional	18 proyectos Se obtuvo una muestra aleatoria estratificada.	Se hicieron visitas de campo a los proyectos. Entrevista a un promedio de cien agricultores o a sus esposas. Discusión entre equipo de investigadores.	El proyecto ha tenido un gran impacto en el bienestar de los agricultores.
Educación	Preliminary Impact Evaluation Report of NGO Educare Training in South Africa	Stevens FI.(1995) Sud África Institucional	32 profesores: 18 recibieron el tratamiento y 14 conformaron el grupo control.	- Cuestionario para los profesores capacitados. - Protocolo de observación del salón de clase. - Información etnográfica complementaria.	Existe un impacto positivo del programa, sin embargo esto se ve afectado por limitaciones metodológicas.
Educación	Improving girls' education in Guatemala	Stromquist NP, Klees S. & Miske S. (1999) Guatemala Institucional	No se menciona	Diseño de trabajo de campo. Entrevistas semi estructuradas, revisión de documentos, observación en escuelas y salones de clase.	Se encontró un impacto parcial del programa.
Educación	Promoting primary education for girls in Guinea	Sutton M, Tietjen K, Bah A & Kamano P. (1999) Guinea Institucional	No se precisa	Revisión documentaria, análisis de datos brindados por la Unidad de Planificación y Estadística de Guinea, entrevista a los actores políticos y del programa en la educación de Guinea y observaciones o entrevistas con los padres, profesores y administrativos de cuatro escuelas de primaria.	Se encontró un impacto positivo del programa

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Agricultura	Agricultural Policy Analysis and Planning: A summary of two recent analyses of a.i.d.-supported projects worldwide	Tilney JS & Riordan JT (1988) Multinacional Institucional	61 proyectos auspiciados por USAID	Se tomó información de dos estudios de evaluación de impacto: Ambos estudios utilizaron las evaluaciones realizadas por AID con el objetivo de determinar el impacto de los proyectos. La recolección de datos fue básicamente a través de censos agrícolas, estudio a familias y consumidores, estudios de producción, y la construcción y uso de otras bases de datos agrícolas.	Los 61 proyectos agrícolas fueron exitosos en sentido estricto para alcanzar sus propósitos, referido especialmente a su capacidad de construcción y capacitación.
Apoyo social	A Synthesis of four legacy/impact studies of USAID assistance to Cameroon	USAID (1994) Camerún Institucional	Proyectos de desarrollo social Auspiciados por USAID	Revisión de la documentación disponible del proyecto, entrevistas al personal clave del proyecto, observaciones a los proyectos en su lugar y la evaluación de los participantes del proyecto.	La intervención de USAID en Camerún ha generado cambios sociales y económicos exitosos y positivos en diversos sectores.
Educación (Salud, Nutrición y Población infantil)	Education and Child Labor: Experimental Evidence from a Nicaraguan Conditional Cash Transfer Program	Maluccio, John A. (2003) Nicaragua Institucional	Se seleccionaron 42 comarcas (pequeñas comunidades con alrededor de 100 familias) utilizando un índice de marginalidad, 21 de ellas recibieron se benefició del programa y las otras 21 se constituyó en el grupo control. Se obtuvo una muestra aleatoria de 1585 familias	Se utilizó la información disponible en el National Population and Housing Census (1995) Evaluación de Diseño aleatorio, se utilizó el estimador de Diferencia en la Diferencia	Se encontró evidencia de que el programa tuvo un efecto significativo y substancial en la matrícula escolar y reclutamiento durante su primer año de operación. Asimismo, permite una sustancial reducción del trabajo infantil para la población en edad escolar.
Educación y salud	Reinsertion of ex-combatants in El Salvador	Creative Associates International (1996)	1000 evaluados: ex combatientes, y grupo control de veteranos.	Se realizó una selección aleatoria estratificada de la muestra de estudio.	No se especifica

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		El Salvador Institucional		Entrevistas estructuradas Aplicación de cuestionarios Entrevista a grupos focales	
Educación	Ikhwelo Pilot Projects: Final Report	Ikhwelonet Consortium (2002) Sudáfrica Institucional	660 entrevistados	Diseño de evaluación longitudinal Entrevistas, encuestas guiadas y grupos de discusión, estudios de caso	Se encontró que el proyecto tiene un impacto positivo.
Capacitación laboral	Impact Evaluation; LAC/ Honduran Training Program.	Van Steenwyk, N. (1984) Honduras Institucional	40 entrevistados: 23 participantes y 17 supervisores	Entrevistas personales a los participantes y supervisores del programa, aplicación de cuestionario	Se encontró impacto positivo del programa
Educación	Project Impact: A low-cost alternative for universal primary education in the philippines	Wooten J, Jansen W, Kohashi Warren M.(1982) Filipinas Institucional	80 entrevistados	Entrevista a padres, profesores, administradores y agentes políticos de la educación local.	El programa obtuvo el impacto esperado, de reducir el costo familiar para la educación básica, además de otros beneficios.
Género y equidad	The impacto Progresas on women`s status and intrahousehold relations	Adato M, De la Brière B, Mindek D. & Quisumbing A. (2000) México Institucional	24, 077 familias Se determinó grupo de tratamiento y grupo control.	Se utilizó información de la Encuestas de las Características Socio Económicas de los Hogares. Diferencia en la diferencia Análisis de regresión	Se encontró un efecto positivo del programa al incrementar la toma de decisiones de la madre en su hogar,
Resolución de Conflictos	El impacto de Progresas sobre las relaciones sociales en la comunidad	Adato M. (2000) México Institucional	230 personas de 70 comunidades en 6 estados. Se determinó grupo de tratamiento y grupo control	Se desarrolló un estudio de evaluación de tipo cualitativo. Grupos focales Entrevistas semi estructuradas a médicos	La participación en el programa favoreció la cohesión social.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Educación y salud	Evaluación de operaciones de Progresá desde la perspectiva de los beneficiarios, las promotoras, directores de escuela y personal de salud	Adato M, Coady D. & Ruel M. (1999) México Institucional	506 localidades: se determinó grupo de tratamiento (n = 2320) y grupo control (n = 186)	La selección de la muestra fue aleatoria. Encuestas cuantitativas de los beneficiarios (1997) Encuestas cuantitativas de las escuelas y clínicas de salud. Entrevista cualitativa semi-estructurada de las escuelas y de las clínicas. Grupos focales a beneficiarios y no beneficiarios.	No se especifica
Educación	Do Crowded Classrooms Crowd Out Learning? Evidence From the Food for Education Programme in Bangladesh	Ahmed AU. & Arends-Kuenning M.(2003) Bangladesh Institucional	600 hogares en 60 aldeas en 30 uniones en 10 thanas, y 110 escuelas en las mismas 30 uniones de las cuales se había extraído la muestra de los hogares. Se determinó grupo control	Estudio de evaluación de corte transversal. Encuestas a escuelas primarias. Administración de una prueba de logro académico a los estudiantes beneficiarios y control. Análisis multivariable	El programa influyó negativamente en el rendimiento promedio de los alumnos el cual sería resultados del efecto compañero negativo.
Nutrición y pobreza	The impact of The International Food Policy Research Institute's research program on rural finance policies for food security for the poor.	Alwang J. (2002) Multinacional (África y Asia) Institucional	80 informantes clave de cuatro países beneficiarios	Diseño de evaluación de estudio de caso y econométrico. Entrevista a informante clave.	Se encontró que el programa tuvo resultados e impacto destacados.
Educación	El impacto de Progresá sobre el rendimiento escolar durante el primer año de operación	Behrman JR, Segueta P. & Todd PE. (2000) México Institucional	500 localidades elegidas aleatoriamente que comprenden a familias pobres que tienen niños inscritos entre 3° y 6° de primaria y 1° a 3° de secundaria Se determinó grupo de	Pruebas de evaluación estandarizadas. Diferencias de medias y proporciones. Censos de 1990 y 1995	A excepción de los los estudiantes de 12 a 14 años, no se encontró un impacto significativo del programa a corto plazo.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
			tratamiento y control.		
Educación	Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA	Attanasio, Orazio; Meghir, Costas; Santiago, Ana (2005) México Institucional	Se utilizó una muestra aleatoria de 506 aldeas: de donde 320 conformaron el grupo de tratamiento y 186 de control.	La información fue obtenida de la línea base de los estudios de hogares (1997, 1998, 1999 y 2000). Diferencia de medias. Diferencia en la diferencia. Modelo de Participación Escolar. Simulación a través del Modelo de Participación Escolar.	Se encontró que el programa tiene un efecto positivo sobre la matrícula de niños, especialmente luego de la escuela primaria.
Educación y salud	Evaluación del impacto de Progresá en la talla del niño en edad preescolar	Behrman JR. & Hoddinott J. (2000) México Institucional	693 niños de 12 a 36 meses de edad. Se determinó grupo de tratamiento y grupo control.	Selección aleatoria de la muestra según estado o localidad. Valores Z. Análisis de Regresión Estimaciones transversales con mínimos cuadrados ordinarios. Estimaciones de los efectos fijos individuales del niño.	Se encontró un impacto significativo que indica del programa en el aumento el crecimiento del niño y reduce la probabilidad de desmedro.
Pobreza y apoyo social	The cost of poverty alleviation transfer programs: a comparative analysis of three programs in Latin America	Caldés N, Coady D. & Maluccio JA. (2004) Multinacional: México, Honduras y Nicaragua Institucional	Tres programas de alivio de la pobreza: Programa Nacional de Educación, Salud y Alimentación (PROGRESA); Programa de Asignación Familiar (PRAF) y Red de Protección Social (RPS)	Estudio de caso Evaluación a través del análisis de la Tasa de Costo de Transferencias (Coady, 2004)	Se encontró que dos de los programas presentan un costo de ejecución bajo.
Pobreza	La aplicación del análisis social costo-beneficio a la evaluación de Progresá	Coady D. (2000) México Institucional	14,856 niños beneficiarios Se determinó grupo de tratamiento y grupo control	Diseño de evaluación pre post. Análisis de regresión Modelo de análisis costo-beneficio.	Se encontró que Progresá es un programa de costo eficiente para conseguir que más niños ingresen a la escuela secundaria.
Educación	A cost-effectiveness analysis of demand- and supply-side education interventions: the case of	Coady DP. & Parker SW. (2002) México	24. 077 familias comprendidas en 506 comunidades elegidas de manera aleatoria.	Estudio de evaluación pre post. La selección de los hogares fue de manera aleatoria.	Se encontró que los subsidios a las demandas educativas son substancialmente más costo-efectivos que la expansión

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Progresá in México	Institucional	Se determinó grupo de tratamiento y grupo control.	Estimadores de regresión de doble diferencia. Análisis de costo efectividad.	educativa.
Pobreza y apoyo social	Análisis del equilibrio general del impacto de las transferencias de Progresá sobre el bienestar	Coady DP. & Lee R. (2000) México Institucional	No se especifica	La información se obtuvo de la matriz de contabilidad de la economía nacional de México (MCEN) con datos de 1996. Análisis de simulación Modelo de equilibrio general calculable.	Los resultados indican ganancias substanciales en el bienestar, derivadas al expandir el programa para incluir a los pobres urbanos.
Nutrición y empleo	Proyecto urbano "alimentos por trabajo" de CARE	Garrett J. (2001) Etiopía Institucional	4 de los 25 barrios beneficiarios del programa	Enfoque de evaluación cualitativa. Entrevista a participantes clave. Revisión de documentación disponible.	El proyecto tuvo un impacto positivo en las comunidades beneficiarias.
Salud y nutrición	El Impacto del Programa de Educación Salud y Alimentación (Progresá) sobre la Salud	Gertler P. (2000) México Institucional	24,000 hogares correspondientes a 125,000 individuos. Se determinó grupo de tratamiento y grupo control.	Diseño de evaluación experimental. Asignación aleatoria de las localidades en tratamiento y control. Aplicación de encuestas cada seis meses por el espacio de dos años. Estimador de doble diferencias.	Se encontró que el programa incrementó significativamente el uso de los servicios de salud de las clínicas públicas para medidas de salud preventiva.
Pobreza y desigualdad social	Pobreza, desigualdad y efectos indirectos del Programa de Educación, Salud y Alimentación	Handa S, Huerta MC, Pérez R. & Straffon B. (2000) México Institucional	506 localidades. Se determinó grupo de tratamiento y grupo control.	Evaluación pre – post. Ecuaciones de regresión de Mínimos Cuadrados Ordinarios (MCO). Aplicación del cuestionario ENCEL 980.	Se encontró un efecto positivo del programa en la disminución de la pobreza y de las desigualdades. Asimismo, se encontró efectos indirectos importantes en términos de la atención a la

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				Diferencia de medias. Estadística T Diferencia en las diferencias	salud de los hogares no beneficiarios.
Pobreza y apoyo social	El impacto de Progresá sobre el consumo: informe final	Hoddinott J, Skoufias E. & Washburn R. (2000) México Institucional	24,000 hogares de 506 localidades seleccionadas por Progresá. Se determinó grupo de tratamiento y grupo control.	Las localidades se seleccionaron de manera aleatoria. Se utilizó información de las encuestas de evaluación ENCEL980, ENCEL99M y ENCEL99N. Se utilizó encuestas de seguimiento. Métodos de regresión.	Se encontró que los beneficiarios del programa mostraron mayores niveles de gastos por consumo, orientados éstos principalmente a productos nutritivos.
Empleo y educación	El impacto de Progresá sobre el trabajo, el ocio y el uso del tiempo	Parker SW. & Skoufias. E. (2000) México Institucional	24,077 hogares comprendidos en 506 localidades. Se determinó grupo de tratamiento y grupo control	Estudio de evaluación pre-post Se utilizó información obtenida de: ENCASEH97, ENCEL98N, ENCEL99M, y la ENCEL99N. Se aplicó una encuesta sobre el uso del tiempo. Modelo de estimación de doble diferencias.	Se encontró un impacto significativo del Programa en la disminución de participación laboral de los niños. No obstante, no ha habido incremento de sus hora libres.
Desarrollo urbano	Property Rights for the Poor: Effects of Land Titling	Galiani, Sebastian, and Ernesto Scharfgrösky. 2005 Argentina Institucional	467 familias seleccionadas aleatoriamente las cuáles conformaron los grupos que recibieron los títulos de propiedad (grupo de tratamiento) y los que no recibieron estos títulos (grupo control)	Selección aleatoria Métodos de evaluación a través de la diferencia de medias, variables instrumentales y el emparejamiento por Propensity Score Matching.	El programa causó un incremento significativo de 28% en el fomento de la vivienda.
Educación y apoyo social	El Programa de Guarderías Comunitarias de Guatemala: Focalización Efectiva de la	Ruel MT. (2003) Guatemala	1,363 hogares con niños de 0-7 años de edad en el área de estudio.	La muestra de estudio se seleccionó de manera aleatoria.	El programa tiene un impacto positivo substancial sobre la dieta de los niños.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Ayuda en Alimentos en Áreas Urbanas	Institucional	Se determinó grupo de tratamiento y grupo control		Complementa también las estrategias de subsistencia de las mujeres pobres de zonas urbanas que trabajan fuera del hogar.
Agricultura	Assessing the impact of Rice Policy Changes in Viet Nam and the contribution of policy research	Ryan JG. (1999) Viet Nam Institucional	3,126 evaluados distribuidos en 17 provincias.	Entrevistas profundas a beneficiarios del programa Se utilizó las estadísticas de series de tiempo de los distritos respecto a su producción agrícola, área, precios de tierras, temporada para la exportación de arroz. Se aplicó cuestionarios. Análisis de regresión, análisis de series de tiempo.	Se encontró que los beneficios económicos obtenidos por la política sólo se aprecian en el largo plazo y de manera evidente.
Nutrición y educación	The Contribution of IFPRI Research and the Impact of the Food for Education Program in Bangladesh on Schooling Outcomes and Earnings	Ryan JG. & Meng X.(2004) Bangladesh Institucional	7, 445 familias y 38,563 individuos. Se determinó grupo de tratamiento y grupo control.	Metodología de evaluación de tipo experimental. Se utilizó información del estudio nacional sobre ingreso y gasto familiar. Propensity Score Matching y Diferencia en las diferencias.	Se encontró un importante efecto positivo del programa sobre la educación e ingreso de los niños y jóvenes participantes, especialmente las mujeres.
Educación	El impacto de Progres a sobre la inscripción escolar	Schultz TP. (2000a) México Institucional	19, 176 niños Se determinó grupo de tratamiento y control.	La selección de las localidades donde opera el programa fue aleatoria. Estudio de evaluación pre_post. Se utilizó información de la encuesta de evaluación ENCEL Estimadores de dobles diferencias. Modelo Probit.	El programa tuvo un impacto positivo sobre la inscripción escolar.
Pobreza y apoyo social.	Is progres a working? summary of the results of	Skoufias E. & McLafferty B. (2001)	24,000 hogares comprendidos de 506 localidades de siete	Estudio de evaluación de tipo cualitativo.	Se encontró un impacto positivo del programa en los

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	an evaluation by IFPRI.	México Institucional	estados	Grupos focales	hogares y comunidades beneficiadas.
Pobreza y apoyo social	Una evaluación del impacto de los apoyos en efectivo de Progresá sobre las transferencias privadas entre los hogares	Teruel G. & Davis B. (2000) México Institucional	24,077 familias, distribuidos en 506 comunidades asignadas de manera aleatoria como grupo de tratamiento y grupo control	Se empleó información de la ENCASEH97 y el conjunto de datos ENCEL. Análisis econométricos. Análisis de regresión.	No se encontró que progresá estaría afectando de manera negativa el apoyo económico de los hogares proveniente de fuentes privadas.
Salud y protección social	The effectiveness of NGO campaigning: lessons from practice	Chapman J. & Fisher T. (2000) Ghana e India Journal	Estudio de dos casos de campañas: La promoción de la lactancia materna en Ghana y la campaña contra el trabajo infantil en las industrias de alfombra en la India.	Se examinó documentos, militantes de las organizaciones y a los beneficiarios. Entrevista semi-estructurada, recortes de diarios.	Se encontró que en las dos campañas se obtuvieron considerables éxitos, sin embargo, esto no sólo sería efecto e la campaña misma.
Apoyo social	Evaluating Decentralized Social Sector Programs: Evidence from Morocco's BAJ	Jacoby HG. (2000) Marruecos Institucional	25,000 personas comprendidas en 3,827 familias.	Diseño de evaluación pre-post. Se utilizó información del Moroccan Living Standard Surveys (1990-91 y 1998-1999) Diferencia en las diferencias. Modelo de Análisis Probit.	Se encontró que los programas no brindaron significativas ganancias a las provincias beneficiarias.
Salud y protección social	An Occupational Health Programme for Adults and Children in the Carpet Weaving Industry, Mirzapur, India: a case study in the informal sector.	Das PK, Shukla KP. & Óry FG. (1992) India Journal	260 personas: 200 tejedores y 60 no tejedores.	La selección de la muestra de tejedores fue aleatoria. Se emparejó la muestra de tejedores con la muestra de no tejedores según edad y situación socioeconómica. Se aplicó cuestionarios.	Se encontró que el programa de salud ocupacional permitió brindar apoyos significativos al hacer cambios técnicos en el ambiente de trabajo y organizar el apoyo de la comunidad para los servicios de salud. Incrementado también la oportunidad de que los niños trabajadores tengan un mejor futuro al brindárseles educación formal o informal.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Educación y apoyo social	Effects on School Enrollment and Performance of a Conditional Transfers Program in Mexico	Dubois P, De Janvry A. & Sadoulet E. (2003) México Institucional	152,000 individuos comprendidos en 26, 000 familias, todos distribuidos en 506 comunidades. Se determinó grupo de tratamiento (320 comunidades) y grupo control (186 comunidades).	Diseño de evaluación experimental. La selección de las comunidades fue de modo aleatorio. Se utilizó la información obtenida por ENCASEH97, ENCEL. Análisis de regresión	Se encontró que el programa tiene un impacto positivo en la continuación educativa, respecto a su impacto en el desempeño fue positivo para el nivel primario y negativo para el nivel secundario.
Empleo	Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching Programa Trabajar II	Jalan, Jyotsna; Ravallion, Martin (2003) Argentina Institucional	2,802 participantes del programa Trabajar (Grupo de tratamiento)	Encuesta de Desarrollo Social (1997) Diferencia en medias y Propensity Score Matching	La ganancia promedio es de 103 dólares, lo cual es la mitad del promedio del salario del programa Trabajar. Las ganancias de ingreso son significativamente mayores para el caso de los jóvenes (15 – 24 años)
Educación	Outcomes of Mandated Preventive Services Programs of Homeless and Truant Children: A Follow-up Study.	Twaite JA. & Tirado D. (1997) Estados Unidos Journal	100 adolescentes que participaron del programa de readaptación.	Se obtuvo información de dos fuentes: registros de las familias de cada adolescente y de la entrevista telefónica a los profesionales del programa. Protocolo conteniendo escalas de medición del comportamiento psicopatológico. Correlación de Pearson. Análisis de Varianza Análisis de Regresión Múltiple.	Se encontró un impacto favorable del programa en la intervención de adolescentes con conducta psicopática.
Salud y apoyo social	Women's Participation in Rural Credit Programmes in Bangladesh and Their Demand for Formal Health Care: Is There a Positive impact?	Nanda P. (1999) Bangladesh Journal	1758 familias distribuidas en 87 distritos Se determinó grupos de tratamiento y grupo control	La selección de los distritos fue aleatoria. Variables instrumentales	Se encontró un impacto positivo de la participación de las mujeres en los programas de crédito sobre sus demandas por una atención de su salud formal.
Educación	The Consequences of Government Program	Pitt MM, Rosenzweig MR. & Gibbons DM.	No se especifica	Se unió las diferentes bases de datos sobre estudios de familias y	Se encontró un impacto positivo de los programas de

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Placement in Indonesia	(1992) Indonesia Institucional		de diferentes períodos, realizados en Indonesia. Análisis de regresión. Razón T de Huber (1967) Mínimos cuadrados.	educación media y salud en la asistencia escolar de los adolescentes. Sin embargo, no se encontró evidencias en el efecto de los programas de planificación familiar.
Apoyo social	Letting Communities Take the Lead: A Cross-Country Evaluation of Social Fund Performance	Rawlings L, Sherburne-Benz L. & Van Domelen J. (2001) Multinacional Institucional	Seis proyectos de apoyo social. 65,000 familias, distribuidas en 1,200 escuelas, centros de salud y proyectos de agua y sanidad.	La selección de los proyectos fue de manera aleatoria. Se utilizó información de: Estudios de hogares, estudios institucionales, evaluaciones cualitativas, estudios de costo, información administrativa de los fondos sociales.	Se encontró un impacto positivo de todos los programas estudiados.
Educación	Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia	Angrist J, Bettinger E. & Kremer M. (2004) Colombia Institucional	4, 044 estudiantes Se determinó grupo de tratamiento y grupo control.	Diseño de evaluación natural experimental pre post. Los ganadores y perdedores del sorteo de becas conformaron los grupos de tratamiento y control respectivamente. Se aplicó la prueba ICFES Análisis de regresión.	Se encontró que el programa incrementó las puntuaciones en los exámenes de los alumnos beneficiarios así como su persistencia en cuanto a dicho rendimiento..
Educación	Remedying Education: Evidence from Two Randomized Experiments in India	Banerjee A, Cole S, Duflo E. & Linden L. (2005) India Institucional	5, 274 escolares de primaria, se estableció grupo de tratamiento y grupo control.	Se evaluó a los programas Balsakhi y Computer – Assisted learning. Muestreo aleatorio estratificado. Evaluación pre post. Variables instrumentales, Diferencia en las diferencias.	Se encontró un impacto positivo de los dos programas evaluados en cuanto el incremento del rendimiento académico de los niños.
Nutrición y educación	Iron Deficiency Anemia and School Participation	Bobonis GJ, Miguel E. & Puri C. (2004)	200 niños pre-escolares. Se determinó tres grupos de	Estudio de evaluación de cohorte longitudinal y diseño cuasi-experimental pre post.	Se encontró efectos positivos del programa en el incremento del peso de los niños y en su

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		India Institucional	tratamiento.	Selección no aleatoria de de los grupos de tratamiento. Se aplicó una Encuesta a Hogares. Puntuaciones Z., evaluación costo-beneficio.	asistencia a la escuela pre - escolar.
Educación y salud	Education and HIV/AIDS Prevention: Evidence from a randomized evaluation in Western Kenya	Duflo E, Dupas P, Kremer M. & Snei S. (2006) Kenya Institucional	328 escuelas. Se determinó grupo de tratamiento (163) y grupo control (165).	Asignación aleatoria de grupo de tratamiento y grupo control. Análisis de regresión	La intervención no produjo una disminución del embarazo adolescente pero sí incrementó la posibilidad que los embarazos adolescentes ocurran dentro del matrimonio.
Educación	Monitoring Works: Getting Teachers to Come to School	Duflo E. & Hanna R. (2006) India Institucional	120 escuelas Se determinó grupo de tratamiento (60) y grupo control (60)	Selección aleatoria del grupo de tratamiento. Cámara fotográfica con comprobación de fecha y tiempo con seguro de alteración. Variables instrumentales.	El programa tuvo impacto en disminuir significativamente la tasa de ausencia a clase de los profesores y el incremento en las notas de exámenes de los alumnos.
Salud	Relative Risks and the Market for Sex: Teenagers, Sugar Daddies and HIV in Kenya	Dupas P. (2006) Kenya Institucional	13,000 estudiantes de 328 escuelas. Se determinó grupo de tratamiento (2,500) y grupo control.	Diseño de evaluación experimental. La selección de las escuelas del grupo de tratamiento fue aleatoria Análisis de regresión, diferencia en las diferencias,	La campaña disminuyó la incidencia de embarazos de parejas adultas entre las adolescentes.
Educación	Evaluación de Impacto del Programa Desarrollo Integral Temprano (PIDI)	Ortiz, SS., Balcazar M, Robison N, Rojas R. & Alurralde L. (2004) Bolivia	8,750 entrevistas con el binomio madre niño, 3,500 correspondiente al grupo de tratamiento y 5,250 al grupo control.	Metodología cualitativa y cuantitativa. Técnica de grupos focales (10) y entrevistas a profundidad (50).	El programa tuvo un efecto positivo y significativo sobre la estimulación temprana de niños

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				Entrevistas Diferencia de medias y otras medidas de emparejamiento.	
Agricultura y desarrollo organizacional	"The Rockefeller Effect"	Gugerty M.K. & Kremer M. (2004) Kenya Institucional	80 grupos de mujeres. Se determinó grupos de tratamiento y grupos control	Estudio de evaluación prospectivo. Entrevista	No se encontró que el fondo de ayuda promueva la solidez y funcionamiento grupal.
Educación	Decentralization: A Cautionary Tale	Kremer M, Moulin S. & Namunyu R. (2003) Kenya Institucional	14 escuelas Se determinó grupo de tratamiento y control	Asignación aleatoria de las escuelas a los grupos de tratamiento y control. Modelo Becker-Rosen Análisis de regresión	El programa incrementó la permanencia escolar y también generó demanda de escolares de otras escuelas cercanas. Se encontró que la participación escolar es sensible a su costo.
Educación	Computer-Assisted Learning: Evidence from a Randomized Experiment	Linden L, Banerjee A. & Duflo E. (2003) India Institucional	111 escuelas, se determinó grupo de tratamiento y grupo control.	Diseño de evaluación pre post. Análisis de regresión,	Se encontró que el programa es muy efectivo.
Salud	Randomised controlled trial of an adolescent sexual health programme in rural Mwanza, Tanzania	Obasi A et al. (2003) Tanzania Institucional	9,654 adolescentes, se determinó grupo de tratamiento y grupo control	Estudio de diseño de cohorte longitudinal.	Se encontró efectividad del programa en cuanto al cambio de conocimientos, actitudes y comportamiento de los adolescentes a favor de una mejor salud sexual.
Apoyo social	Women and Micro-Finance Programs	Anolin ALC. (2000) Filipinas	Estudio de caso	Se utilizó tres metodologías de investigación: revisión de la literatura, conducción de grupos de discusión focal y entrevista a	Se encontró que el programa tuvo impacto favorable en la economía y en el empoderamiento de las

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional		informantes clave.	participantes.
Salud y educación	Break the Silence Talk about AIDS	Awasum D, Sienché C. & Obwaka E. (1999) Kenya Institucional	750 jugadores y aficionados.	Entrevistas	Se encontró que la campaña resultó ser una estrategia efectiva y oportuna.
Apoyo social	The Impact of Three Microfinance Programs in Uganda	Barnes C, Gaile G. & Kibombo R. (2001) Uganda Institucional	1,332 personas, se determinó grupo de tratamiento y grupo control.	Evaluación de los programas: Foundation for International Community Assistance, Foundation for Credit and Community Assistance, y Promotion of Rural Initiatives and Development Enterprises. La selección aleatoria de la muestra de estudio. Cuestionario. Análisis de Varianza y Chi Cuadrada	Se encontró que los programas alcanzan a sus grupos objetivo. Así mismo tiene un impacto positivo.
Educación	Measuring the Effects of Behavior Change Interventions in Burkina Faso with Population – Based Survey Result	Baya B, Sangli G. & Maiga A. (2004) Burkina Faso. Institucional	720 encuestados distribuidos en 145 Aldeas. Se determinó grupo de tratamiento y grupo control mujeres y padres en espera de un hijo o con hijos menores de 12 meses, así como líderes comunitarios	Estudio de evaluación pre -post. Selección aleatoria por estratos de las aldeas de estudio. Entrevista cualitativa. Seis cuestionarios.	Se encontró que los efectos del programa en la población beneficiaria fue efectiva.
Nutrición y educación	Medium-Term Effects of the Oportunidades Program Package, including Nutrition, on Education of Rural Children Age 0-8 in 1997	Behrman JR, Parker SW. & Todd PE. (2004) México Institucional	2,500 habitantes distribuidos en 320 comunidades. Se determinó grupo de tratamiento y grupo control	Se utilizó información de Rural Evaluation Survey (ENCEL, 2003) Diferencia en las Diferencias, Se utilizó el emparejamiento de	Se encontró que el programa tuvo un mayor impacto con los niños que pertenecen a un grupo mayor de edad.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				grupos para evaluar el impacto del programa.	
Empleo y salario	The Impact of Minimum Wages in México and Colombia	Bell. LA (1995) México y Colombia Institucional	Trabajadores de empresas industriales formales.	Se utilizó información panel sobre las empresas industriales del sector formal. Estimador de mínimos cuadrados ordinarios, análisis de regresión	Se encontró que el salario mínimo no es efectivo para el sector forma en México pero sí lo es para el caso de Colombia.
Pobreza	Evaluación de Impacto en Programas de Superación de Pobreza, el Caso del Fondo de Inversión Social (FOSIS) de Chile	Paker C, Rivas G. & Cauas E. (1999) Chile Institucional	Se procedió a seleccionar 117 proyectos ejecutados en el año FOSIS 1997 en 16 comunas (8 de la VI Región y 8 de la IX Región), sobre los cuales se hizo un análisis descriptivo en carpetas de archivo. Para un análisis en profundidad se seleccionó una muestra estratificada intencional de 6 comunas en las cuales se estudiaron la totalidad de localidades en las cuales hubo intervención FOSIS durante 1997	Evaluación expost con enfoque territorial. Diseño metodológico comparativo, con inclusión de técnicas cuantitativas y cualitativas. Se utilizó la entrevista a informantes calificados	Los resultados revelan la importancia del concepto de pobreza - capacidades para asumir tareas de superación de pobreza. También esos procesos validan un modelo metodológico de evaluación de impacto territorial de programas sociales, ofreciendo adicionalmente un conjunto de indicadores multidimensionales, herramientas, y procedimientos metodológicos para medir impacto en comunidades pobres.
Pobreza y apoyo social	Comparing Effects of General Subsidies and Targeted Transfers on Poverty: Robustness Analysis Using Data Set from Tunisia	Bibi, S. (1998) Túnez Institucional	No se especifica la dimensión de familias	Se utilizó información del estudio de hogares de 1990. Análisis de regresión no paramétrica. Se utilizó la simulación	Se encontró que el programa de subsidios generales no tiene un impacto significativo a favor del alivio de la pobreza, a diferencia de un proyecto de transferencia focalizado.
Pobreza y apoyo social	Conditional Cash Transfers, Schooling and Child Labor: Micro-Simulating Bolsa Escola	Bourguignon F, Ferreira HG. & Leite PG. (2003) Brasil	60,000 familias	Se utilizó información del estudio de hogares PNAD de 1999. Método de simulación para evaluar el impacto del programa brasileño <i>Bolsa Escola</i> en el	Se encontró un fuerte impacto del programa en la asistencia escolar, pero una ausencia de impacto en la reducción de los niveles de pobreza e inequidad.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional		bienestar y comportamiento.	
Agricultura y apoyo social	Land reform and its Impact on Livelihoods: Evidence from eight land reform groups in the Northern Cape Province of South Africa	Bradstock, A. (2005) Sudáfrica Institucional	118 Familias beneficiarias provenientes de ocho grupos de reformas de tierras.	La muestra fue obtenida de manera aleatoria. Informantes clave, entrevistas de grupos focales y clasificación socio-económica. Ingreso per cápita	No se encontró evidencia de un impacto favorable del programa en el ingreso familiar.
Educación	A Longitudinal Study of the Impact of Integrated Literacy and Basic Education Programs on Women's Participation in Social and Economic Development in Nepal	Burchfield S, Hua H, Baral D. & Rocha V. (2002) Nepal Institucional	1000 mujeres Se determinó grupo de tratamiento y grupo control.	Diseño de investigación cualitativa y cuantitativa de tipo longitudinal pre post. Entrevistas a profundidad a una sub muestra de 20 mujeres. Se utilizó métodos de triangulación.	Se encontró que participaron de los programas obtuvieron mayores puntuaciones que las del grupo control.
Convivencia social	Jamaa Wazima Project: A Household Livelihood Security Project in Siaya and Busia Districts	CARE (2004) Kenya Institucional	800 personas, se determinó grupos de tratamiento y grupo control.	Diseño de investigación cualitativo y cuantitativo. Se utilizó una selección de muestra sistemática y por estratos. Grupos de discusión Focal, entrevistas a informantes clave, visitas de campo y cuestionarios. Análisis de contenido, criterios y de tendencias.	Se encontró que el programa tuvo efectos positivos en las actividades que implicó el trabajo en grupo con entidades: agua y agricultura.
Salud	Evaluation of GAVI Immunization Services Support Funding	Chee G, Fields R, Hsi N. & Schott W, (2004) Multinacional: países pobres Institucional	Seis países elegidos como estudios de caso	Revisión de documentos de GAVI, estudio de casos profundos.	No se obtuvo una diferencia significativa entre los países beneficiarios y los que conformaron el grupo control

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Pobreza y apoyo social	Community Programs and Women's Participation: The Chinese Experience	Coady D, Dai X. & Wang L. (2001) China Institucional	6,628 familias distribuidas en 57 aldeas. Se determinó grupo de tratamiento y grupo control	Se seleccionó aleatoriamente las aldeas de tratamiento y control. Técnicas de regresión multivariada de mínimos cuadrados ordinarios.	Se encontró que el programa incrementó substancialmente la participación e ingreso familiar de las mujeres y también generó beneficios sociales positivos.
Educación	Monetary Incentives for Teachers and School Performance. The evidence for Chile	Contreras D, Flores L. & Lobato F. (2003) Chile Institucional	5, 044 escuelas. Se determinó grupo de tratamiento y grupo control	Se utilizó un modelo de evaluación econométrico. Suma de cuadrados de dos etapas	Se encontró evidencia del efecto positivo del programa en el desempeño de la enseñanza.
Salud y educación	Evidence of behaviour change following a hygiene promotion programme in Burkina Faso	Curtis V, Kanki B, Cousens S, Diallo I, Kposehouen A, Sangaré M. & Nikiema M. (2001) Burkina Faso (África) Journal	500 madres de niños de 0 a 35 meses de edad.	Evaluación pre post. Observación estructurada de los comportamientos de higiene. Se utilizó el análisis de regresión logística, <i>The Huber-White sandwich estimator</i>	Se encontró que el programa es efectivo respecto a los cambios de conducta generados.
Apoyo social	Access to Credit and Its Impact on Welfare in Malawi	Diagne A. & Zeller M. (2001) Malawi (África) Institucional	4.699 familias distribuidas en 45 aldeas. Se determinó grupos de tratamiento y grupo control	Cuestionario sobre créditos y ahorros. Análisis econométricos y de tendencia.	Se encontró un gran efecto positivo, aunque no estadísticamente significativo, del programa en el bienestar de las familias.
Empleo	Evaluación del Impacto de los Programas de Bolsa Escolar en el Trabajo Infantil en Brasil	Ferro, Andrea, Rodrigues y Ana Kassouf (2004) Brasil	La muestra fue de 378837 niños.	Se utilizó el método experimental probit, en el cual la variable dependiente es la decisión de trabajar, que asume el valor uno cuando el niño trabaja y cero cuando no trabaja.	Tuvo impacto significativo.
Salud y apoyo social	Randomized trial of a statewide home visiting program to prevent child abuse: impact in reducing	Duggan A, Fuddy L, Burrell L, Higman SM, McFarlane E, Windham A. & Sia C. (2002)	643 familias. Se determinó grupo de tratamiento y grupo control	Entrevista anual de evaluación clínica a madres.	El programa de visita a hogares no disminuyó los factores de riesgo mayores hacia el abuso infantil.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	parental risk factors	Estados Unidos Institucional			
Agricultura y apoyo social	The Impact of Land Reform on Commercial Farm Workers' Livelihoods	Farm Community Trust of Zimbabwe (2001) Zimbawe Institucional	Cinco haciendas comerciales en el Este y Oeste de Mashonaland	Diseño de evaluación de estudio de caso. La metodología del Enfoque de Economía Familiar. Grupos de discusión focal y entrevistas a informantes clave. Guía de entrevista semi-estructurada.	Los efectos de la reforma varió de manera significativa según las haciendas estudiadas.
Pobreza y Apoyo social	Do Microfinance Programs Help Families Insure Consumption Against Illness?	Gertler P, Levine DI. & Moretti E. (2002) Indonesia Institucional	Estudio panel de 7,000 familias.	Se utilizó información del Estudio de la Vida Familiar en Indonesia (IFLS). Aplicación de cuestionario, auto reportes.	Se halló la importancia de las instituciones de préstamos bancarios en ayudar a las familias para asegurarse ante los riesgos a su salud.
Educación	Empowering parents to improve education: evidence from rural Mexico	Gertler P, Patrinos H. & Rubio-Codina M. (2006) México Institucional	6,038 escuelas primarias rurales no indígenas. Se determinó grupo de tratamiento y grupo control.	Se utilizó la información administrativa de la cobertura de CONAFE (1991-2003), del Censo Escolar Mexicano Se utilizó el estimador de diferencia en las diferencias.	Se encontró un impacto positivo en la promoción de resultados a través la gestión basada en la escuela.
Salud	An Evaluation of Post-campaign Knowledge and Practices of Exclusive Breastfeeding in Uganda	Gupta N, Katende C. & Bessinger R. (2004) Uganda Journal	1,766 mujeres y 1, 057 varones.	Estudio de evaluación post test. La información se obtuvo del Estudio de Evolución DISH (1999). Se siguió un proceso de muestreo	Se encontró que la campaña tuvo un significativo efecto sobre el conocimiento de los adultos respecto a la lactancia materna exclusiva, sin embargo, no se encontró lo mismo respecto a su práctica.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				de dos etapas. Entrevistas personales. Técnicas de regresión logística múltiple.	
Nutrición	Change in Food and Nutrient Consumption Among the Ultra Poor: Is the CFPR/TUP Programme Making a Difference?	Haseen, F. (2006) Bangladesh Institucional	400 familias: 200 muy pobres y 200 no muy pobres.	Estudio de evaluación pre post. Cuestionario Entrevista	Se encontró un efecto significativo del programa en la calidad de dieta de las familias beneficiarias.
Pobreza y políticas de apoyo social	The Impact of Minimum Wage Legislation in Developing Countries where Coverage is Incomplete	Jones, P.(1997) Ghana Institucional	No se menciona.	Se utilizó información de: Yearbook of Labour Statistics, African Employment Report (1990), Penn World Tables 5.6. Análisis transversal y de series de tiempo.	Se encontró que las políticas de salario mínimo en Ghana durante 1970 y 1980 llevaron a la reducción del sector formal
Infraestructura y pobreza	Meta-Analysis to Assess Impact of Watershed Program and People's Participation	Joshi PK, Jha AK, Wani SP, Joshi L. & Shiyani RL. (2005) Multinacional Institucional	311 estudios de caso sobre programas de divisorias de aguas	Diseño de evaluación meta-analítico. Análisis costo beneficio.	Se encontró que los programas de divisorias de aguas están rejuveneciendo y revolucionando las regiones lluviosas.
Pobreza y apoyo social	Do the poor benefit from devolution policies? Evidence from Malawi's forest co-management program.	Jumbe Ch.BL. & Angelsen A. (2006) Malawi Journal	404 familias	Se utilizó la información de hogares de las reservas forestales de Chimaliro y Liwonde. La muestra se seleccionó de manera aleatoria. Grupos de discusión focal y	Se encontró una fuerte evidencia del impacto del programa en el incremento de la ganancia económica de las mujeres y de los participantes de bajos ingresos.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				entrevista a informantes clave. Propensity Score Matching y las Técnicas de Descomposición.	
Salud y educación	Impact of Sex and HIV Education Programs on Sexual Behaviors of Youth in Developing and Developed Countries	Kirby D, Laris BA. & Roller L. (2005) Multinacional Institucional	83 evaluaciones del programa	Estudio de evaluación meta-analítico.	Se encontró que los programas tuvieron mucho más probabilidad de tener un impacto positivo que negativo en el comportamiento de sus participantes.
Educación	Evaluación del Impacto en las Instituciones Escolares de los Proyectos apoyados por el Instituto para la Investigación Educativa y el Desarrollo Pedagógico "IDEP", de Bogotá	Torres G, Isaza L, & Chávez LM. (2004) Colombia Journal	Se estableció un marco muestral de 140 proyectos de los 151 financiados entre 1997 y 2001. Se realizó un muestreo estratificados según año de convocatoria, modalidad de convocatoria, sujetos que se convocan y área temática. La muestra final es de 18 proyectos	Se estableció como fuente de información el investigador, el director del proyecto o coordinador del proyecto. Instrumentos: guías de análisis documental de las convocatorias y proyectos, guías de entrevista individual, guías para entrevistas en grupo focal y cuestionario para evaluar procesos cognitivos	Existe impacto de los proyectos en la fundamentación de las prácticas pedagógicas e investigativas de la mayoría de los participantes.
Empleo	The Impact of Training on Women's Micro-Enterprise Development.	Leach, F, Abdulla S, Appleton H, El-Bushra, Cardenas N, Kebede K, Lewis V. & Sitaram S. (2000) Multinacional (Etiopía, India, Perú y Sudan) Journal	78 mujeres	Se utilizó una metodología de estudio de casos. Entrevista estructurada y la observación.	Se encontró evidencia del impacto positivo que tiene el entrenamiento o capacitación de las mujeres en el incremento de sus ingresos y de su valoración personal.
Pobreza	Does Foreign Aid Reduce Poverty? Empirical Evidence from Nongovernmental and Bilateral Aid	Masud N. & Yontcheva B. (2005) Multinacional	Dos muestras para cada sub estudio (87 países y 76 países)	Análisis de Regresión Diseño meta-analítico	Se encontró que la ayuda de las ONG disminuye significativamente la mortalidad infantil pero no se encontró lo mismo respecto al

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional			analfabetismo.
Pobreza	Policy Expectations and Programme Reality: The Poverty Reduction and Labour Market Impact of Two Public Works Programmes in South Africa.	McCord, A. (2004) Sud África Institucional	4,792 personas pertenecientes a 676 familias. Se determinó grupo de tratamiento y grupo control	Diseño de evaluación de estudios de caso de dos programas: Limpopo y KwaZulu. Se utilizó información de Labour Force Survey (2003) La selección de la muestra fue aleatoria Aplicación de cuestionario. Diferencia en la Diferencia.	Se encontró impactos positivos de ambos programas
Salud	The impact on condom use of the "100% Jeune" social marketing program in Cameroon	Meekers D, Agha S. & Klein M. (2004) Camerú Journal	Dos muestras de jóvenes de 15 a 24 años evaluados en los años 2000 (n = 2,097) y 2002 (N = 3,536)	Análisis de regresión logística	La campaña de salud a través de los medios de comunicación e interpersonal tuvo un impacto positivo en la promoción del uso del condón y la prevención de las ETSs.
Medio ambiente	The social impacts of payments for environmental services in Costa Rica: A quantitative field survey and analysis of the Virilla watershed	Miranda M, Porras IT. & Moreno IL. (2003) Costa Rica Institucional	45 terratenientes. Se determinó grupo de tratamiento y grupo control	Se evaluó el programa a través del enfoque de Sustentos Sostenibles. Aplicación de cuestionario a través de visitas de campo.	Se encontró un impacto positivo del programa en el presupuesto familiar
Salud y apoyo social	An examination of Public Sector subsidies to the Private Health Sector: A Zimbabwe Case Study	Mudyarabikwa, O. (2000) Zimbabwe Journal	5 gerentes	Análisis cualitativo de la literatura y los documentos políticos sobre los sectores públicos y la salud privada. Entrevista a gerentes de instituciones	Se encontró evidencia de un mayor impacto de los subsidios sobre la equidad y los beneficios del consumidor si son directamente proporcionados al consumidor.
Pobreza y apoyo social	NAV Evaluation Report - The Integrated Rural Development Program 1994 – 2004 Thua Thien Hue Province.	Irene Norlund, Thu Nhung Mlo Duon Du, Ngo Huu Toan (2004) Vietnam Journal	17 comunidades beneficiarias.	Revisión de los documentos e informes de NAV. Evaluación participativa de las comunidades beneficiarias. Cuestionario cualitativo.	El impacto del programa ha sido considerable y mayor de lo que se había esperado.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Educación	Impact Study of the New Horizons Program in Egypt	North South Consultants Exchange (2003) Egipto Institucional	73 personas: 19 participantes para los grupos de discusión focal y 64 para las entrevistas.	Cuestionarios cualitativos. Grupos de discusión focal Entrevistas a profundidad.	Se encontró un impacto positivo del programa
Salud	Immunization Status and Child Survival in Rural Ghana	Nyarko P, Pence B. & Debpuur C. (2001) Ghana Institucional	17,701 niños	Se utilizó información del Sistema de Vigilancia Demográfica de Navrongo (1993). Se consideró la información respecto a los niños nacidos entre octubre el 1 de 1994 y el 31 de diciembre de 1999. Técnicas de regresión	Se encontró que el programa de inmunización tiene efectos positivos en la aumentar la sobre vivencia infantil en regiones de alta mortalidad infantil.
Salud	Paying Health Personnel in the Government Sector by Fee-For-Service: A Challenge to Productivity and Quality, and a Moral Hazard	Pannarunothai, S. & Kittidilokku, S. (1997) Tailandia Journal	2,943 profesionales de salud	Estudio de evaluación pre post. Grupos de discusión focal, entrevistas a profundidad. Se seleccionó una muestra a través de una técnica multietápica. Cuestionario postal auto administrado.	Se encontró que el sistema de pago por servicios de honorarios incrementó la productividad de los profesionales de servicios de salud.
Educación	Evaluating a Targeted Social Program When Placement is Decentralized	Ravallion, Martin, and Quentin Wodon (1998) Bangladesh Institucional	No se menciona	Deferencia de medias, variables instrumentales, otros métodos de emparejamiento.	La asistencia a la escuela se incrementó significativamente en 24% como resultado del programa
Salud	Impact of a Community-based Comprehensive Primary Healthcare Programme on Infant and	Perry HB, Shanklin DS. & Schroeder DG. (2003)	4,072 personas. Se determinó grupo de tratamiento y grupo control	Estudio de evaluación prospectivo. Se utilizó información de los	Se encontró que el programa de provisión basada en la comunidad fomentó la sobre vivencia infantil en las

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Child Mortality in Bolivia	Bolivia Journal		registros del programa. Diferencia de grupos	regiones pobres.
Salud, agua potable y sanidad	Social Returns From Drinking Water, Sanitation and Hygiene Education: A Case Study of Two Coastal Villages in Kerala	Pushpangadan, K.(2002) India Institucional	Dos aldeas	Evaluación a través del estudio de caso. Análisis costo beneficio	Los beneficios del programa son mayores si se toma en cuenta el enfoque de las competencias.
Nutrición y apoyo social	Food Aid and Child Nutrition in Rural Ethiopia	Quisumbing AR. (2003) Etiopía Institucional	1,500 familias distribuidas en 15 aldeas	Se utilizó información del Estudio de las Familias Rurales en Etiopía. Se utilizó el panel de datos. Aplicación de cuestionarios. Análisis de tendencia, puntuaciones Z y análisis de regresión	Se encontró que ambos programas tuvieron un efecto positivo en el peso y talla de los beneficiarios-
Pobreza y apoyo social	Impact Assessment of CFPR/TUP: A Descriptive Analysis Based on 2002-2005 Panel Data	Rabbani M, Prakash VA, Sulaiman M. (2006) Bangladesh Institucional	5,626 familias. Se determinó grupo de tratamiento y grupo control.	Se estableció un panel de las familias Se utilizó el enfoque del sustento sostenible Diferencia en la diferencia	Se encontró un impacto definitivo del programa sobre la base de recursos de la familia participante del programa.
Salud y educación	The Impact of Life Skills Education on Adolescent Sexual Risk Behaviors	Magnani, R. et al. (2003) Sudáfrica Institucional	2,222 jóvenes de 14 a 22 años	No se menciona	Se encontró un efecto substancial del programa a corto y mediano plazo respecto a áreas tales como conocimiento, confianza en el uso del condón y uso del condón.
Pobreza y empleo	Poverty and Employability Effects of Workare Programs in Argentina	Ronconi L, Sanguinetti J, Fachelli S, Casazza V. & Franceschelli I. (2006)	3,930 personas Se determinó grupo de tratamiento (n =655) y grupo control (n =3.275).	Se utilizó la información brindada por el Instituto Nacional de Estadística y Censos acerca de los Estudios de Hogares Permanentes.	Se encontró que para una gran fracción de los beneficiarios el programa generó dependencia y no incrementó su capital humano.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Argentina Institucional		Propensity Score Matching, Diferencia en las diferencias y análisis de regresión	
Pobreza y apoyo social	Does subsidized childcare help poor working Women in urban areas? Evaluation of a Government-sponsored program in Guatemala city	Ruel MT, De la Brière B, Hallman K, Quisumbing A. & Coj N. (2002) Guatemala Institucional	500 niños y 1,400 familias Se determinó grupo de tratamiento y grupo control	Diseño de control de caso del grupo de niños Selección aleatoria de la muestra de familias. Se utilizó métodos de análisis cuantitativos y cualitativos. Entrevistas estructuradas Observación en hogares Grupos de discusión focal	Se encontró que el impacto nutricional del programa es positivo y significativo
Pobreza y apoyo social	Evaluation of Echo's 1999 to 2002 Funded Actions in Sudan	Schuftan C, Van der Veen A. & Lothe P. (2003) Sudan Institucional	25 proyectos de desarrollo social financiados por la European Comisión Humanitarian Office (ECHO)	Revisión exhaustiva de documentos y entrevistas personal de ECHO y de las organizaciones socias.	Los proyectos Echo han logrado desarrollar a las poblaciones vulnerables en diferentes aspectos. Dicho impacto es difícil de cuantificar debido a la ausencia de una línea base.
Empleo y derechos humanos	The Impact of Social Labelling on Child Labour in India's Carpet Industry	Sharma AN, Sharma R. & Raj N. (2000) India Institucional	47 exportadores distribuidos en 35 aldeas	La selección de la muestra fue a través de técnicas aleatorias e intencionales. Se aplicó un cuestionario Grupos de discusión focal	Se encontró un impacto limitado de la intervención a través del etiquetamiento social.
Apoyo social y empleo	Conditional Cash Transfers, Adult Work Incentives, and Poverty	Skoufias E. & Di Maro V. (2006) México	24,000 familias distribuidas en 506 localidades. Se determinó grupo de tratamiento y grupo control	Diseño de evaluación experimental pre post. Se utilizó información de la Encuesta de Características Socioeconómicas de los Hogares).	Se encontró que el programa no tiene efectos significativos en la participación de la fuerza laboral adulta y el tiempo libre.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional		Estimador de diferencias en las diferencias. Análisis de regresión.	
Salud	An experimental evaluation of recovery management checkups (RMC) for people with chronic substance use disorders	Dennis M, Scott CK, Funk R. (2003) Estados Unidos Journal	448 personas adultas.	Diseño de investigación experimental. La selección de los grupos de tratamiento y de control se realizó de manera aleatoria. Los datos se adquirieron a través de diversas fuentes: entrevista estandarizada, prueba de orina y prueba de saliva.	Se demostró la importancia del protocolo de intervención Recovery Managment Checkups como medio de favorecer resultados a largo plazo en personas con desordenes crónicos en el consumo de sustancias
Pobreza	Evaluating of the Impact of Conditional Cash Transfers on Schooling: An Experimental Analysis of Hondura's PRAF Program	Glewwe, Paul; Olinto, Pedro (2004) Honduras Institucional	Se seleccionaron 70 municipalidades de manera aleatoria para la formación de cuatro grupos (uno de ellos el grupo control). Se obtuvo un total de 5,748 familias	La información se obtuvo a través de la aplicación de cuestionarios escolares. Se realizó análisis de simulación a través del modelo del proceso de transición escolar de Harkov. Diferencia de medias Diferencia en la diferencia Selección aleatoria Se realizaron estimaciones econométricas	Se encontró un efecto estadísticamente positivo del impacto de la demanda de intervención pero no en la entrega de la intervención.
Salud	The 'healthy brothel': the context of clinical services for sex workers in Hillbrow	Stadler J. & Delaney S. (2004) Sudáfrica Institucional	Trabajadoras sexuales	Diseño de evaluación cualitativo Grupos de discusión focal y entrevistas a profundidad a trabajadoras sexuales.	Se encontró que la campaña (Clínica móvil) tuvo un efecto positivo en la búsqueda de un comportamiento saludable de las trabajadoras sexuales.
Salud	Increased protected sex	Stanton BF, et al.	515 jóvenes.	Diseño de evaluación	Se encontró que el programa

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	and abstinence among Namibian youth following a HIV risk-reduction intervention: a randomized, longitudinal study	(1998) Namibia (África) Journal	Se determinó grupo de tratamiento y grupo control	experimental pre post. La asignación a los grupos de tratamiento y control fue aleatoria. Aplicación de cuestionarios. Chi Cuadrado.	“My future is My Choice” redujo el comportamiento de riesgo de VIH entre los participantes sexualmente inexpertos.
Salud	Impact of the Heart Campaign: Findings from the youth surveys, 1999-2000	Underwood C, Hachonda H, Serlemitsos E. & Bharath U. (2001) Zambia Institucional	901 jóvenes (368 hombres y 533 mujeres)	Diseño de evaluación cuasi-experimental pre post. Cuestionarios de línea base y post test. Análisis estadísticos bivariados y multivariados.	Se encontró un efecto positivo del programa.
Pobreza y apoyo social	Impact Assessment of Poverty Alleviation Programmes from Human Rights Perspective	Bandyopadhyay KR. & Mukherjee AN. (2005) India Institucional	Distrito de Bolangir, Orissa.	Estudio de Caso. Se utilizó un diseño de evaluación desde el enfoque de los Derechos Humanos, se estableció los siguientes cuatro principios: no-discriminación, participación y empoderamiento, transparencia y evaluación, y realización progresiva de los derechos humanos.	No se especifica
Calidad de vida	An Assessment of Community-Based Health Financing Activities in Uganda	Derriennic Y, Wolf K. & Kiwanuka-Mukiibi P. (2005) Uganda Institucional	No se especifica	Estudio de evaluación cualitativo. Entrevista semi-estructuradas a informantes claves y grupos de discusión focal.	Se encontró que el proyecto generó un desarrollo significativo en la calidad de vida de sus beneficiarios.
Educación, salud, nutrición y población	Evaluating Preschool Programs when Length of Exposure to the Program Varies: A Nonparametric	Behrman, Jere; Cheng, Yingmei; Todd, Petra (2003)	Tres grupos de comparación: niños de la comunidad que fueron presentados el programa pero no participan, niños de	Estudio de diseño cuasi-experimental. Diferencia en la diferencia	El programa tiene un impacto acumulado positivo y significativo sobre las puntuaciones de las pruebas

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Approach	Bolivia Institucional	comunidades similares en las que el programa no ha sido presentado y niños que están en el programa por menos de un mes.	Variables instrumentales Otros métodos de Emparejamiento.	para los niños de 37 a 64 semanas. El impacto marginal es positivo y significativo para los niños mayores de 42 meses.
Educación, protección social, suministro de agua y sanidad	Evaluating Social Funds: A Cross-Country Analysis of Community Investments	Rawlings, Laura; Sherburne-Benz, Lynne; Van Domelen, Julie (2004) Armenia, Bolivia, Honduras, Nicaragua, Perú y Zambia, (multinacional) Banco Mundial (Libro)	Familias beneficiarias de los programas (grupo de tratamiento) y familias no beneficiarias (grupo control)	Análisis de impacto multinacional. Se utilizó el método de Diferencia de medias, Propensity Score Matching. Se realizaron los correspondientes análisis de regresión	El programa tuvo un efecto positivo y significativo en la asistencia escolar, un descenso significativo de las familias que informan una incidencia de enfermedad y la probabilidad de informar desocupación debido a enfermedad fue menor.
Información (infraestructura)	Evaluating the impact of an electronic business system in a complex organizational setting: the case of Central Contractor Registration	Morel, JA. (2002) Estados Unidos Journal	Caso único	Caso único	Queda demostrado que la evaluación de impacto del programa puede ser obtenido, y que puede ser de un costo razonable
Educación	Evaluating the Impact of Conditional Cash Transfer Programs: Evidence from Latin America	Rawlings, Laura B., and Gloria M. Rubio. (2003) Brasil, México, Honduras, Jamaica, y Nicaragua (multinacional) Institucional	Niños entre las edades de 7 a 14 años.	Diseño cuasi experimental Uso de la Encuestas de Familias. Diferencia de medias y otros métodos de emparejamiento.	Existe una clara evidencia del éxito del programa al incremental la tasa de matrículas, favorecer el cuidado de la salud preventiva e aumentar el consumo familiar.
Educación, transporte, suministro de agua y	Evaluating the Impact of Infrastructure Rehabilitation Projects on	Lokshin, Michael and Ruslan Yemtsov (2003)	La selección de la muestra no fue aleatoria, incluyendo 249 aldeas de población rural	Los análisis se basan en dos fuentes de datos: a) El estudio oficial de familias en Georgia	El número de aldeas que informan un incremento en el número de estudiantes de

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
sanidad	Household Welfare in Rural Georgia	Georgia Institucional		(SGHH, 1996) y del Estudio de Infraestructura Comunitaria Rural (2002). Diferencia en la diferencia y Propensity Score Matching	secundaria se incrementó significativamente en comparación con las comunidades en control.
Educación	Evaluating the Impact of PETI on Child Labor Supply and Schooling Demand in Rural Northeastern Brazil: The Case of Pernambuco, Bahia and Sergipe	Sedlacek G, Yap, Y. & Orazem P. (2000) Brasil Institucional	Municipalidades con familias beneficiarias (grupo de tratamiento) y Municipalidades con familias no beneficiadas (grupo de tratamiento.)	Se utilizó el estudio de hogares. Diferencia de medias y otros métodos de emparejamiento	No se reportó hallazgos con niveles significativos.
Educación: gestión del sector público	Evaluating the Impact of School Decentralization on Educational Quality	Galiani, Sebastian; Schargrodsy, Ernesto (2002) Argentina Institucional	Alumnos del quinto de secundaria de Escuelas públicas (grupo de tratamiento) y escuelas privadas (grupo control)	Se utilizó la información brindada por a partir de las evaluaciones que realizó Sistema de Evaluación de la Calidad Educativa de los niños del quinto de secundaria (1994-1998). Se utilizó en método de evaluación de diferencia en la diferencia.	Se encontró que el desempeño en las escuelas públicas se promueve significativamente con la descentralización.
Salud	Evaluating the impact of training in psychosocial interventions: a stakeholder approach to evaluation – part II	Forrest S, Masters H. & Milne V. (2004) Reino Unido Journal	16 participantes	Diseño cualitativo Se utilizó la entrevista semi-estructurada. El análisis de los datos fue conducido sobre una base temática y comparativa y el programa de análisis cualitativo NUDIST.	El estudio mostró la importancia del proceso de reconstrucción de la Intervención Psicológica en los participantes para la práctica cotidiana y la importancia de los resultados de capacitar en actitudes y enfoques. Existe poca evidencia de que los encuestados perciban el uso extensivo de la Intervención Psicológica como el objetivo final de los cursos de capacitación.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Suministro de agua y sanidad	Evaluating the Impact on Child Mortality of a Water Supply Project and Sewerage Expansion in Quito: Is Water Enough?	Galdo, Virgilio; Briceno, Bertha (2004) Ecuador Institucional	Las unidades de análisis son las mujeres de 15 a 49 años de edad	Se utilizó información secundaria de los censos 1990 y 2001, realizado por el Instituto Nacional de Estadística y Censos. También se utilizó la información obtenida a través de la Encuestas de Condiciones de Vida (1998). Diferencia de medias Diferencia en la diferencias Propensity score matching.	Se encontró un impacto negativo significativo en la mortalidad infantil
Salud	An impact evaluation of student teacher training in HIV/AIDS education in Zimbabwe	Chifunyisea T, Benoyb H, & Mukiibi B. (2002) Zimbabwe Journal	1572 estudiantes de educación participaron en el llenado de los cuestionarios. 207 estudiantes de educación participaron en los grupos de discusión focal	Se utilizó cuestionarios de línea base y seguimiento, y guías para grupos de discusión focal.	Hubo un incremento en el conocimiento de la prevención del Virus de Inmunodeficiencia Humana (VIH) y en la habilidad de los profesores para temas de salud reproductiva y sexual. Sin embargo, como los estudiantes estuvieron expuestos a otro material de VIH fuera del programa, no todos los cambios se deben al programa.
Salud	Evaluation of a Model Self-Help Telephone Program: Impact on Natural Networks	Goodman, C. (1990) Canadá Journal	81 participantes fueron seleccionados aleatoriamente para conformar una red telefónica o a una serie de lecturas profesionales accesibles a través del teléfono, y luego fueron reasignados al componente alternativo luego de tres meses.	Estudio experimental pre post en su modalidad de diseño en serie de tiempos. 40 cuidadores de pacientes con Alzheimer fueron mantenidos en el estudio para la recolección la información. Instrumentos: a) The Memory and Problem Behavior Checklist, b) The Brden Interview, c) The Mental Health Index, d) Percieved Social Support for Caregiving and	Seis meses de aplicado el programa dejó a los participantes con una acrecentada información y apoyó la satisfacción y desarrollo el compromiso de las familias y amigos como apoyos emocionales.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				Social Conflict. Los datos fueron sometidos al ANOVA	
Salud	Evaluation of a Peer Assessment Approach for Enhancing the Organizational Capacity of State Injury Prevention Programs	Hunter WH, Schmidt ER, & Zakocs R. (2005) Holanda Journal	Siete directores de programas de evaluación de impacto	Entrevista telefónica semi estructurada. Revisión documentaria. Criterios del Programa State Technical Assesment Team (STAT)	La evaluación de impacto reveló que luego de un año, los siete programas estatales para la prevención de daños habían actuado sobre 81% de las recomendaciones recibidas durante las visitas de STAT. Todos los programas reportaron ganancias en visibilidad y credibilidad dentro del departamento de salud estatal con otras unidades y agencias.
Pobreza	Evaluation of Rhode Island's Family Independence Program (FIP)	Dryden A, Queral M. & Tauchen H. (2001) Reino Unido Institucional	29253 familias lideradas por padres o madres soleteros.	Se utilizó información administrativa de diversas fuentes, incluyendo información del estado del registro de apoyo en efectivo, y de los informes de empleo e ingresos del programa de Seguro de Desempleo. Se siguió un diseño de estudio longitudinal, utilizándose asimismo los estimadores de efectos combinados y de efectos aleatorios.	Se encontró que el mayor impacto del FIP fue el incremento de la probabilidad de que el apoyo corriente y en efectivo a los beneficiarios sí funcionaría. Sin embargo, el ingreso estimado en la ganancia trimestral debido al FIP fue relativamente modesto.
Desarrollo urbano	Evaluation of Sites and Services Projects: The Evidence from El Salvador	Bamberger, Michael; Gonzalez-Polio, Edgardo; Sae-Hau, Umnuy (1982) El Salvador Institucional	El grupo de tratamiento estuvo integrado por 196 familias	Estudio de diseño cuasi experimental. Para compensar la falta de equivalencia del grupo control se utilizó el análisis de regresión que controló las diferencias iniciales.	El proyecto tuvo un impacto positivo y significativo sobre la ganancia por trabajador.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Salud	Evaluation of the impact of the Baby-Friendly Hospital Initiative on Rates of Breastfeeding	Braun M.; Giugliani E, Matos M, Giugliani C, Proenco A, & Machado A, (2003) Brasil Journal	Primer cohorte: 187 bebés nacidos en 1994 Segundo cohorte: 250 bebés nacidos en 1999.	Estudio observacional pre-post de dos cohortes. Entrevista personal y a través del teléfono acerca de la dieta usual de sus niños. La información fue sometida a análisis de regresión coxial y la prueba Chi cuadrada para la diferencia de grupos.	El estudio muestra un significativo incremento de la tasa de lactancia materna, especialmente la de tipo exclusiva, luego de la implementación de la Baby-Friendly Hospital Initiative (BHFI).
Salud	Evaluation of the Pharmaceutical Industry Investment Program	Productivity Commission (2003) Australia Institucional	Participantes del programa de los años 1998-99 a 2001-02	Diseño de evaluación cuasi experimental pre post. Análisis del error de predicción.	La comparación del nivel de producción entre quienes participaron y no participaron en el programa evidencia un efecto positivo del programa. Sin embargo, el análisis de predicción, la comparación entre la actividad producida antes y luego de la aplicación del programa, y la comparación entre empleo, inversión y exportaciones, todos estos revelan debilidades o ausencias en el impacto del programa.
Empleo	Evaluation the impact of SENA on earnings and employment	Gaviria A. & Núñez JA. (2003) Colombia Institucional	10000 jefes de hogar localizados alrededor de 75 municipalidades, tanto del área rural como urbana.	Diseño de estudio post tratamiento, sin comparación. Se utilizaron dos fuentes de datos: la Encuestas Nacional de Calidad de Vida (1997) y el estudio conducido por el Servicio Nacional de Aprendizaje (SENA). Se escogió un grupo control para cada participante a través del promedio de aproximación por	Los resultados muestran de manera consistente que el impacto de SENA para los servicios de capacitación es negativo: ganancias de participación menores a 0.13 veces del sueldo mínimo.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				semejanzas.	
Educación, Salud, Nutrición y población, suministro de agua y sanidad	Ex-post Evaluation of the Honduran Social Investment Fund (FHIS 2)	Walker, Ian; Del Cid, Rafael; Ordonez, Fidel; Rodriguez, Florencia (1999) Honduras Institucional	El grupo de tratamiento están conformado por los beneficiarios del programa y el grupo control	La información es de fuente secundaria, proveniente del Bi-annual Household Survey de la Oficina de Estadística y Censos de Honduras y de los estudios de hogares. Diferencia en la diferencia y comparación "Pipeline". Se utilizó en análisis multivariado para controlar las diferencias entre grupo de tratamiento y control.	El programa tuvo un impacto positivo y significativo en la proporción de niños entre 8 y 9 años que cursan el grado escolar de acuerdo a su edad, existió mayor probabilidad en la búsqueda de atención médica en casos de enfermedad y un menor y significativo índice de diarrea.
Salud, Nutrición y Población	The Impact of PROGRESA on Health	Gertler, Paul. (2000) México Institucional	Una muestra 506 comunidades, distribuidas aleatoriamente para conformar los grupos de tratamiento (320) y control (186)	Diseño experimental. Estimador Diferencia en la Diferencia para evaluar los cambios producidos por el programa.	El programa incrementó significativamente la utilización de las clínicas de salud pública para el cuidado de prevención, asimismo, existen significativos avances en la salud de los niños y adultos.
Salud	Findings from a retrospective process evaluation of five HIA studies	Jackson N, Taylor L, & Quigley R. (citados por Quigley et al, (2003)) Inglaterra Journal	Estudios previos	Diseño exploratorio Entrevistas cualitativas detalladas fueron tomadas a informantes clave en los estudios de evaluación de impacto en salud (Health Impact Assessment = HIA)	Los principales hallazgos muestran preocupación por la necesidad del seguimiento y evaluación de las actividades relacionadas. Sin embargo, ninguno de los cinco estudios de HIA, y de los entrevistados, había estado incluido o había sido sujeto a una evaluación externa o independiente.
Educación	Assessing the Impact and Effectiveness of the Advanced Technological Education (ATE) Program.	Hanssen C., Gullickson A. & Lawrenz, F. (2003)	128 Directores o responsables de los proyectos.	Se utilizó una encuesta estructurada. Se midieron los indicadores en	El programa está impactando en gran número de estudiantes a través de los esfuerzos de promoción del programa.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Estados Unidos Institucional (Libro)		dos dimensiones: estado actual del programa y tendencia del programa. No se utilizó grupo control.	El programa ATE está estableciendo exitosas relaciones de colaboración con otro tipo de instituciones u otros programas ATE. Los proyectos de ATE están utilizando la evaluación para guiar las actividades del programa, pero existe todavía un significativo vacío en la información acerca de la calidad de los productos del proyecto.
Suministro de Agua y sanidad	Has Private Participation in Water and Sewerage Improved Coverage? Evidence from Latin America	George RG.; Kosec, K.; Wallsten, SJ. (2004) Argentina (multinacional) Institucional	Familias de Argentina, Bolivia y Brasil residentes en 18 ciudades donde se privatizó el sistema de agua y alcantarillado (grupo de tratamiento) y 28 ciudades donde no hubo privatización (grupo control)	Se emparejó las regiones en función de las poblaciones. Estimador de Diferencia en la diferencia	El programa Participación del Sector Privado incrementa la proporción de habitantes con conexiones de agua por tubería y de alcantarillado. Sin embargo, este incremento deja de ser significativo si se le hace una comparación en el tiempo y con el grupo control.
Salud, Nutrición y Población	Health Behavior and the Design of Public Health Programs: Evidence from Randomized Evaluations	Miguel, Edward; Kremer, Michael (2003) Kenya Journal	75 escuelas primarias divididas aleatoriamente en tres grupos. Se determinó grupos de tratamiento y control según etapa de la evaluación	Diseño experimental pre post. Se utilizó el estimador Diferencia en la diferencia	Las tasas de tratamiento fueron reducidas significativamente a un 62%. La intervención de compromiso verbal redujo significativamente el consumo de fármacos por casi seis puntos porcentuales.
Pobreza	Hidden Impact? Ex-Post Evaluation of an Anti-Poverty Program	Chen S. & Ravallion M. (2003) China	2000 hogares	Evaluación pre post en serie de tiempos a través del método de emparejamiento por similitudes para el análisis del grupo control	Se encontró un retorno promedio del desembolso del programa de 9 a 10%, superior al impacto de la

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional		y grupo que participó en el programa y análisis de la diferencia entre las diferencias.	asistencia del gobierno para áreas pobres. Sin embargo, se encontró que la mitad del ingreso obtenido es ahorrado, de modo que el impacto del programa es mucho menos evidente en el consumo de los participantes.
Pobreza	Hidden Impact? Household Saving in Response to a Poor-Areas Development Project.	Ravallion, Martin and Shaohua Chen. (2005) China Institucional	35 aldeas pobres de tres provincias de China: Guangxi, Guizhou y Yunan. El grupo control lo integraron aquellas aldeas que no están en el programa pero se encuentran dentro de la provincia donde interviene el programa.	La línea base se estableció a partir de los estudios realizados por el Rural Household Survey Team (RHS de 1996 -2000) en 2000 familias distribuidas en 200 aldeas. Método de Diferencia en las Diferencias para evaluar el impacto del programa sobre pobreza e ingreso. Propensity Score Matching para disminuir la heterogeneidad entre las aldeas en tratamiento y las de control.	Se encontró un incremento significativo en el ingreso y ahorro.
Pobreza	Household Effects of African Community Initiatives: Evaluating the Impact of the Zambia Social Fund	Chase, Robert S., and Lynne Sherburne-Benz (2001) Zambia Institucional	La muestra es de 20950 familias que pertenecen a 99 comunidades donde al programa ha intervenido.	Se utilizó la información de hogares del Estudio de Monitoreo sobre Condiciones de Vida. Se utilizó dos diseños de evaluación: a) Comparación Pipeline y b) Propensity Score Matching. Diferencia de Medias	Se encontró evidencia del alcance del programa hacia las familias pobres, particularmente en las áreas rurales. En comparación con el grupo control, el programa incrementó de manera significativa la asistencia escolar y el gasto familiar en educación. Asimismo, su impacto fue mayor en la zona rural que la urbana.
Pobreza	Impact Evaluation of a Conditional Cash Transfer	Maluccio, John A.; Flores, Rafael (2004)	Se seleccionaron a 42 comarcas, de las cuáles se	La selección de comarcas control y tratamiento fueron a través de	Al comparar varias dimensiones de capital

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Program: The Nicaraguan Red de Protección Social	Nicaragua Institucional	obtuvo 21 para el grupo de tratamiento y 21 para el grupo control. Se realizó una selección aleatoria estratificada de 1585 familias.	un índice de marginalidad basado en el National Population and Housing Census. Se tomaron en cuenta para esta evaluación tres estudios de familia: La línea Base La evaluación utilizó un diseño aleatorio y midió el impacto utilizando el estimador Diferencia en la Diferencia.	humano, incluso de estatus nutricional, salud y educación, estos programas han sido capaces de influir en muchos de los indicadores destacados dentro de las estrategias de reducción de la pobreza.
Educación, salud, nutrición y población, suministro de agua y sanidad	Impact Evaluation of Social funds. An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund	Newman, John, Meno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa, and Jose Luis Evia. (2002) Bolivia Journal	Muestra aleatoria de 200 escuelas fueron, 86 para el grupo de intervención y 124 para el grupo control.	Diferencia en la diferencia, Propensity Score Matching, Otros Métodos de Emparejamiento.	El programa Social Investment Fund (SIF) tuvo un impacto positivo y significativo sobre parte de las escuelas con facilidades de sanidad, número de textos por alumno, la cifra de salones en buenas condiciones. Sin embargo, el programa tuvo un efecto significativo negativo en la tasa de deserción escolar y el número de alumnos matriculados por colegio.
Educación, Salud, Nutrición y Población, Suministro de Agua y Sanidad	Impact Evaluation of Social Funds. The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund	Pradhan, Menno, and Laura B. Rawlings. (2002) Nicaragua Journal	Se seleccionó una muestra no aleatoria. Se determinó grupo de tratamiento y grupo control.	Los datos analizados provienen del estudio del Fondo de Inversión de Emergencia Social de Nicaragua (1998) y Estimador de Diferencia en la Diferencia. Propensity Score Matching Otros métodos de emparejamiento.	El programa tuvo un impacto significativo en la matrícula escolar primaria. Por otro lado, el impacto de la intervención en salud no es conclusivo y el programa tuvo un significativo impacto en la proporción de familias con baño.
Educación	Impact Evaluations of	Chase, Robert S.	Se obtuvo una muestra	Diferencia de medias	El programa Armenia Social

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Social Funds. Supporting Communities in Transition: The Impact of the Armenian Social Investment Fund	(2002) Armenia Journal	estratificada de las familias según zona de residencia: sísmica, en conflicto y no especificada. Se determinó grupo de familias en tratamiento y grupo de familias aspirantes (control).	Se utilizó el método Propensity Score matching para disminuir sesgos de selección. La información proviene de un estudio de familias integrado (1998-1999) que incluye información de ingreso, gasto, educación y salud familiar. Se realizaron los correspondientes análisis de regresión.	Investment Fund tuvo un efecto positivo y significativo en el gasto familiar de la escuela, así como un efecto positivo y significativo en la matrícula escolar.
Educación	Impact of a vocational counselor on employment-related outcomes among methadone patients	Appela PW, Smith R, Schmeidler JB, & Randell J (2000) Estados Unidos Journal	619 pacientes de dos clínicas. Clínica donde se mantuvo la metadona (Clínica 1; N = 364) y la clínica con la que se realizó la comparación (Clínica 2; N = 358)	Diseño cuasiexperimental. La información adicional acerca de los servicios de educativo-vocacionales brindados a los pacientes fueron obtenidos de los reportes ingresados al Servicios de Alcoholismo y Abuso de Sustancias por el Consultor de Rehabilitación Vocacional. Se utilizaron análisis de regresión logística.	Se concluyó que el establecimiento de un consejero de rehabilitación vocacional en la clínica que mantuvo la metadona es una intervención educativa vocacional de costo efectivo. Los servicios educativos-vocacionales se acrecentaron significativamente en la Clínica 1; la presencia educativa-vocacional se incrementó de 53 a 56% en la clínica 1 y disminuyó en la clínica 2, 45-43%.
Educación	Assessing the Impact and Effectiveness of the Advanced Technological Education (ATE) Program.	Coryn CL, Gullickson AR, & Hanssen CE (2004) Estados Unidos. Institucional	Se evaluó a 16 centros del programa	El impacto se determinó a través de la comparación entre el número de estudiantes que completan el programa y los estudiantes que no lo concluyen.	Los centros del programa ATE impactan preactiva y positivamente en los estudiantes y la fuerza tecnológica de los Estados Unidos. El número de alumnos que completan el programa exceden a los que lo abandonan.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
Justicia	Impact of Legislation Requiring Moped and Motorbike Riders to Wear Helmets	Ballart X. & Riba C. (1995) España Journal	La muestra comprende el número de accidentes en pequeñas motocicletas y casco (grupo de tratamiento) y en motocicletas grandes (grupo control).	Se realizaron tres observaciones de series de tiempo con grupo control no equivalente. El grupo control no equivalente permite medir los efectos de la medida de legislativa respecto al uso obligatorio del casco a los motociclistas y el número de daños serios y el de muertes. Se utilizó la técnica de Análisis de Series ARIMA con el fin de estimar cambios significativos en la tendencia del número del número de víctimas con daños graves en el grupo de tratamiento.	El impacto de la adopción de la medida legislativa grande y permanente pero el hábito de uso de casco de los motociclistas no ha sido influido por lo campaña a través de los medios ni por las posibles sanciones.
Pobreza	Will welfare reform influence marriage and fertility? Early evidence from the ABC demonstration	Fein, D. (2001) Estados Unidos Journal	1547 mujeres para el análisis de matrimonio y 1027 para el análisis de la fecundidad.	Evaluación de diseño experimental. Se asignó aleatoriamente a los beneficiarios y nuevos postulantes de cinco oficinas piloto, tanto hacia el grupo de tratamiento como al grupo control. Se realizó una evaluación aleatoria de las participantes en el programa.	Se encontró evidencia de que los cambios políticos en un estado sí tienen efectos en el matrimonio y el cuidado de los niños. El programa A Better Chance (ABC) tuvo impactos positivos en la cohabitación marital en mujeres menores de 25 años y aquellas con menos de 12 años de instrucción.
Salud	Impact of Providing a Medical Home to the Uninsured Evaluation of a Statewide Program	Gill JM, Bittner H, Townsend B, & Mainous AG. (2005) Estados Unidos Journal	Se eligieron a 795 personas que participan del programa.	Se estableció un diseño de cohorte retrospectivo donde se compara el cuidado de la salud de una línea base a seis meses luego de haberse integrado al programa. Se utilizó la entrevista personal o por teléfono. Se desarrollaron análisis de	Se encontró una disminución en la proporción respecto a las visitas del departamento de emergencias pero no de hospitalización, así como un incremento significativo en la satisfacción con el cuidado. El programa CHAP de Delaware está asociado con

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				tendencia y comparación de resultados a través de la prueba de McNemar's	un importante desarrollo a través de muchas mediciones de la utilización del cuidado de la salud y representa un método exitoso y fácilmente financiable para que los estados fomenten el cuidado de la salud de sus poblaciones no aseguradas..
Educación	Improving Primary School Education in India: An Impact Assessment of DPEP-Phase I	Jalan, Joytsna; Glinskaya, Elena (2004) India Institucional	La selección de los beneficiarios al programa no fue aleatoria, constituyéndose una muestra total integrada por las familias de 40 distritos. Asimismo, el impacto fue evaluado a través de dos submuestras: grupo de tratamiento y grupo control.	Evaluación del programa District Primary Education Program (DPEP, fase I.) Diferencia en la Diferencia Propensity Score Matching para el control del sesgo de selección.	El programa tuvo un impacto positivo y significativo sobre la asistencia a la escuela y sobre el mayor nivel de educación lograda.
Educación	Incentives to Learn	Kremer M, Edward M, Thornton R. & Ozier O. (2004) Kenya Institucional	Se seleccionó una muestra aleatoria de 127 escuelas primarias, 63 fueron invitados a participar en el programa y 64 sirvieron como grupo control.	La información sobre las puntuaciones en los exámenes fueron obtenidas del District Education Offices (DEO) en los Distritos de Busia y Teso. Se utilizó el Diseño de Regresión Discontinua para evaluar el impacto del programa.	El programa tuvo un impacto positivo y significativo sobre las puntuaciones del examen, la asistencia escolar y la asistencia de los profesores a clases.
Salud y nutrición	Increased Weight Gain with Mass Deworming Given During Child Health Days In Uganda	Alderman, Harold; Sebuliba, Isaac; Konde-Lule, Joseph; Hall, Andrew (2003) Uganda Institucional	Se realizó un muestreo aleatorio de grupos con tres estratos, obteniéndose 50 parroquias. Se determinó grupo de tratamiento y grupo control.	Se utilizó el método de evaluación de Diferencia en la Diferencia. Un cuestionario de línea base fue aplicado a 2250 familias por Instituto de Salud Pública. Se aplicó el análisis de regresión multivariado.	La provisión periódica de fármacos antihelmínticos tuvo un impacto significativo sobre el índice de ganancia de peso en los niños menores de cerca de 10% sobre lo esperado si el tratamiento se realiza dos veces al año.
Educación	Investing In Children Through The Food For	Ahmed, Akhter, Carlo Del Ninno, and Omar	930 niños en edad escolar de 6 a 13 años de edad.	Variabes instrumentales y otros métodos de emparejamiento.	El programa tuvo un efecto positivo significativo en la

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Education Program	Haider Chowdhury. (2004) Bangladesh Institucional		Con grupo control Diferencia de medias y estadísticas descriptivas.	matrícula escolar. Las mujeres tienen una mayor probabilidad de matricularse que los varones.
Pobreza	Jamaica's Food Stamp Program: Impacts on Poverty and Welfare	Ezemenari, Kene; Subbarao, K (1998) Jamaica Institucional	El panel de estudio estuvo integrado por 986 familias. Se determinó grupo de tratamiento y grupo control.	Los datos considerados provienen del Jamaican Survey of Living Conditions (JSLC, 1989-1991) Diferencia en la diferencias y simulación contrafactual Análisis de dominancia estocástica. Marginal propensity to consume.	El programa redujo significativamente las brechas de pobreza. Sin embargo no sucede lo mismo con la incidencia de la pobreza.
Educación	Limiting Child Labor Through Behavior-Based Income Transfers: An Experimental Evaluation of the PETI Program in Rural Brazil	Yap, Yoon-Tien, Guilherme Sedlacek, and Peter Orazem. (2002) Brasil Institucional	Seis municipalidades divididas en grupo de tratamiento y grupo control. Se seleccionó de manera aleatoria de cada municipalidad a 200 familias con al menos un niño. Lo cual hace un total de 1200 familias.	Diferencia de medias y otros métodos de emparejamiento	Se encontró que el programa incrementa de manera significativa el tiempo de permanencia en la escuela en los tres Estados. Asimismo, reduce la probabilidad de trabajo infantil.
Salud, Nutrición y Población	Longitudinal Evaluation of Uganda Nutrition and Early Child Development Program	Alderman, Harold; Britto, Bia; Siddiqi, Arjumand (2004) Uganda Institucional	Se realizó un muestreo aleatorio de grupos con tres estratos, obteniéndose una muestra de 50 parroquias. Se determinó grupo de tratamiento y grupo control.	Se realizó una evaluación longitudinal de análisis profundo. Se utilizó el estimador Diferencia en la Diferencia. Se hicieron análisis de regresión y de comparación a través de las puntuaciones Z.	Las comunidades beneficiarias tienen un significativo descenso en la probabilidad de abandono temprano de la lactancia con pecho. El programa tuvo un efecto positivo y significativo en el consumo de alimentos nutritivos. Asimismo, el programa

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
					<p>incrementó el consumo de medicinas antiparasitarias.</p> <p>Por último, existe un efecto favorable del programa en la educación y relaciones sociales de los niños.</p>
Transporte de Energía Eléctrica y otras Energías.	Assessing the Impact of Transport and Energy Infrastructure on Poverty Reduction	<p>Cook, Cynthia C., Tyrrell Duncan, Somchai Jitsuchon, Anil K. Sharma, Wu Guobao. (2004)</p> <p>India</p> <p>Institucional</p>	Se estratificaron cuatro grupos de estudio en términos de la distancias de la carretera (a menos de 0.5 km está el grupo de tratamiento y a más de 0.5 km, el grupo control) y de acuerdo al acceso de los hogares a la electricidad (Aquellos con acceso fueron el grupo de tratamiento, y aquellos sin acceso fueron el grupo control)	<p>Se utilizó información secundaria de la Nacional Simple Survey (1987/88, 1993/94 y 1999/2000).</p> <p>También se aplicaron cuestionarios y entrevistas.</p> <p>Diferencia en la Diferencia y Propensity Score Matching.</p> <p>Uso del modelo probit para analizar la intervención individual sobre la pobreza.</p> <p>Se utilizó la prueba T para evaluar la significancia de las medias de consumo entre los subgrupos maestres.</p>	El acceso a la electricidad tuvo un negativo y significativo impacto sobre la pobreza en las provincias de Panchmahal y Kuchchh.
Salud, Nutrición y Población	Medium and Long Run Effects of Nutrition and Child Care: Evaluation of a Community Nursery Programme in Rural Colombia	<p>Attanasio, Orazio P., and A. Marcos Vera-Hernandez. (2004)</p> <p>Colombia</p> <p>Institucional</p>	No informa.	<p>Se usó de información de tipo secundaria proveniente del programa Familias en Acción (2002).</p> <p>Se aplicaron cuestionarios detallados para obtener información de la Línea Base.</p> <p>Diferencia en la diferencia Variables instrumentales (distancia de la residencia de la familia al programa y la distancia</p>	La participación en el programa Hogares Comunitarios de Bienestar Familiar tiene un efecto positivo significativo sobre la altura de los niños, sobre la probabilidad de estar en la escuela, así como sobre la probabilidad de que sus madres estén empleadas y con más horas laborales.

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				de la residencia de todas las familias del pueblo hacia el programa) Otros métodos de emparejamiento. Análisis de puntuaciones Z, desviación estándar,	
Desarrollo	Microcrédito, Evaluación de impacto. Caso: Brasil y Chile	Aroca, P. (2002) Brasil – Chile Institucional	796 observaciones (81= grupo tratamiento y 715= grupo control)	Se definieron dos grupos: las personas que recibieron microcrédito y las que no (grupo control). Se utilizó los puntajes de predisposición y las técnicas de emparejamiento.	Los resultados para los datos brasileños muestran un alto impacto positivo de los programas de microcrédito, especialmente para aquellos asociados con los bancos. En el caso Chileno, la evidencia es más débil para el microcrédito administrado por los bancos. En lo referente a los programas de las ONG, la evidencia sugiere que su impacto sobre el ingreso promedio de sus clientes es realmente negativo.
Salud	Multi-media campaign exposure effects on knowledge and use of condoms for STI and HIV/AIDS prevention in Uganda	Bessinger R, Katende C, & Gutpa N (2004) Uganda Journal	Este estudio comprendió una muestra de 1697 mujeres y 900 hombres.	Se utilizó la información del estudio de evaluación Delivery of Improving Service for Health de los años 1997 y 1999. Para evaluar la asociación entre la exposición a la campaña se utilizó el análisis de regresión logística.	La exposición a la campaña Behavior Change Communication estuvo más fuertemente asociada con el mayor conocimiento del condón. Las campañas que utilizan múltiples canales de comunicación de masa serían más efectivas en la promoción del conocimiento de la salud sexual.
Suministro de Agua y Sanidad con efectos en educación	Nicaragua Ex-Post Impact Evaluation of the Emergency Social	Banco Mundial (2000) Nicaragua	Se obtuvo una muestra constituida por 1312 familias que conformaron el grupo	La evaluación se realizó a partir de tres fuentes de datos principales: Living Standards	Los resultados de la evaluación del impacto familiar demuestran un

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
	Investment Fund (FISE)	Institucional	beneficiado y no beneficiado con el programa.	Measurement Survey (1998), La FISE Facilities Survey, y The Qualitative Beneficiary Assesment. La comparación de los beneficiarios y no beneficiarios de FISE se realizó a través de la comparación contrafactual. Estimador de Diferencia en la Diferencia. Propensity Score Matching Otros métodos de emparejamiento.	significativo y robusto impacto de las inversiones en educación primaria del programa sobre los resultados e indican un significativo impacto de la inversión en agua del programa sobre los resultados en salud.
Educación	Nicaragua's School Autonomy Reform: Fact or Fiction?	King, Elizabeth M.; Ozler, Berk; Rawlings, Laura B. (1999) Nicaragua Institucional	Se evaluó cerca de 200 escuelas del nivel primario y secundario.	Se utilizó información de estudios de escuelas y hogares (1995 y 1997), en el mismo grupo de escuelas. Un Cuestionario especial fue diseñado para conocer acerca de la toma de decisiones en la escuela. Se utilizó un modelo econométrico para comparar las escuelas beneficiarias del programa de las no beneficiarias.	La participación en el programa tuvo un impacto positivo y significativo sobre la toma de decisiones para la escuela primaria y secundaria. El programa tuvo un efecto positivo y significativo en la influencia percibida de los directores tanto en los asuntos pedagógicos como administrativos.
Educación	Organizational learning: A cross-national pilot-test of effectiveness in children's services	Ortner DK, Cook P, Sabah Y, & Rosenfeld J. (2005) Israel - Estados Unidos (multinacional) Journal	131 niños americanos y 132 niños de Israel participaron en la línea base. Mientras que 153 niños de norteamericanos y 153 niños de Israel participaron al final del período de estudio.	Se utiliza un diseño cuasi-experimental y longitudinal. El aprendizaje organización fue medido con la Organizational Learning Assesment Scale (OLAS); mientras que la percepción de la eficacia	Los resultados indican que los mejoramientos en el aprendizaje organizacional ocurrieron en el caso de recibir el tratamiento y no en los otros. Hubo también una relación significativa entre las puntuaciones en las escalas de

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
				<p>organizacional de los trabajadores fue medida con la Worker Empowerment Scale (WES), y la estimación del comportamiento individual que tiene el profesor del alumnos se evaluó mediante la Behavioral Rating Index for Children.</p> <p>Se utilizaron análisis de correlación no-paramétrica y paramétrica.</p>	<p>aprendizaje organizacional y las mediciones de satisfacción laboral y emponderamiento entre el personal. Los resultados fueron confirmados sólo para el caso de los programas de Estados Unidos.</p>
Educación	Philippines: Student Outcomes in Philippine Elementary Schools: An Evaluation of Four Experiments	Tan, Jee-Pang, Julia Lane, and Gerard Lassibille (1999) Filipinas Journal	Se seleccionaron 40 escuelas ubicadas en áreas con bajos ingresos. De estas, 20 fueron integraron el grupo de tratamiento y 20 el grupo control.	<p>Se utilizó un diseño de evaluación aleatorio.</p> <p>Se utilizó el estimador de Diferencia en la Diferencia.</p> <p>El control de diferencias entre los grupos de tratamiento y control se realizó a través del análisis multivariado.</p>	<p>Las tasas de deserción disminuyeron significativamente en las escuelas que recibieron material de aprendizaje de diferente grado.</p> <p>La alimentación escolar con la relación padre-profesor tuvieron un impacto positivo y significativo sobre las puntuaciones alcanzadas en matemática.</p>
Pobreza y desarrollo	Primed by the Euro: The Impact of a Referendum Campaign on Public Opinion and Evaluations of Government and Political Leaders	De Vreese, CH.(2004) Holanda Journal	962 encuestados 23 noticias obtenidas de manera aleatoria.	Análisis de contenido de las noticias en los medios y un estudio panel.	El inconveniente del gobierno para manejar el referéndum pobremente fue penalizado por el referéndum. La exposición a los titulares de medios de comunicación que cubrían extensivamente el referéndum y ofrecieron evaluaciones negativas de los líderes políticos estimularon la disminución del nivel de aprobación de los líderes políticos por ciudadanos

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
					políticamente menos involucrados.
Salud	Program evaluation with unobserved heterogeneity and selective implementation: the Mexican PROGRESA impact on child nutrition	Behrman JR. & Hoddinott J (2001) México Institucional	Grupo de niños que recibe el tratamiento (n=336) y grupo de niños que no recibe el tratamiento (n=325)	Diseño comparativo transversal en su primera fase. Comparación de Medias entre grupo de tratamiento y grupo control.	Las estimaciones indican que PROGRESA no tuvo efectos en el estado nutricional del niño preescolar. Sin embargo, a través del análisis de regresión se encontró que el programa está relacionado significativamente con el crecimiento en estatura y disminuye las probabilidades de atrofio en la niñez.
Educación, Salud, Nutrición y Población.	Progesa and its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI	Skoufias, Emmanuel (2001) México Institucional	Se obtuvo una muestra aleatoria de 24,000 familias de 506 localidades, éstas fueron distribuidas para formar el grupo de tratamiento y el grupo control.	Se utilizó el diseño aleatorio controlado del programa y la disponibilidad de información de panel antes y después de la intervención. Se empleó el estimador Diferencia en la Diferencias para medir el impacto del programa.	PROGRESA incrementó de manera significativa la matrícula escolar, asimismo redujo la tasa de deserción escolar y la probabilidad de trabajar en los escolares.
Empleo	Assisting the Transition from Workfare to Work: A Randomized Experiment	Galasso, Emanuela; Ravallion, Martin; Salvia, Agustin (2001) Argentina Institucional	Un total de 953 familias participaron en la línea base del estudio. Dos grupos de tratamiento y un grupo control	La evaluación utilizó un diseño de medición experimental. Se aplicó el Permanent Household Survey. Método de evaluación diferencia en la diferencia, variables instrumentales y grupos aleatorios-	Existe un efecto significativo para los participantes del grupo de tratamiento respecto a la posibilidad de ser empleados en el sector privado.
Educación	Progressing Through Progesa: An Impact Assessment of a School Subsidy Experiment	Behrman, Jere; Sengupta, Piyali; Todd, Petra (2001) México	506 aldeas rurales fueron asignadas aleatoriamente para participar en el programa o servir como grupo control.	La información fue obtenida de la línea base de los estudios de hogares (1997 y 1998). El impacto del subsidio del progesa se midió a través del	Se halló que el programa reduce efectivamente la deserción escolar y favorece el paso hacia el siguiente nivel de estudios, particularmente de la primaria ha la

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Institucional		Modelo de Transición Escolar de Markov aplicado a la información experimental. Simulación del impacto del programa.	secundaria. Si los niños participarían en el programa a edades de 6 a 14 años, experimentarían un incremento de 0.6 años en el promedio de logro educacional y un incremento de 19% en el porcentaje de jóvenes que asisten a la escuela secundaria
Suministro de agua y sanidad.	Reforming Urban Water Supply: The Case of Chile	Shirley, Mary M., Colin Xu L., and Ana Maria Zuluaga (2000) Chile Institucional	No se menciona	Diferencia de medias y simulación contrafactual. Para la elaborar la comparación contrafactual de los años 1989 y 1998, se proyectaron los parámetros clave en base a sus tendencias lineales desde 1981 a 1988.	No se informó sobre ganancias significativas respecto a la privatización del agua.
Educación	Remedying Education: Evidence from Two Randomized Experiments in India	Banerjee, Abhijit, Shawn Cole, Esther Duflo & Leigh Linden. (2004) India Institucional	Se realizó una selección aleatoria estratificada de la muestra según lengua de instrucción, género y proporcionalidad alumnos-profesores. El estudio evaluó 62 escuelas en total. Se determinó grupo de tratamiento y grupo control.	Diseño de evaluación experimental. Se analizó el efecto de los programas Balsakhi y Mumbai. Diferencia de medias, diferencia en la diferencias, variables instrumentales.	Se encontró que los programas Balsakhi y Mumbai tienen efectos positivos y significativos tanto en las puntuaciones de las pruebas de matemáticas como de lenguaje.
Educación	Retrospective Vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya	Glewwe, Paul; Kremer, Michael; Moulin, Sylvie; Zitzewitz, Eric (2000) Kenya Institucional	Muestra integrada por 89 escuelas que recibieron de manera aleatoria los rota folios (grupo de tratamiento) y 89 escuelas que no los recibieron (Grupo control).	Evaluación experimental de tipo retrospectiva y prospectiva. Diferencia en la diferencia	Las estimaciones prospectivas acerca del uso de rota folios en el desempeño académico de los niños en Kenya no muestran impacto en el aprendizaje. El análisis sugiere que las

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
					regresiones más retrospectivas tendrían mayor sobrestimación de los efectos de un programa que brinda rotafolios a gran escala.
Educación	School Meals, Educational Achievement, and School Competition: Evidence from a Randomized Evaluation	Vermeersch, Christel; Kremer, Michael (2004) Kenya Institucional	50 escuelas fueron divididas aleatoriamente en dos grupos de 25 cada una: grupo de tratamiento y grupo de comparación.	Se utilizó un diseño de selección aleatorio. Estimación de impacto utilizando un <i>intention-to-treat-estimator</i> . Diferencia en la diferencia	El programa tuvo un efecto positivo y significativo sobre el promedio de participación escolar y en su desempeño en los exámenes de evaluación; asimismo se encontró influencia positiva significativa en el peso de los estudiantes varones
Educación	Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment	Duflo, Esther (2001) Indonesia Institucional	Se analizó a dos grupos: la muestra total integradas por 152,989 varones con edad promedio de 9.98 años de educación completa, y el grupo de quienes reciben salario: que suman un total de 60.633. Se determinó grupo de tratamiento y grupo control.	Se utilizó la información provinientes del Intercensal Survey of Indonesia (SUPAS, 1995). Variables instrumentales, Two Stage Least Squares Análisis de Regresión	El programa incrementó significativamente los años de educación de los niños de 2 a 6 años de edad en 0.12 años para la muestra total y en 0.20 años para la muestra de quienes perciben salario. El salario se incrementó significativamente en 1.5% como resultado del programa.
Mercado laboral y Empleo	Social Protection in a Crisis: Argentina's Plan Jefes y Jefas	Galasso, Emanuela and Martin Ravallion. (2003) Argentina Institucional	Se dividió la muestra entre los que se benefician del programa (grupo de tratamiento) y los que solicitaron pero no se benefician del tratamiento (grupo control)	Se utilizó la información de la Encuesta Permanente de hogares (2001 y 2002) Diferencia de medias. Diferencia en diferencia. Propensity Score Matching	El programa tiene un impacto positivo y significativo en el ingreso familiar e individual.
Salud	Social Support and Psychological Functioning Among High-Risk Mother: The impact of the Baby Love Maternal Outreach Worker Program	Navaie-Waliser M, Martin SL, Tessaro I, Campbell MC, & Cross AW. (2000) Estados Unidos	419 madres embarazadas, 221 pertenecen al grupo de tratamiento y 198 al grupo control.	Se utilizaron entrevistas estructuradas cara a cara. Escala de autoestima de Ronsenberg (1981)	El estudio sugiere que la intensidad del apoyo es un componente importante de los programas de visita a domicilio que se centran en desarrollar las funciones

Área de estudio	Título de la evaluación	Autores / lugar de aplicación / año	Muestra	Metodología	Conclusiones
		Journal		Diferencia de medias	psicológicas de las madres.
Educación	South Africa: A Review of Two Evaluations of the Application of the READ Primary Schools Program in the Eastern Cape Province of South Africa.	Schollar, Eric (2001) Sudáfrica Journal	Se utilizó una muestra de tratamiento de 850 escolares y una muestra control de 360 escolares. La selección de la muestra fue a través del método de conteo por intervalo simple.	Diseño pretest postest. Se utilizó el método de evaluación Diferencia en la Diferencia.	El programa tuvo un impacto positivo y significativo en la lectura y escritura.
Educación	Systemic Education Evaluation Evaluating the Impact of Systemwide Reform in Education	Jonathan A. Supovitz (2005) Estados Unidos Journal	20,000 estudiantes de las 14 provincias más grandes del estado de Florida. 126 directores de escuela., 96 de escuelas elementales y 25 de nivel secundario.	Estudio de tipo no experimental. La medición del impacto del programa Duval se basó en el logro obtenido por los estudiantes., el promedio de puntuación alcanzado en lectura y matemáticas así como la comparación entre las escuelas fue a través de la observación de cinco años: 1999, 2000, 2001, 2002 y 2003. Se aplicaron análisis descriptivos, modelos de regresión de mínimos cuadrados. Visitas estructuras a centros educativos para la observación de temas particulares.	Las estimaciones de impacto podrían no reflejar el verdadero impacto del programa Duval. Aunque utilizar un modelo de efectos de corrección reduce el monto de desviación de las características estables de la escuela, el monto que permanece luego aplicar los controles basados en los modelos es desconocido

Tal como se observa en la Tabla N°3.3, se han realizado evaluaciones de impacto en distintos ámbitos, siendo necesario, por tanto, resumir la información de las mismas utilizando la vía estadística. Los resultados se presentan en el siguiente subtítulo.

3.2.2.1. Revisión sistemática de las tendencias de evaluación de impacto

En esta sección se presenta la revisión sistemática de las tendencias de evaluación de impacto de los programas sociales en el mundo. De una muestra no aleatoria de 437 informes de evaluación, se realiza un análisis estadístico de sus principales características, tendencias metodológicas, áreas de aplicación e impacto. En una primera sección se presenta la descripción de los estudios, en general, mientras que en una segunda se realiza una comparación entre las Evaluaciones de Impacto (EI) realizadas en el Perú, frente a las realizadas en el resto del mundo.

3.2.2.1.1. Evaluación de impacto: Perspectiva global

Según el continente o región donde tuvo lugar el estudio de evaluación de impacto (n=433), el mayor porcentaje de ellos se realizó en Latinoamérica – incluyendo México- (37.4%), África (18.7%), Asia (16.1%), América del Norte (13.2%) y Europa (8.8%). Al respecto, los tres primeros continentes están comprendidos –principalmente- por países pobres o en vías de desarrollo; mientras que los dos últimos están integrados por países desarrollados. Por otro lado, 4.2% de los estudios de impacto se han realizado en diversos continentes y 0.2% no especifican el lugar de evaluación.

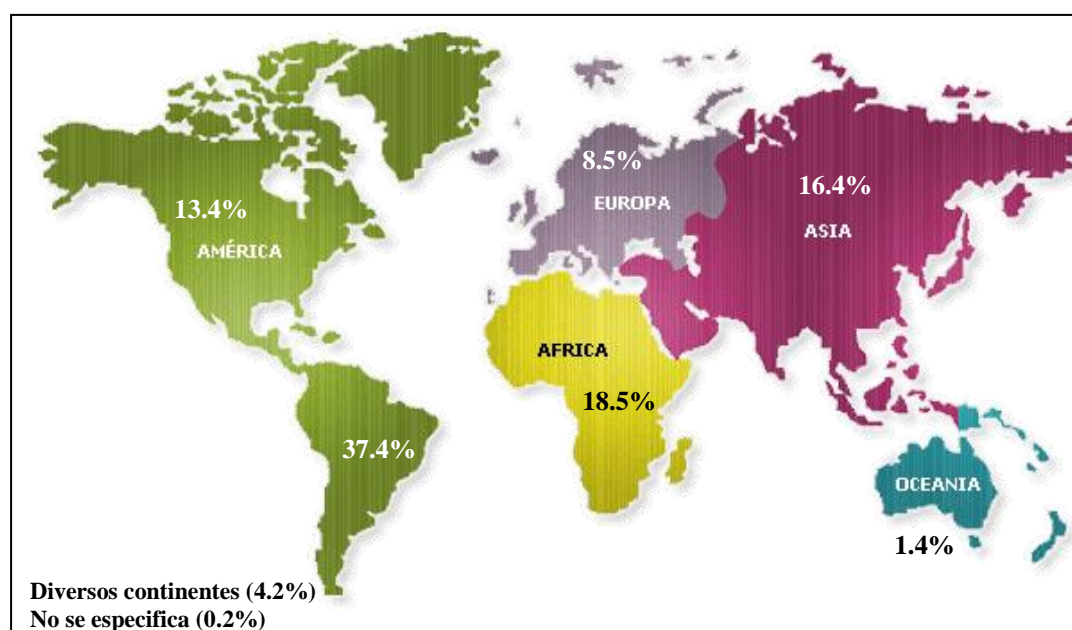


Figura N° 3.1. Distribución porcentual de la evaluación de impacto de los programas sociales según continente (Fuente: Elaboración propia).

Respecto a los países en donde se realizaron las evaluaciones de impacto, se destacan México y Perú, para el caso de Latinoamérica; Kenya para el caso del continente africano; Bangladesh e India para Asia; Estados Unidos para América del Norte; Reino Unido para Europa y Australia para Oceanía. Estos valores se presentan en la Tabla 3.4.

Tabla N°. 3.4. Países donde se realizaron las evaluaciones de impacto (Valores expresados en frecuencias) (Fuente: Elaboración propia).

<i>Latino América</i>		<i>África</i>		<i>Asia</i>	
México (33)	Chile (3)	Kenya (16)	Burkina Faso (2)	Bangladesh (15)	Armenia (2)
Perú (33)	Costa Rica (3)	Sudáfrica (7)	Camerún (2)	India (12)	Georgia (2)
Argentina (16)	El Salvador (3)	Uganda (6)	Liberia (2)	Indonesia (10)	Korea (2)
Colombia (15)	Jamaica (2)	Ghana (5)	Mozambique (2)	Filipinas (8)	Nepal (2)
Brasil (14)	República Dominicana (2)	Marruecos (5)	Egipto (1)	China (4)	Singapur (1)
Honduras (7)	Haití (1)	Malawi (4)	Eritrea (1)	Israel (4)	Tailandia (1)
Nicaragua (6)	Panamá (1)	Tanzania (4)	Magadascar (1)	Pakistán (3)	Turquía (1)
Bolivia (5)	Paraguay (1)	Túnez (4)	Namibia (1)	Vietnam (3)	
Ecuador (4)	Multinacional (13)	Zambia (4)	Sierra Leona (1)	Multinacional (1)	
Guatemala (4)		Etiopía (3)	Sudán (1)		
		Guinea (3)	Turquía (1)		
		Zimbabwe (3)	Zaire (1)		
<i>Europa</i>		<i>América del norte</i>		<i>Oceanía</i>	
Reino Unido (19)	Dinamarca (1)	Estados Unidos (53)		Australia (6)	
Suecia (4)	Italia (1)	Canadá (4)			
Holanda (3)	Polonia (1)	Multinacional (1)			
España (2)	Rusia (1)				
Alemania (2)	Multinacional (2)				
Bulgaria (1)					

En cuanto a la fuente de publicación de los artículos o informes de evaluación de impacto, solamente el 23.6% proviene de revistas científicas internacionales; la gran mayoría (75.8%) proviene de informes institucionales (*Working Papper*) de organismos tales como el Banco Mundial, Banco interamericano de Desarrollo, Institutos de Investigación de Universidades, entre otros.

Tabla N° 3.5. Fuente de publicación de los artículos o informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Institucional	328	75.8
Journal	103	23.8
No se especifica	2	0.4
Total	433	100.0

Según se muestra en la Figura 3.2, los estudios de evaluación de impacto de programas se han incrementado significativamente hacia fines de las década de los

noventa y principios de la presente década. Por tanto, la gran mayoría (77.3%) de estos estudios han sido publicados en los últimos seis años; 15.2% durante la década de los noventa y sólo 7.4% desde 1976 hasta 1989. Este dato demuestra que la tendencia creciente de la evaluación de impacto es geométrica.

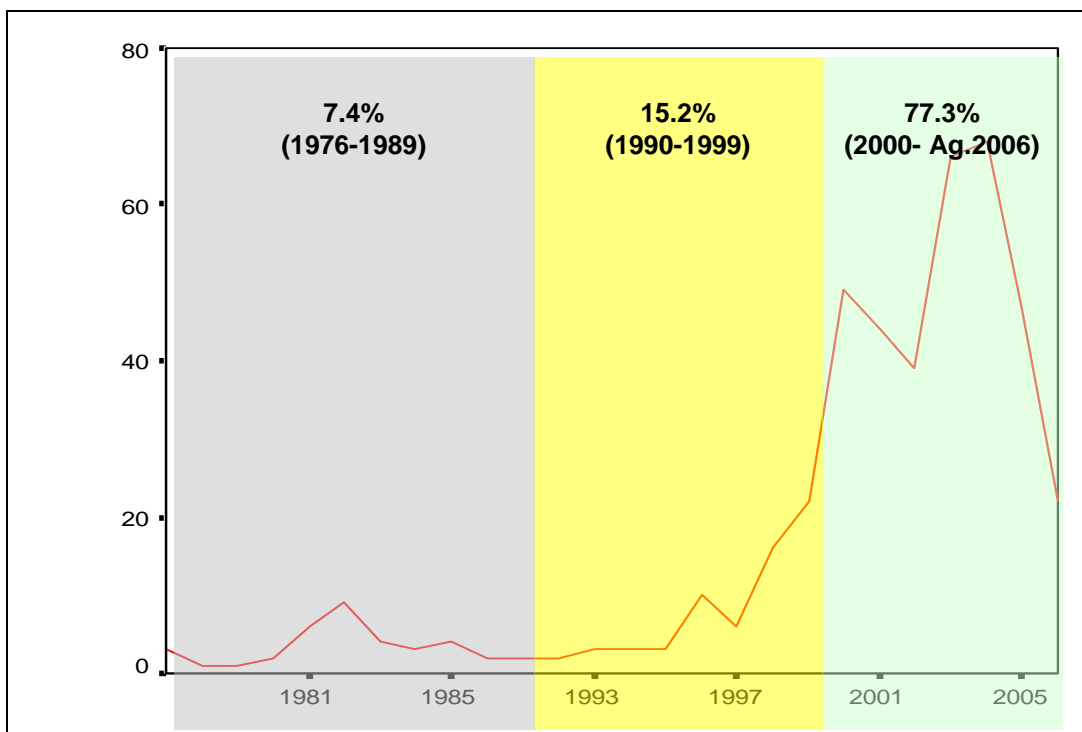


Figura N° 3.2. Año de publicación de los artículos o informes de evaluación de impacto (Fuente: Elaboración propia).

Tal como se aprecia en la Tabla 3.6, las principales áreas de estudio e interés de las evaluaciones de impacto son los programas educativos (26.3%), de salud (21.5%), de pobreza extrema (14.6%) y de empleo (11.4%). En menor porcentaje se han centrado en la evaluación de programas alimentarios, infraestructura y electrificación, agricultura y ganadería, entre otros.

Tabla N° 3.6. Área donde se realizó la evaluación de impacto (Fuente: Elaboración propia).

Áreas	Frecuencia	Porcentaje
Educación	113	26.1
Salud	94	21.7
Pobreza extrema	64	14.8
Empleo	49	11.3
Alimentación	19	4.4
Infraestructura y electrificación	18	4.2
Agricultura y ganadería	18	4.2
Agua y sanidad	16	3.7
Justicia	14	3.2
Micro crédito e ingresos familiar	13	3.0
Calidad de vida, convivencia social y medio ambiente	8	1.8
Desarrollo urbano (títulos propiedad)	4	.9
No se especifica	3	.7
Total	433	100.0

Por otro lado, según se observa en la Tabla 3.7, que entre las poblaciones beneficiadas por los programas sociales se destacan las familias (48.1%) y los estudiantes de educación básica (11.4%). La elección de estas poblaciones beneficiarias guarda correspondencia con las áreas de estudios preponderantes en las evaluaciones de impacto señaladas anteriormente: educación, salud y pobreza extrema.

Tabla N° 3.7. Población beneficiaria de los programas considerados en los artículos o informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Familias	210	48.1
Estudiantes de educación básica	50	11.4
Niños e infantes	41	9.4
Adultos y ancianos	38	8.7
Jóvenes	28	6.4
Mujeres	26	5.9
Profesionales o expertos	13	3.0
Países	10	2.3
Pacientes o personas discapacitadas	8	1.8
Estudiantes de educación superior	5	1.1
No se especifica	8	1.8
Total	437	100.0

Según se observa en la Tabla 3.8, los datos requeridos para el proceso de análisis cuantitativo o cualitativo del impacto de los programas, fueron obtenidos principalmente de personas (51.9%) o familias (18.9%). Otros tipos de unidades de análisis considerados lo constituyen las delimitaciones territoriales o gubernamentales (distritos, comunidades, países), organizaciones o instituciones, programas o proyectos sociales; incluyéndose, también, estudios previos sobre evaluación de impacto de programas sociales, bajo un enfoque revisionista.

Tabla N° 3.8. Unidad de análisis considerada por los artículos o informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Personas	227	51.9
Familias	79	18.1
Distritos, comunidades, localidades o países	36	8.2
Centros, instituciones u organizaciones	36	8.2
Programas o proyectos sociales	28	6.4
Estudios previos	12	2.7
No se especificó	19	4.3
Total	437	100.0

De acuerdo con la naturaleza de la recolección de datos, se encontró que 49.9% de evaluaciones han usado datos primarios, mientras que 37.5% han usado datos secundarios. Se demuestra así la importancia de los datos secundarios en el desarrollo de estos estudios. Estos valores se presentan en la Tabla 3.9.

Tabla N° 3.9. Origen de los datos utilizados para el análisis de la evaluación de impacto (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Primarios	218	49.9
Secundarios	164	37.5
No se especificó	55	12.6
Total	437	100.0

Según se muestra en la Tabla 3.10, los medios o instrumentos de recolección de datos empleados para el proceso de EI comprenden, principalmente, los registros de datos secundarios (35.9%) y los cuestionarios (32%). Asimismo, un porcentaje importante hizo uso de la entrevista (13.7%). Estos dos últimos instrumentos se han utilizado exclusivamente para la recolección de datos primarios.

Tabla N° 3.10. Instrumento utilizado para la recolección de datos en los artículos o informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Registros de datos secundarios	157	35.9
Cuestionario	140	32.0
Entrevista	60	13.7
Fichas de observación	10	2.3
No se precisó	70	16.0
Total	437	100.0

Por otro lado, el mayor porcentaje de artículos o informes de evaluación de impacto reportan el empleo de la metodología cuasi-experimental (41%). También es significativo el porcentaje de artículos que siguen el método pre-experimental (19%) y de análisis cualitativo (14.2%). Aquellos informes de evaluación que han utilizado la metodología experimental propiamente dicha solamente constituyen sólo el 9.4%. Se observa también la aparición de informes de metodología meta-analítica (1.6%), siendo estas muy recientes. Se verifica, entonces, la primacía del diseño cuasi-experimental en la EI.

Tabla N° 3.11. Metodología de investigación empleada en los artículos o informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Cuasi-experimental	179	41.0
Pre-experimental	83	19.0
Análisis cualitativo	62	14.2
Experimental	41	9.4
Meta-análisis	7	1.6
No se especifica	65	14.9
Total	437	100.0

En la Tabla 3.12 se muestra que los informes de evaluación de impacto que han usado grupo control superan en gran medida a aquellos que no lo han hecho (n=225 vs. n=84). Al respecto, como era de esperarse, son más frecuentes los estudios con grupo control que han utilizado la metodología experimental (n=40) o cuasi-experimental (n=172), mientras que es menor si se trata de artículos que siguen la metodología pre-experimental, de análisis cualitativo o meta-análisis.

Por otro lado, los estudios experimentales refieren, en primer lugar, el uso de muestreo probabilístico (n=36), mientras que los estudios cuasi-experimentales reportan con mayor frecuencia el muestreo tanto probabilístico (n=48) como no probabilístico (n=54). Los estudios pre-experimentales, de análisis cualitativo y meta-análisis fueron los que, en mayor número, no utilizaron muestreo probabilístico. Estos valores se presentan en la Tabla 3.13.

Tabla N° 3.12. Metodología de investigación empleada según el uso de grupo control en los artículos o informes de evaluación de impacto (Fuente: Elaboración propia).

Metodología	Uso de grupo control		Total (Fr)
	No (Fr)	Sí (Fr)	
Experimental	1	40	41
Cuasi-experimental	6	172	178
Pre-experimental	54	7	61
Análisis cualitativo	20	5	25
Meta-análisis	3	1	4
<i>Total</i>	84	225	309

Tabla N° 3.13. Frecuencia de artículos o informes de evaluación de impacto según la metodología de investigación y el tipo de muestreo empleado

Metodología	Tipo de muestreo empleado				Total
	Probabilístico	No probabilístico	Sin muestreo	Censo	
Experimental	36	2	2		40
Cuasi-experimental	48	54	26	3	131
Pre-experimental	11	10	15	1	37
Análisis cualitativo	7	2	15		24
Meta-análisis		1	2		3
<i>Total</i>	102	69	60	4	235

Según el tamaño de las muestras empleadas, un alto porcentaje de informes de EI no informan al respecto (23.6%), no obstante, se observa que es mayor la tendencia de utilizar muestras superiores a los 1,000 sujetos (33.4%). Estos valores se presentan en la Tabla 3.14.

Tabla N° 3.14. Tamaño de muestras reportadas en los informes de evaluación de impacto (Fuente: Elaboración propia).

Intervalos	Frecuencia	Porcentaje
Menos de 49	46	10.5
50 – 99	28	6.4
100 – 299	51	11.7
300 – 499	22	5.0
500 – 999	41	9.4
1000 a 2999	45	10.3
3000 a más	101	23.1
No se especifica	103	23.6
Total	437	100.0

Entre los diseños de evaluación se destacan, en primer lugar, el uso de la diferencia en las diferencias (24.3%), seguido por los diseños pretest- posttest (15.3%) y de diferencias sólo posttest (15.1%). Llama la atención que uno de los diseños cuasi-experimentales más robustos (discontinuidad de la regresión) sea uno de los menos usados (0.9%).

Tabla N° 3.15. Diseños de evaluación reportados en los informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Diferencias en diferencias	106	24.3
Pretest - posttest	67	15.3
Diferencias sólo posttest	66	15.1
Series de tiempo interrumpidas	38	8.7
Costo beneficio	18	4.1
Estudio de caso	16	3.7
Descripción de medias y porcentajes	12	2.7
Discontinuidad de la regresión	4	.9
No se especifica	110	25.2
Total	437	100.0

Respecto al uso de la técnica de variables instrumentales, se encontró que dicha técnica fue señalada solamente en 7.6% de informes de EI véase la Tabla 3.16. A pesar que es una técnica poderosa para controlar sesgos por efectos de variables endógenas, su uso se ve restringido.

Tabla N° 3.16. Uso de variables instrumentales en los informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Uso</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
No	392	89.7
Sí	33	7.6
No se especifica	12	2.7
Total	437	100.0

De igual manera, es reducido el porcentaje de artículos o informes de evaluación de impacto donde se señaló la utilización de técnicas de emparejamiento de grupos muestrales. Así, sólo el 19.7% de los artículos informan del uso de esta técnica para controlar sesgos de selección.

Tabla N° 3.17. Uso de emparejamiento de grupos en los informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Uso</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
No	341	78.0
Sí	86	19.7
Total	427	97.7
No se especifica	10	2.3
Total	437	100.0

No obstante existe un alto porcentaje de informes de evaluación que no hacen referencia a los análisis estadísticos aplicados (45.3%), la mayoría de artículos que sí informan, han utilizado ecuaciones de regresión (38%), estadísticas de comparación de medias (8.2%) y las estadísticas de tipo descriptivas (5.5%). Definitivamente, las ecuaciones de modelamiento lineal (regresión) son las más populares en la EI, debido quizá por su flexibilidad y simplicidad.

Tabla N° 3.18. Análisis estadísticos empleados en los informes de evaluación de impacto (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Ecuaciones de regresión	166	38.0

Estadísticas de comparación de medias	36	8.2
Estadísticas descriptivas (correlación, porcentajes, etc.)	24	5.5
Análisis categórico cualitativo	6	1.4
ARIMA´s	4	.9
Ecuaciones estructurales	3	.7
No se especificó	198	45.3
Total	437	100.0

En cuanto a los resultados de la evaluación, un porcentaje significativo de informes de EI (79.2%) concluyó que los programas sociales evaluados han generado impacto en la calidad de vida de sus beneficiarios. No obstante, existe un importante porcentaje (16.5%) que concluyó la ineficacia del programa, en la medida que no produjo impacto significativo. En términos de proporción, por lo menos 2 de cada 10 programas sociales evaluados no han producido efecto positivo alguno en los beneficiarios.

Tabla N° 3.19. Impacto alcanzado por el programa o proyecto social evaluado (Fuente: Elaboración propia).

<i>Impacto</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Tuvo impacto	346	79.2
No tuvo impacto	72	16.5
No se especifica	19	4.3
Total	437	100.0

En síntesis, la principal fuente de difusión y producción de EI en el mundo son los Organismos Internacionales, quienes han aportado más del 75% de la producción de informes. Esta es la razón por la cual los informes provienen más de países en desarrollo y se centran en aspectos educativos, de salud, pobreza y empleo, por cuanto son el objetivo e interés de estos organismos. Adicionalmente, las EI han aumentado geométricamente durante los últimos 30 años, concentrándose más del 77% en el periodo 2000-2006. En términos metodológicos, un importante porcentaje de EI (37.5%) utiliza datos secundarios, y en el caso de los primarios, el instrumento más usado es el cuestionario. El diseño más usado es el cuasi-experimental (41%), principalmente el de diferencias en diferencias. El tamaño de la muestra es muy variable (33.4% supera las mil unidades) y se centra principalmente en personas y familias. En cuanto al uso de técnicas de control especiales, su uso es muy limitado para el caso de las variables instrumentales (7.6%) y limitado para el *matching* (19.7%). La técnica estadística más empleada es la regresión (38%). Finalmente, los resultados indican que por lo menos 2 de cada 10 programas evaluados no muestran impacto positivo alguno.

3.2.2.1.2. Evaluación de impacto: Perú vs el mundo

A continuación se presentan los diferentes **análisis comparativos** entre los estudios de EI realizados en nuestro país y los correspondientes al resto del mundo. De este modo, se observa que los primeros están representados por el 7.6% del total de informes sobre evaluación de impacto; mientras que los segundos por el 91.5%. Asimismo, se tienen tres informes que no refieren el país donde se ejecutó

la evaluación del respectivo programa social; por lo tanto, el análisis comparativo se limita a 433 informes. Estos valores se presentan en la Tabla 3.20.

Tabla N° 3.20. País donde fueron realizados los programas sociales contenidos en los informes de evaluación de impacto (Fuente: Elaboración propia).

<i>País</i>	<i>Frecuencia</i>	<i>Porcentaje</i>
Perú	29	7.6
Otros países	404	91.5
No específica	3	0.9
Total	436	100.0

Según se muestra en la Tabla 3.21, todos los artículos o informes sobre evaluación del impacto aplicados en nuestro país han sido publicados o archivados en fuentes institucionales⁵⁰. En el caso de los países extranjeros, se encontró un significativo porcentaje de informes que han sido publicados en revistas de investigación (25.3%). Esto hecho es un indicador del menor apoyo tanto académico como económico a la realización de estudios de evaluación de impacto en nuestro medio nacional.

Tabla N° 3.21. Fuente de publicación de los artículos o informes de evaluación de impacto en el Perú y otros países (Fuente: Elaboración propia).

<i>Fuente</i>	<i>Otros países (n = 404)</i>	<i>Perú (n= 29)</i>
Institucional	74.3	100.0
Revistas científicas	25.3	-
No se especifica	0.4	-
Total	100.0	100.0

Por otro lado, en la Figura N° 3.3 se muestra que la tendencia histórica de publicación de artículos o informes de evaluación de impacto es muy similar entre el Perú y el resto de países. No obstante, llama la atención la ausencia de estudios de evaluación de impacto en nuestro país durante el intervalo de tiempo comprendido entre fines de la década de los ochenta y mediados de los años noventa. Situación coincidente con los graves problemas de índole político y económico que atravesó el país, marcados por el terrorismo y la dictadura.

⁵⁰ Ello no excluye que algunos autores nacionales han publicado sus informes vía artículo de investigación en *Working Pappers* de instituciones extranjeras. Ejm: Chong A. & Galdo J. (2006) y Valdivia M. (2004).

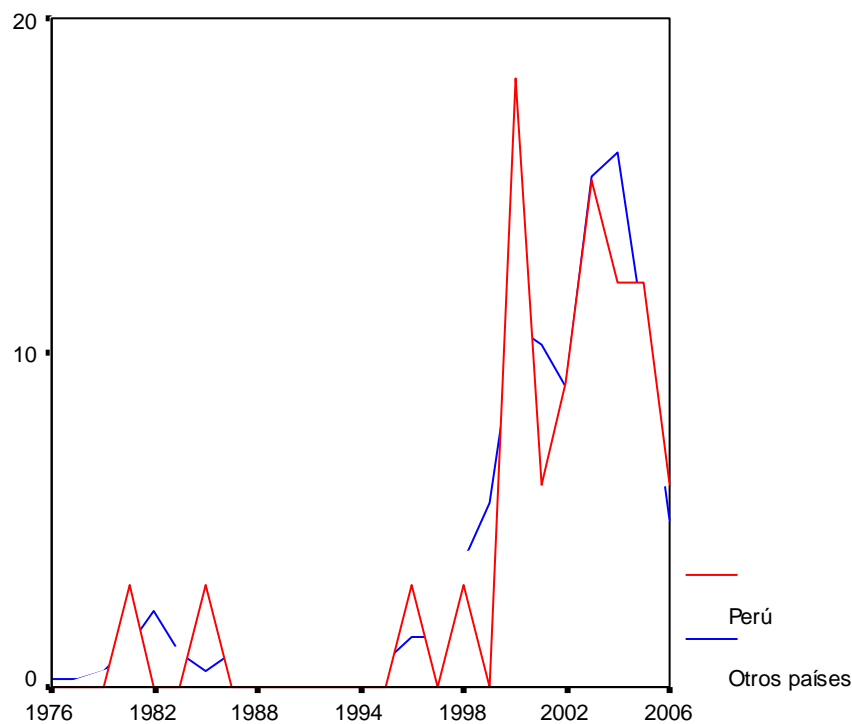


Figura N° 3.3. Año de publicación de los artículos o informes de evaluación de impacto en el Perú y otros países (Fuente: Elaboración propia).

Respecto a la comparación entre las áreas de estudio de los artículos o informes sobre EI tanto en el Perú como en el mundo, se encontró que la diversidad temática es característica en ambas circunscripciones; sin embargo, existe diferencia en la jerarquía que ocupan dichas áreas. Así, en el caso de los programas sociales de países extranjeros, sobresale la evaluación de impacto en educación (27.3%), salud (23%) y pobreza extrema (15%); mientras que en nuestro país se destacó, en primer lugar, la evaluación de programas sociales referentes a empleo (18%) y, en segundo lugar, la pobreza extrema (15.2%). Véase la Tabla 3.22.

La información de la Tabla 3.23 nos indica que, tanto en el Perú como en otros países, la EI de programas sociales está orientada hacia la población en general, más específicamente a las familias (Perú =54.5% y otros países = 47.8%). Asimismo, en nuestro país, es mayor el porcentaje de artículos sobre programas enfocados en poblaciones de niños y jóvenes (15.2% y 9.1%) si lo comparamos con los artículos del resto de países (9% y 6.3%). Por otro lado, en nuestro medio se carece de artículos o informes sobre la evaluación de impacto de aquellos programas sociales dirigidos a beneficiarios ancianos.

Tabla N° 3.22. Áreas de evaluación de impacto en el Perú y otros países (Fuente: Elaboración propia).

<i>Áreas</i>	<i>Otros países (n = 400)</i>	<i>Perú (n = 29)</i>
Agua y sanidad	3.5	6.1
Empleo	11.0	18.2
Educación	27.3	12.1
Salud	23.0	6.1
infraestructura y electrificación	3.5	12.1
Justicia	3.5	---
Alimentación	3.5	15.2
Pobreza extrema	15	12.1
Desarrollo urbano (Títulos de propiedad)	1.0	---
Micro crédito e ingresos familiar	2.8	6.1
Calidad de vida, convivencia social y medio ambiente	1.8	3.0
Agricultura y ganadería	3.8	6.1
No se especifica	0.5	3.0
Total	100.0	100.0

Tabla N° 3.23. Población beneficiaria de los programas evaluados en el Perú y otros países (Fuente: Elaboración propia).

<i>Beneficiarios</i>	<i>Otros países (n = 404)</i>	<i>Perú (n = 29)</i>
Población en general (familias)	47.8	54.5
Niños y niñas	9.0	15.2
Estudiantes de educación básica	11.3	9.1
Adultos mayores y ancianos	9.3	---
Jóvenes	6.3	9.1
Pacientes o personas discapacitadas	1.6	3.0
Mujeres	6.0	6.1
Profesionales o expertos	3.3	---
Países	2.5	---
No se especifica	1.8	3.0
Total	100.0	100.0

De manera coincidente, en nuestro país y el resto del mundo, los artículos indican, en mayor medida, que las evaluaciones de impacto de los programas sociales utilizaron como unidades de análisis a personas (51.5%) o familias (18.2%). No obstante, para el caso de nuestro país es significativamente mayor el porcentaje de los artículos que no especifican dicha información (12.1%). Véase la Tabla 3.24.

Tabla N° 3.24. Unidad de análisis considerada en los informes de evaluación de impacto en el Perú y otros países (Fuente: Elaboración propia).

<i>Unidades</i>	<i>Otros países (n = 404)</i>	<i>Perú (n = 29)</i>
Familias	18.3	18.2
Distritos, comunidades, Localidades o países	8.5	6.1
Personas	51.5	51.5
Estudios previos	3.0	---
Centros, instituciones u organizaciones	8.5	6.1
Programas sociales	6.5	6.1
No se especifica	3.7	12.1
Total	100.0	100.0

Aún cuando es alto el porcentaje de artículos de EI que no especifica la procedencia de sus datos de análisis, se encontró que en el Perú la preponderancia de los datos primarios sobre los secundarios es evidente, situación parecida a los

estudios de impacto de otros países. Sin embargo, que 30.3% de informes nacionales no especifiquen la naturaleza del origen de sus datos es un indicador de serias deficiencias metodológicas. Véase la Tabla 3.25.

Tabla N° 3.25. Origen de los datos utilizados para el análisis según se indica en los informes de evaluación de impacto en el Perú y otros países (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Otros países (n = 404)</i>	<i>Perú (n =29)</i>
Primarios	50.3	39.4
Secundarios	38.5	30.3
No se especifica	11.3	30.3
Total	100.0	100.0

Se encontró que tanto en los estudios de evaluación de impacto de programas sociales ejecutados en nuestro país como en el resto del mundo existe mayor preferencia por el uso de los registros de datos secundarios, en primer lugar; y la utilización de cuestionarios y entrevistas, en segundo y tercer lugar. No obstante, es significativamente alto el porcentaje de artículos que no brindaron información al respecto, siendo mayor en el caso de las evaluaciones realizadas en nuestro país (33.3% vs. 14.8%).

Tabla N° 3.26. Instrumentos utilizados para la recolección de datos según los informes de evaluación de impacto en el Perú y otros países (Fuente: Elaboración propia).

<i>Instrumentos</i>	<i>Otros países (n = 404)</i>	<i>Perú (n =29)</i>
Cuestionario	32.0	27.3
Fichas de observación	2.3	3.0
Registros de datos secundarios	36.8	30.3
Entrevista	14.3	6.1
No se especifica	14.8	33.3
Total	100.0	100.0

Asimismo, el porcentaje de informes de evaluación de impacto nacionales que no hacen mención a su metodología empleada es alto en comparación a los artículos del resto de países (36.4% vs. 13.3%). No obstante, tanto en nuestro país como en el resto del mundo las EI siguieron principalmente la metodología cuasi-experimental. Además, es significativo el porcentaje de artículos sobre evaluaciones de impacto realizadas en nuestro país donde se utilizó una metodología pre-experimental (24.2%). De igual forma, nuestro país carece de la publicación de artículos o informes respecto a evaluaciones de impacto con metodologías experimental o de meta-análisis.

Tabla N° 3.27. Metodología de investigación empleada en la evaluación de impacto en el Perú y otros países (Fuente: Elaboración propia).

<i>Métodos</i>	<i>Otros países (n = 404)</i>	<i>Perú (n =29)</i>
Experimental	10.3	---
Cuasi-experimental	41.8	30.3
Pre-experimental	18.5	24.2
Análisis cualitativo	14.5	9.1
Meta-análisis	1.8	---
No se especifica	13.3	36.4
Total	100.0	100.0

En la Tabla 3.28 se observa que los porcentajes de informes de evaluación de impacto donde no se menciona el diseño empleado son significativamente altos, siendo mayor en nuestro país (42.4%). Asimismo, son dos los diseños más frecuentes utilizados en nuestro país: diseño de diferencia en diferencias (doble diferencias) y diseño pretest-postest; mientras que en los informes internacionales se agrega a estos dos el diseño de las diferencias sólo postest. Por otro lado, se observa la ausencia de informes nacionales con diseños de discontinuidad de la regresión, descripción de medias y porcentajes y los estudios de casos.

Tabla N° 3.28. Diseños de evaluación utilizados en la evaluación de impacto en Perú y otros países (Fuente: Elaboración propia).

<i>Tipos</i>	<i>Otros países (n = 404)</i>	<i>Perú (n =29)</i>
Diferencias en diferencias	24.8	18.2
Discontinuidad de la regresión	1.0	---
Pretest – postest	15.0	18.2
Series de tiempo interrumpidas	9.0	6.1
Diferencias sólo postest	16.3	3.0
Descripción de medias y porcentajes	3.0	---
Costo beneficio	3.5	12.1
Estudios de caso	4.0	---
No se especifica	23.5	42.4
Total	400	100.0

La Tabla 3.29 muestra que el porcentaje de informes donde se menciona el empleo de muestras probabilísticas es similar para el Perú y el resto de países (21.2% y 24.3% respectivamente). Sin embargo, en el caso de nuestro país, el 75.8% no contiene información al respecto.

Tabla N° 3.29. Tipo de muestreo empleado en la evaluación de impacto en Perú y otros países (Fuente: Elaboración propia).

<i>Tipo</i>	<i>Otros países (n = 404)</i>	<i>Perú (n =29)</i>
Probabilístico	24.3	21.2
No probabilístico	18.0	---
Sin muestreo	14.8	3.0
Censo	1.0	---
No se especifica	42.0	75.8
Total	100.0	100.0

Asimismo, se encontró que en nuestro país un alto porcentaje de artículos no informan sobre las técnicas utilizadas para favorecer o determinar dicha evaluación; mientras que en el resto de países sólo se encontró una significativa falta de referencias respecto al uso o no de grupo control. No obstante, los estudios de evaluación de impacto en nuestro país como en el resto de mundo hacen referencia, en un porcentaje significativo, al uso de grupos control en sus procesos de análisis (48.5% y 54% respectivamente).

Tabla N° 3.30. Técnicas de control de variables extrañas utilizadas en la evaluación del impacto de los programas sociales del Perú y otros países (Fuente: Elaboración propia).

Técnicas	Otros países (n = 404)			Perú (n =29)		
	Sí	No	No específica	Sí	No	No específica
Grupo control	54.0	22.3	23.8	48.5	15.2	36.4
VARIABLES INSTRUMENTALES	7.5	91.3	1.3	9.1	69.7	21.2
Emparejamiento	19.3	80.0	0.8	27.3	51.5	21.2

Respecto a los tipos de análisis estadísticos utilizados en las evaluaciones de impacto, se encontró que tanto en nuestro país como en el resto de países existe un alto porcentaje que no hacen mención al respecto, siendo mayor para el caso de nuestro país (69.7% vs. 43.5%). No obstante, tanto las evaluaciones de impacto realizadas en nuestro país como en los otros países emplearon, en su mayoría, ecuaciones de regresión (21.2% y 39.5% respectivamente).

Tabla N° 3.31. Análisis estadísticos empleados en la evaluación de impacto en el Perú y otros países (Fuente: Elaboración propia).

<i>Tipos</i>	<i>Otros países (n = 404)</i>	<i>Perú (n =29)</i>
Ecuaciones de regresión	39.5	21.2
Ecuaciones estructurales	0.8	---
Estadísticas de comparación de medias	8.8	---
ARIMA's	1.0	---
Estadísticas descriptivas (correlación, porcentajes, etc.)	5.0	9.1
Análisis categórico cualitativo	1.5	---
No se especifica	43.5	69.7
Total	100.0	100.0

Respecto al impacto obtenido por los programas sociales, se observa un porcentaje significativamente alto de informes que no hacen mención sobre este aspecto (24.2%). No obstante, en nuestro país el porcentaje de programas o proyectos sociales evaluados que obtuvieron un impacto positivo es menor que en el resto de países (63.6% vs. 80.8%) o inespecífico (24.2% vs. 2.8%).

Tabla N° 3.32. Impacto alcanzado por el programa social evaluado en Perú y otros países (Fuente: Elaboración propia).

<i>Impacto</i>	<i>Otros países (n = 404)</i>	<i>Perú (n =29)</i>
Positivo	80.8	63.6
Negativo	16.5	12.1
No se especifica	2.8	24.2
Total	100.0	100.0

En síntesis, las únicas fuentes de difusión y producción de EI en el Perú son las Instituciones Nacionales y Organismos Internacionales, quienes han aportado el 100% de la producción de informes. La principal área de interés es el empleo y la lucha contra la pobreza. Similar al ámbito internacional, las EI en el Perú han aumentado geométricamente durante la última década, con una disminución en la década del 90 quizá debido a la coyuntura política. En términos metodológicos, las EI nacionales subinforman u omiten importantes aspectos metodológicos en sus informes, dando indicios de deficiencias. Igual que en el extranjero, el diseño más frecuente es el cuasi-experimental, aunque en nuestro país no se han realizado

diseños experimentales ni meta-analíticos; si se han empleado más técnicas de control de sesgos como el *matching* o las variables instrumentales. En cuanto al impacto de los programas evaluados, en el Perú se reportan menos programas con efectos negativos, pero 24.2% no especifica adecuadamente impacto alguno, hecho debido a las deficiencias metodológicas mencionadas.

3.2.2.2. Desarrollo metodológico: Nuevas estrategias

Los métodos de evaluación clásicos basados en estudios costo-beneficio, costo-eficiencia o de estudios de impacto, han sido utilizados principalmente por gobiernos y organismos internacionales en su necesidad de evaluar intervenciones de política, con la idea de brindar elementos objetivos sobre la racionalidad de los programas y evitar su fracaso. Uno de los campos en donde estas evaluaciones han sido más utilizadas es precisamente el campo de las políticas activas de empleo y, en particular, en los programas de capacitación y formación para el trabajo en grupos objetivo vulnerables, como pueden ser los jóvenes de bajos ingresos o los desempleados de larga duración. Sin embargo, desde medios académicos y, posteriormente, desde su adopción por organismos multilaterales y consultores independientes, ha sido creciente la utilización de métodos más elaborados que implican el uso de técnicas estadísticas y métricas más avanzadas que los métodos más clásicos. La localización de problemas serios de simultaneidad y de sesgos importantes en las estimaciones de los indicadores que medían el impacto de los programas, por parte de académicos y estudiosos del tema, llamaron la atención hacia problemas antes ignorados. La necesidad de contar con métodos rigurosos de validación y, a la vez, los avances en la teoría estadística y de la medición social, pusieron al alcance de la práctica una cantidad de opciones metodológicas que prometían ser métodos que aportarían mayor rigurosidad a la hora de discutir la propia existencia de varios programas de empleo y formación (Diez de Medina, 2003). Este es pues el origen de la metodología cuasi-experimental.

Desde entonces, el uso de esta metodología se ha extendido por todo el mundo, tal como se ha demostrado en el subtítulo anterior, estableciéndose una etapa de desarrollo “normal” basado en los aportes originales de Campbell y Stanley (1966) y sus modificaciones. Sin embargo, recientemente se han venido publicando, en diversas revistas internacionales, aportes significativos que salen de ese “parámetro normal” de la metodología cuasi-experimental y que requieren una revisión detallada para captar sus posibilidades de aplicación en los ámbitos de evaluación, los cuales requieren constantemente ser optimizados.

De lo dicho, a continuación se resumen algunos aportes metodológicos novedosos útiles para la evaluación de impacto de programas sociales publicados en revistas internacionales.

Por ejemplo, Bustelo (2004) analiza la importancia del establecimiento de las normas, estándares y códigos en la creación de una cultura de evaluación, para orientar a los profesionales sobre la práctica profesional y guiar la orientación ética. Según Bustelo, es importante analizar cuestiones relacionadas con los propósitos y la concepción de la evaluación, cuestiones relacionadas con la naturaleza valorativa de la evaluación, cuestiones relacionadas con la gestión de la

evaluación: tipos y calendario, cuestiones relacionadas con la inversión en evaluación y cuestiones relacionadas con la comunicación de los resultados.

Vivo (2006) muestra el uso de simulaciones estadísticas computarizadas para analizar la inferencia causal de la evaluación de impacto de los programas. Estos procedimientos son útiles para comparar mediante simulación las limitaciones y bondades de los principales diseños de evaluación de impacto. Se usa la simulación experimental computarizada (uso de datos simulados, inferencia causal, aleatorización, diferencia en diferencias).

Michael, Schwerina, Michaelb, Glaserc & Farrar (2002), describen un sistema de evaluación de programas que puede ser aplicado a una amplia gama de programas de Calidad de Vida, donde la visión es holística y de objetivo único. Las medidas de evaluación fueron diseñadas específicamente para cada programa consistiendo en un grupo clave de ítems que evaluaron cómo cada programa influye en los resultados específicos del programa.

Bifulco (2002) propone alternativas para determinar los sesgos de la autoselección en las evaluaciones cuasi-experimentales de la reforma educativa. El autor demuestra que desarrollar una estrategia de variable instrumental puede ser utilizada para promover los estimadores comunes de valor agregado cuando sólo se tenga a disposición, medidas post tratamiento. De esta forma, las amenazas potenciales para la validez interna que poseería un programa con participantes auto-seleccionados en comparación con aquél donde los participantes fueron obtenidos de manera aleatoria, se controlarían mediante la estrategia de la variable instrumental, pues brindaría estimaciones de los impactos del modelo similar a las brindadas por el estimador diferencia en la diferencia.

Shu-Hong Zhu (1999) propone un método para obtener un grupo control aleatorio donde parece imposible. Según el autor, aun en los trabajos de campo aparecen alternativas que permiten al investigador obtener un grupo control aleatorio sin comprometer los estándares éticos o afectar la probabilidad de una potencial inclusión del participante en el estudio. Analizando el diseño utilizado en un gran proyecto de salud en California, *The California Smokers' Helpli*, demostró la posibilidad de obtener grupos de control aleatorios en caso de programas de intervención social.

Curria, Kinga, Rosenbaumb, Law, Kertoyc & Jacqueline (2005) proponen un modelo de los impactos de las sociedades investigadores en servicios de salud para ayudar a los investigadores, evaluadores, miembros de la comunidad, y a las unidades de financiamiento a reflexionar sobre los impactos desde un modo amplio y multidimensional. Este modelo de sistemas de procesos orientados representa tres grandes dominios del impacto a medio plazo (incrementar el crecimiento, favorecer el estudio de habilidades y utilizar la formación). El modelo brinda a las asociaciones investigadoras herramientas para demostrar su responsabilidad y para facilitar las operaciones e impactos.

Madueño, M, Linares, J. & Zuria, A. (2004) han elaborado un instrumento estandarizado de identificación de beneficiarios para programas sociales en el Perú. El instrumento es estimado sobre la base de un conjunto de variables relacionadas con el nivel de bienestar ("proxy mean test") agrupadas en cuatro dimensiones de análisis: (a) características socio-demográficas, (b) características de la vivienda,

(c) tenencia de activos durables y (d) de locación. Así, mediante el uso de técnicas multivariantes (basadas en el análisis de componentes principales y escalas óptimas), se define un sistema de puntuaciones por variables y categorías de variables que posibilitan el cálculo del IBEH y, por ende, un ordenamiento de la población. Finalmente, los autores plantean el uso del método de conglomerados para estratificar grupos homogéneos al interior pero heterogéneos entre sí y determinar los umbrales de bienestar para la elegibilidad de los hogares a los subsidios públicos. Los estimados muestran que el IBEH tiene una mejor capacidad de discriminación que otros indicadores tradicionales utilizados para la segmentación de la población (Ej. gasto per cápita, nivel de pobreza, entre otros), lo cual se traduciría en resultados más eficientes en términos de reducción de errores de focalización. Asimismo, el menor costo relativo en la recopilación de información y el uso de información de mejor calidad y de fácil verificación favorecen su aplicación como instrumento de focalización individual en ámbitos urbanos.

IV. ESTRATEGIAS METODOLÓGICAS DE EVALUACIÓN DE IMPACTO: MODELO ARYSIS

En la revisión sistemática se ha encontrado que la principal fuente de difusión y producción de EI en el mundo son los Organismos Internacionales, quienes han aportado más del 75% de la producción de informes, centrándose en aspectos educativos, de salud, pobreza y empleo. Se ha encontrado también que las EI han aumentado geométricamente durante los últimos 30 años, concentrándose más del 77% en el periodo 2000-2006. Siendo en definitiva, un método muy útil y cada vez de mayor aceptación.

En términos metodológicos, se ha encontrado que el diseño más usado es el cuasi-experimental (41%), principalmente el de doble diferencias; asimismo, se está empezando a usar técnicas especiales de control, como el caso de las variables instrumentales (7.6%) y el *matching* (19.7%). Estas estrategias están vigentes y tienen cada vez mayor aceptación y desarrollo.

Por eso, tales estrategias metodológicas se han integrado en un modelo de evaluación denominado ARYSIS, el cual intenta unificar los diversos procedimientos metodológicos dentro de una perspectiva teórica uniforme.

El modelo de Evaluación de Impacto ARYSIS se ha construido bajo el enfoque de la Teoría General de los Sistemas, el cual está caracterizado por ser permanente, cíclico, autorregulado y abierto. Se concibe al Modelo de Evaluación de Impacto ARYSIS como un conjunto dinámico de principios, parámetros teóricos, criterios metodológicos y procedimientos estandarizados, organizados y regulados en subsistemas, que se dirigen hacia la identificación y obtención de datos pertinentes y válidos.

El Sistema de Evaluación ARYSIS tiene una lógica basada en la evaluación de resultados, el cual debe orientarse al aumento del conocimiento sustantivo como herramienta clave en la gestión. Se persigue convertir los datos recogidos en información útil para la implantación de los proyectos, buscando optimizar el impacto y la eficiencia en la gestión futura. La siguiente figura comprende los principales elementos constitutivos del Sistema ARYSIS, los cuales serán presentados y analizados en los acápites siguientes.

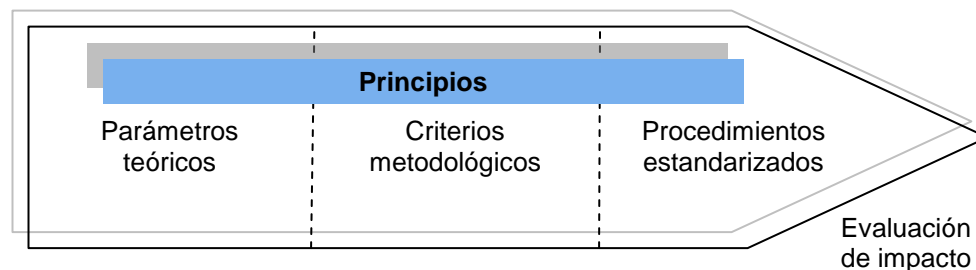


Figura N° 4.1. Modelo de Evaluación de Impacto ARYSIS (Fuente: Elaboración propia).

Tal como se muestra en la Figura 4.1, el Modelo ARYSIS tiene un contenido transversal, de naturaleza ética y actitudinal constituida por los principios que rigen los demás componentes. En sentido programático, se encuentran los parámetros teóricos, primer elemento constitutivo de naturaleza cognoscitiva orientada a fundamentar las acciones del evaluador. Sin estos parámetros teóricos, los criterios metodológicos no serán lo suficientemente exhaustivos ni tendrán la suficiente garantía de validez. Los criterios metodológicos, por su parte, definen los procedimientos estandarizados que se seguirán para realizar la evaluación de impacto de la forma más eficaz y eficiente posible.

4.1. PRINCIPIOS

Todo modelo sistémico contiene principios que rigen sus elementos constitutivos⁵¹. El modelo ARYSIS no es la excepción y se basa en los principios propuestos por la *American Evaluation Asociation* (AEA, 1994; 2004) y por otras instituciones internacionales.

Debe recordarse que las normas, estándares, guías orientativas y/o códigos en evaluación, surgen a partir de la década de los años ochenta y noventa en algunos países, cumpliendo una doble función⁵²: a) por un lado, orientan a los

⁵¹ Según la XXII edición del Diccionario de la Real Academia Española, el término "principio" significa, entre otros, "punto que se considera como primero en una extensión o cosa", "base, origen, razón fundamental sobre la cual se procede discuriendo en cualquier materia", "causa, origen de algo", "cualquiera de las primeras proposiciones o verdades fundamentales por donde se empiezan a estudiar las ciencias o las artes". Así, los principios son ideas, postulados éticos o escritos fundamentales que condicionan y orientan la acción evaluativa.

⁵² Los primeros códigos en elaborarse en el campo de la evaluación de programas y políticas fueron en el ámbito anglosajón. El primero y más importante de todos, las normas del Joint Committee, en su primera edición en 1981 y su segunda en 1994. Luego le siguieron los códigos de la Asociación Americana de Evaluación, 1994; la Asociación Canadiense de Evaluación, 1996; y la Asociación Australiana de Evaluación, 1997. También se aprobaron en 2001 unas orientaciones de evaluación por parte de la Asociación Africana de Evaluación, que son una adaptación a las normas del Joint Committee y en las modifican aproximadamente un 40% de las normas previamente existentes. En el ámbito europeo, se han aprobado códigos u orientaciones de las Sociedades de Evaluación francesa (2003), alemana (2001), suiza (2000) y británica. Mientras los modelos suizo y alemán siguen la estructura de las normas del Joint Committee, el francés está formulado en formato de grandes principios éticos (pluralismo, independencia, competencia, respeto, transparencia y responsabilidad), y el inglés hace una interesante separación de normas según vayan dirigidas a

profesionales sobre la práctica profesional y son, además, una guía de orientación ética; y b) por otro, son un indicador de la institucionalización de la función de evaluación en un determinado país y, en parte, de la consolidación de la evaluación como ámbito académico y profesional (Bustelo, 2004).

Al respecto, ARYSIS tiene dos tipos básicos de principios: a) los principios del evaluador y b) los principios del proceso de evaluación. Los principios del evaluador son cinco interrelacionados y no jerárquicos, mientras que los principios del proceso de evaluación son cuatro principios jerárquicos (véase la siguiente figura).

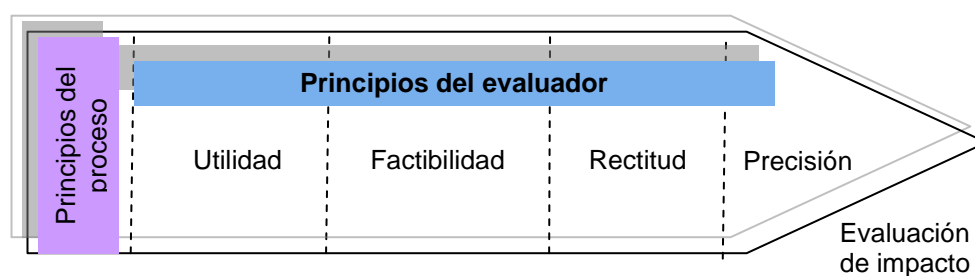


Figura N° 4.2. Principios de la Evaluación de Impacto ARYSIS (Fuente: Elaboración propia).

4.1.1. Principios del evaluador

La Asociación Americana de Evaluación (AEA, 2004) ha elaborado una guía de principios para el evaluador. Estos principios han sido consagrados para orientar a los evaluadores en el resguardo de su conducta ética y de sus competencias como investigadores. De este modo, los principios sirven para orientar a los evaluadores sobre la sistematicidad en la investigación, las competencias para desempeñarse como evaluador; la integridad y honestidad en los estudios; el respeto a las personas; y la responsabilidad ante el bienestar general y público de todos aquellos que participan de estos estudios. Estos principios se relacionan constitutivamente tal como se muestra en la siguiente figura.

evaluadores, comanditarios, participantes, o implicados en una auto-evaluación. Recientemente el Grupo de Evaluación de las Naciones Unidas (2005) ha elaborado los *Standards for Evaluation in the UN System*.

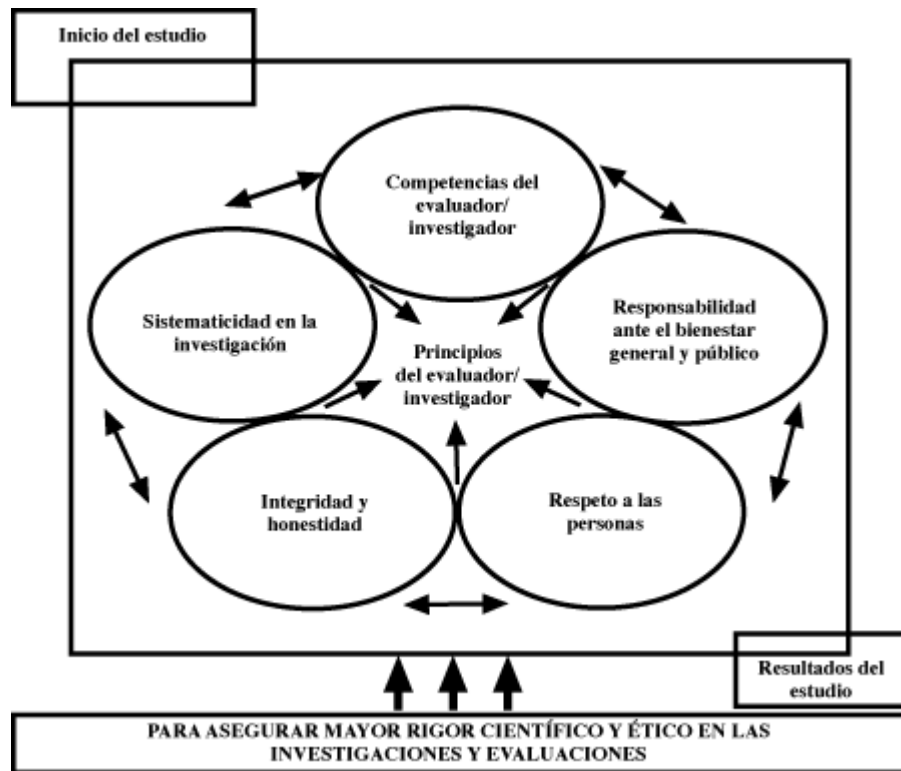


Figura N° 4.3. Principios básicos del evaluador de impacto en ARYSIS (Fuente: Elaboración propia).

En estos principios no solo se reconoce la debida solidez científica que debe poseer un evaluador, sino que también se destacan valores tales como la honestidad, integridad, respeto y responsabilidad frente al acto de evaluar.

Debido a que tienen una relación simétrica funcional, ninguno de estos principios tiene supremacía sobre el otro, por el contrario, su integración e interrelación contribuyen a marcar el componente ético de la evaluación. A continuación se detalla cada principio incluido.

4.1.1.1. Principio de sistematicidad en la evaluación

Este principio apunta al criterio valor de “verdad” y “validez”, indicando que el evaluador debe adherirse a los más exigentes estándares metodológicos en la conducción de su estudio para incrementar así la validez, precisión y credibilidad de la información producida por la evaluación.

En la búsqueda de esta sistematicidad, el evaluador deberá explorar con las personas que requieren el estudio, cuáles son las fortalezas y debilidades de las interrogantes que se pretenden resolver, de tal manera que oriente adecuadamente el estudio para responder a dichas interrogantes. Lo anterior se complementa con la formalidad de comunicar a los clientes y audiencias participantes, las técnicas de recolección y análisis de datos para permitirles comprender, interpretar y criticar

los resultados de la evaluación, procurando un lenguaje simple pero efectivo y con el debido acercamiento al contexto donde tuvo lugar el estudio⁵³.

4.1.1.2. Principio de las competencias para desempeñarse como evaluador

El evaluador deberá poseer las habilidades, destrezas y experiencia necesarias para conducir las tareas propias de una evaluación. No es suficiente con conocer el tema, se requiere experiencia en el desarrollo de una evaluación. Sin dejar de desconocer la experiencia, se requiere también de la debida capacitación como conductor de evaluaciones y especialista en las temáticas centrales de dichas evaluaciones.

La evaluación de impacto es una actividad muy especializada y requiere, por tanto, la confluencia de muchas competencias. Si por algún motivo el evaluador reconoce que no posee alguna de las competencias requeridas para el desarrollo del estudio, debe declinar su participación en favor de su prestigio y de la credibilidad de la evaluación. Esto también lo debería impulsar a estar en constante perfeccionamiento.

4.1.1.3. Principio de la integridad y honestidad frente a los estudios

El evaluador deberá ser honesto frente a las negociaciones que debe considerar respecto a los costos del estudio, las limitaciones de la metodología a emplear y las restricciones de los resultados para ser repetidos o aplicados a otros participantes distintos al contexto donde se realizó el estudio.

Por otro lado, la integridad y honestidad del evaluador se prueba cuando éste es capaz de expresar sus temores, dudas, inconvenientes frente a aquellos que solicitaron la evaluación (clientes), promoviendo las debidas instancias de conversación, con el fin de atenuar las posibles diferencias y así poder atender mejor a los requerimientos de la evaluación. Ahora, si no logra superar las diferencias que tiene con los clientes o audiencias, antes de iniciar las etapas más cruciales de la evaluación, debería dejar el estudio a tiempo. El evaluador no deberá, bajo ningún motivo, distorsionar los resultados para favorecer al cliente.

4.1.1.4. Principio de respeto a las personas

Este principio apela a la dignidad humana como valor esencial de toda persona. Indica que el evaluador deberá respetar a las personas que participan de la evaluación en su amplio sentido, cautelando la confidencialidad de los datos individuales, el debido consentimiento informado de los entrevistados y la

⁵³ Michel Patton (1997) es uno de los principales precursores del mejoramiento de la utilización de los resultados de las evaluaciones. Indica que un factor que atenta contra el uso de los resultados, es la casi nula o vaga explicación de los procedimientos empleados para recoger los datos y la forma en que se analizan, quedando muchas veces los clientes y audiencias con estudios que solo han servido para incrementar el número de libros de una estantería o simplemente justifican que ha sido tomada con anterioridad a la realización del estudio.

evitación de daños y perjuicios morales producto de esa participación. Entonces, el evaluador deberá buscar el máximo beneficio, reduciendo cualquier tipo de daño innecesario que pueda ocurrir durante y después de ocurrido el estudio.

Adicionalmente, el evaluador deberá conducir su estudio y comunicar sus resultados respetando la dignidad y valor intrínseco que tienen las personas, más allá de buscar obsesivamente un resultado que tienda a beneficiar sus personales intereses o los de su cliente. Lo anterior también implica respetar las naturales diferencias entre los participantes, tales como la cultura, la religión, sexo, discapacidades, edad, entre otras cosas más.

4.1.1.5. Principio de responsabilidad ante el bienestar general y público

El evaluador deberá ser responsable frente a los potenciales efectos colaterales que pueda tener su estudio. Este principio invita a los evaluadores a manejarse con una amplia perspectiva valórica al momento de comunicar los resultados del estudio, evitando propiciar estereotipos raciales o poblacionales.

Adicionalmente, el evaluador no debe caer en la tentación de responder a los particulares intereses del organismo que solicita la evaluación en detrimento del grupo poblacional, es decir, no se debe prestar al desarrollo de evaluaciones que de alguna u otra forma pueden dañar el medio social o la integridad moral de algunos grupos poblacionales sujetos de estudio.

4.1.2. Principios para el proceso de evaluación

El Modelo ARYSIS también tiene principios que guían el proceso de la evaluación de impacto de los programas. Estos principios están basados en los propuestos por el *Joint Committee on Standards for Educational Evaluation* (1994).

El cumplimiento de estos principios mejora sensiblemente la calidad de las evaluaciones. En efecto, estos principios parten de la base de que la calidad de una evaluación está asociada a cuatro atributos básicos: a) utilidad, b) factibilidad, c) rectitud y d) precisión.

4.1.2.1. Utilidad

El principio de utilidad busca garantizar que una evaluación ofrezca la información y las condiciones necesarias para que las personas afectadas o involucradas por la evaluación usen sus resultados para mejorar. Este atributo aparece deliberadamente en primer lugar, ya que sin utilidad ninguno de los otros principios es relevante. Este principio tiene 7 indicadores:

1. Identificación de los interesados. Las personas involucradas o afectadas por la evaluación deben ser identificadas para poder considerar sus intereses.

2. Credibilidad del evaluador. Las personas que realizan la evaluación deben de ser tanto confiables como competentes para realizar la evaluación, con el fin de que los resultados de la evaluación alcancen el máximo de credibilidad y aceptación.
3. Selección de la amplitud y profundidad de la información. La información a colectar debe ser seleccionada para responder las preguntas pertinentes acerca del programa y corresponder a los intereses de los clientes y otros beneficiarios. Pero sin incluir aspectos secundarios o irrelevantes.
4. Identificación de los criterios de evaluación. Los enfoques, procedimientos y racionalidad utilizados para interpretar los resultados deben ser claramente descritos, de manera que los criterios que sustentan las conclusiones sean claros.
5. Claridad en los informes. Los informes de evaluación deben describir claramente el programa a evaluar, incluyendo su contexto, los propósitos, los procedimientos y los resultados de la evaluación, de tal manera que se ofrezca la información esencial de manera comprensible.
6. Puntualidad de los informes y difusión de resultados. Los resultados intermedios relevantes y los informes de la evaluación deben de ser difundidos entre los usuarios potenciales, de tal forma que puedan ser usados oportunamente.
7. Impacto de la evaluación. Las evaluaciones deben ser planeadas, ejecutadas y reportadas de manera que motiven una respuesta por parte de los afectados, a fin de aumentar la probabilidad de que la evaluación sea de calidad óptima.

4.1.2.2. Factibilidad

El principio de factibilidad pretende asegurar que la evaluación será realista, prudente, diplomática y frugal. Bajo este principio se entiende que:

1. Los procedimientos de la evaluación deben ser prácticos a fin de reducir los contratiempos al mínimo durante los procesos de colecta de información (practicidad).
2. El proceso de evaluación debe ser planeado y ejecutado previendo las diferentes posiciones de los grupos de interés, a fin de que se pueda lograr su cooperación y evitar hasta donde sea posible, los intentos de estos grupos por dificultar o restringir el proceso de evaluación, así como desvirtuar y minimizar los resultados de la misma (viabilidad política).
3. La evaluación debe ser eficiente y producir información lo suficientemente valiosa como para que los recursos utilizados en ella puedan ser justificados (costo razonable).

4.1.2.3. Rectitud

El principio de rectitud está diseñado para asegurar que la evaluación sea efectuada de manera legal y éticamente correcta, así como con el debido respeto hacia el bienestar de los que están involucrados en el proceso de evaluación y de todos aquellos que pudieran ser afectados por sus resultados. Tiene 8 indicadores:

1. Orientación al servicio. Las evaluaciones deben ser diseñadas para ayudar a las organizaciones a dirigir y atender de manera efectiva las necesidades del mayor número posible de participantes.
2. Acuerdos formales. Las obligaciones adquiridas por las partes involucradas en el proceso de evaluación (lo que ha de hacerse, cómo, por quiénes y cuándo) deben ser expresadas por escrito, de manera que las partes involucradas se adhieran a todas las condiciones de los acuerdos o puedan renegociar de manera formal.
3. Respeto a los derechos humanos. Las evaluaciones deben ser diseñadas y operadas para respetar y proteger los derechos y el bienestar de las personas.
4. Interacción humana. Los evaluadores deben respetar la dignidad humana así como considerarla en su interacción con otras personas relacionadas con la evaluación, de modo que los participantes no se vean intimidados o perjudicados.
5. Dictámenes equilibrados y justos. La evaluación debe ser íntegra y equitativa en el examen y registro de las fortalezas y debilidades del programa a evaluar, a fin de que las fortalezas y los problemas puedan ser ponderados con buen juicio y prudencia.
6. Divulgación de resultados. Los encargados del proceso de evaluación deben asegurarse que el conjunto de resultados, con las limitaciones pertinentes, lleguen a las personas afectadas por la evaluación, así como cualquier otra que tenga derecho legal de conocer sus resultados.
7. Conflicto de intereses. Los conflictos de intereses deben atenderse de manera abierta y honesta, a fin de que esto no arriesgue el proceso y resultados de la evaluación.
8. Responsabilidad administrativa. El destino y gasto de los recursos utilizados en el proceso de evaluación deben reflejarse mediante procedimientos contables transparentes; es decir, los encargados de la evaluación están obligados a ser prudentes y éticamente responsables, de que los recursos sean contabilizados y gastados adecuadamente.

4.1.2.4. Precisión

Las normas de precisión están diseñadas para asegurar que la evaluación revele o muestre la información técnica requerida, sobre los indicadores que fundamentan el valor o mérito del programa que está siendo evaluado. Tiene 11 indicadores:

1. Descripción del programa. El programa que está siendo evaluado debe ser descrito y documentado de forma clara y precisa, de modo que esté plenamente identificado.
2. Análisis contextual. El contexto en el que existe el programa debe ser examinado con suficiente detalle, a fin de que los factores de influencia puedan ser identificados.
3. Descripción de propósitos y procedimientos. Los propósitos y procedimientos de la evaluación deben ser analizados y descritos con suficiente detalle a fin de que estén perfectamente identificados.
4. Fuentes de información confiables. Las fuentes de información utilizadas en la evaluación de un programa deben ser descritas con suficiente detalle a fin de que la pertinencia de la información esté sustentada.
5. Información válida. Los procedimientos para recopilar la información deben elegirse, desarrollarse e implementarse, asegurándose así que la información obtenida sea válida dentro del propósito de la evaluación.
6. Información confiable. Los procedimientos para recopilar la información deben elegirse, desarrollarse e implementarse para que se asegure que la información obtenida es lo suficientemente confiable para el proceso de evaluación.
7. Información consistente. La información recopilada, analizada y reportada en la evaluación debe ser verificada sistemáticamente a fin de detectar y corregir errores en forma oportuna.
8. Análisis de información cuantitativa. La información cuantitativa utilizada en una evaluación debe ser analizada de manera apropiada y sistemática para que las preguntas de la evaluación sean efectivamente respondidas.
9. Análisis de información cualitativa. La información cualitativa utilizada en una evaluación debe ser analizada de manera apropiada y sistemática para que las preguntas de la evaluación sean efectivamente respondidas.
10. Conclusiones fundamentadas. Las conclusiones a las que se llega en un proceso de evaluación deben estar fundamentadas explícitamente para que los interesados en la evaluación puedan apreciarlas y estimarlas correctamente.
11. Informe imparcial. Los procedimientos para informar a los interesados deben evitar las distorsiones ocasionadas por sentimientos personales y sesgos en cualquier fase de la evaluación, a fin de que los informes (verbales o escritos) reflejen fielmente los hallazgos de la evaluación.

4.2. PARÁMETROS TEÓRICOS

El Modelo de Evaluación de Impacto ARYSIS se fundamenta en un único parámetro teórico básico. Éste parámetro se refiere a la validez causal de los resultados. Debido a que si la evaluación de impacto mide los efectos producidos por el programa, ¿cómo se puede estar seguro o tener la certeza que los efectos medidos corresponden causalmente al programa?, ¿no podrían ser, acaso, consecuencia de otras situaciones y no del programa? Erróneamente se puede atribuir causalidad a un programa cuando en realidad los efectos medidos son consecuencia de otras variables o circunstancias. Así, el gerente social especialista en evaluación debe utilizar un parámetro teórico que le permita controlar – previamente- el nivel de validez de su modelo de evaluación.

4.2.1. La inferencia causal

Es evidente que no todos los diseños de investigación social están orientados a determinar las causas de determinados fenómenos o las consecuencias de ciertas intervenciones. Sin embargo, en una disciplina aplicada como la gerencia social este tipo de planteamientos tienen una importancia fundamental. En efecto, las preguntas de investigación que motivan la mayoría de los estudios en las ciencias sociales son de naturaleza causal. Por ejemplo: ¿Cuál es la eficacia de un medicamento en una población dada?, ¿Qué fracción de muertes de una enfermedad pudieron haberse evitado por un tratamiento o política dada?, ¿La descentralización de las escuelas mejora la calidad de las mismas?, son preguntas motivadas por las preocupaciones políticas y por las necesidades gubernamentales de decidir el destino de los programas sociales (Galiani, 2006).

El objetivo del análisis estadístico estándar (típicamente probabilidad y otras técnicas de estimación) es inferir parámetros de una distribución, a partir de muestras obtenidas de esa distribución. Con la ayuda de tales parámetros, uno puede: 1) Inferir asociación entre variables, 2) estimar la probabilidad de eventos pasados y futuros, 3) actualizar la probabilidad de eventos a la luz de nueva evidencia o nuevas mediciones. Sin embargo, el análisis causal va un paso adelante. Su objetivo es inferir aspectos del proceso de generación de datos para deducir no solamente la probabilidad de eventos bajo condiciones estáticas, sino también la dinámica de eventos bajo condiciones cambiantes. Esta capacidad incluye: 1) predecir los efectos de intervenciones, 2) predecir los efectos de cambios espontáneos, 3) identificar las causas de eventos reportados; demostrando que los conceptos causales y de asociación no se mezclan (Galiani, 2006).

La palabra causa no está en el vocabulario de la teoría de probabilidad estándar. Toda teoría de probabilidad permite decir que dos eventos están mutuamente correlacionados o son dependientes – lo que significa que, si se encuentra uno, se puede esperar encontrar el otro. De lo dicho, los científicos que buscan explicaciones causales para fenómenos complejos o racionales para decisiones políticas deben, por tanto, suplir el término de probabilidad por causalidad. Por ello, en los últimos años se han desarrollado procedimientos

diversos para intentar contrastar modelos causales a partir, incluso, de estudios de carácter descriptivo, tales como el análisis de ruta (*path-analysis*) o el análisis de modelos de ecuaciones estructurales lineales (Tuijman y Keeves, 1997; Cook y Campbell, 1979; Julmes, 2004) o el modelo *Neyman-Rubin potential outcome* (Rubin, 1974).

4.2.1.1. El problema de la validez de las evaluaciones

No es secreto que la investigación sobre gerencia social y administración de programas ha tenido una serie de falencias y debilidades (Cordray, 1986; Wienert, 1997). Por ejemplo, en 1984, la Oficina de Auditoria General de los Estados Unidos descartó 55 de 61 informes de investigación⁵⁴ porque tenían serios problemas metodológicos y, consiguientemente, escasa validez (Cordray, 1986). En otra área, el Departamento de Educación de Estados Unidos solicitó al Consejo Nacional de Investigación que evaluara los principales estudios sobre educación bilingüe, encontrando resultados frustrantes (Meyer y Conrad, 1992): *"Los estudios no autorizan la conclusión de que ningún tipo de programa sea superior a ningún otro, ni que los programas sean igualmente efectivos. Incluso si alguno de los programas hubiera sido superior, los estudios, tal y como se planificaron y llevaron a cabo, podrían haber fracasado en identificar tal circunstancia"*. (p.104). *"Los estudios [revisados] adolecen de una excesiva atención al uso de elaborados métodos estadísticos para intentar solventar las deficiencias en los diseños de investigación. Los métodos de análisis estadístico no pueden corregir los fallos en la conceptualización, diseño y ejecución de los estudios"*. (p. 104).

Es claro que esta baja calidad de investigaciones amenaza la reputación de la evaluación de programas (Aliaga, 2000; Cordray, 1986) y dispone que diversos grupos políticos y "académicos" arremetan diversos argumentos contra la evaluación de impacto. Así pues surgen algunas frases como estas:

- *"Los métodos de investigación en ciencias sociales no son adecuados (o capaces) de explicar la compleja realidad social y humana"*. Este argumento es típico de quienes creen únicamente en "modelos alternativos", tales como la investigación cualitativa o interpretativa (Keeves, 1997; Walker y Evers, 1997). Aunque estos planteamientos parecen prometedores, aún no se ha demostrado su utilidad y validez para la medición de impacto. En efecto, aunque desde la perspectiva cualitativa se han desarrollado criterios de calidad, su aceptación aún es desigual; y a lo mucho se la considera como complemento de la investigación cuasi-experimental de impacto y no como su reemplazo (Campbell, 1974; Cronbach, 1974).
- *"Los métodos de investigación en ciencias sociales están aún poco evolucionados y no se han desarrollado suficientemente como para ser capaces de detectar la eficacia de los distintos tratamientos"*. El desarrollo de cierta proporción de investigaciones bien controladas

⁵⁴ Analizaban la influencia de los programas sociales sobre el peso de los niños al nacer.

y fiables (como, por ejemplo, en el caso de 6 de los 61 estudios revisados por la Oficina de Auditoría General de los Estados Unidos, 1984) permite descartar esta alternativa. En todo caso, el problema no es la falta de métodos de investigación evolucionados, sino su inadecuada utilización y carencia de conocimientos especializados sobre los mismos (Sáez et al, 1994; Suárez et al, 1995).

- “*El trabajo empírico en ciencias sociales supone complicaciones específicas que hacen muy difícil llevar a cabo investigaciones válidas*”. Algunos de los problemas de este tipo han sido analizados en los últimos treinta años, básicamente gracias al trabajo de Campbell y sus colaboradores (Campbell & Stanley, 1966; Cook & Campbell, 1979; Campbell, 1986; Cook, Campbell & Peracchio, 1990), instigados por planteamientos críticos como los de Cronbach (1982), de tal modo que se han desarrollado procedimientos específicos, como es el caso de los diseños cuasi-experimentales, que posibilitan, cuando menos, una comprensión global del proceso de investigación en las particulares circunstancias que se dan en las ciencias sociales.

A mi entender, todos estos argumentos son superficiales y desconocen el nivel de desarrollo actual de la teoría evaluativa, por cuanto la rechazan con creencias poco fundadas y más cargadas de sesgos dogmáticos. De lo dicho, pienso que la principal razón por la cual se encuentran evaluaciones de impacto de baja calidad es porque los profesionales e investigadores no han comprendido o aplicado adecuadamente los conocimientos desarrollados sobre evaluación de impacto; y, más aún, desconocen los condicionantes de su validez que, conjuntamente con el desarrollo de los procedimientos de investigación (particularmente la cuasi-experimentación), se han desarrollado en las últimas décadas. Recuérdese que ya Campbell & Stanley (1966) sostenían que *“una capacitación más a fondo de los investigadores educacionales en técnicas modernas de estadística experimental permitiría elevar la calidad de la experimentación pedagógica”* (pág. 14).

Centrarse en el tema de la validez es importante porque el conocimiento de sus tipologías puede ayudar a comprender mejor las ventajas y desventajas de cada tipo de diseño a la hora de aplicarlo a una cuestión determinada (Mark, 1986; Vara, 2006). Ahora, hablar de validez de una evaluación de impacto es otro tema complejo y muchas veces, por desgracia, la literatura la trata de una forma confusa y heterogénea (Aliaga, 2000). A modo de ejemplo:

- Buendía, Colás & Hernández Pina (1997), cuando se refieren a la validez de la investigación experimental, distinguen dos grandes tipos, la validez interna y la validez externa, dentro de la cual engloban la validez de población, la validez ecológica, y la validez de constructo.
- Por su parte Latorre, Del Rincón & Arnal (1996) plantean que *“hay que tener en cuenta tres tipos de validez: la interna, que exige el*

control; la externa (generalización) y la conceptual (coherencia)” (p. 127).

- Colás & Buendía (1994) señalan que los diseños experimentales deben responder a tres clases de validez: validez interna, validez externa, y validez ecológica.
- Bisquerra (1989) distingue por un lado entre validez interna y validez externa, añadiendo un apartado residual de "otros tipos de validez", en los que incluye la validez ecológica, la validez de población, la validez de conclusión estadística y la validez de constructo.
- López-Barajas, López López & Pérez Juste (1987) distinguen también entre validez interna y validez externa, dentro de la cual distinguen entre validez de muestras, de variables y ecológica.
- Pérez Juste (1985), en su revisión de los diseños experimentales, hace referencia sólo a la validez interna y a la validez externa, siguiendo el primer modelo planteado por Campbell y Stanley (1966).

Así, existe una enorme variedad de planteamientos sobre validez, generando mayor confusión en los lectores. Sin embargo, es importante resaltar que casi todas estas propuestas están basadas en el modelo original de Campbell y Stanley (1966), por lo que su estudio se torna obligatorio, principalmente desde una perspectiva histórica. Por ello, a continuación se analizará brevemente la evolución en el campo de la validez e inferencia causal.

4.2.1.2. Modelos de la validez causal

En términos funcionales, una relación causal existe cuando hay tres condiciones: a) precedencia temporal (las causas han de acontecer antes que las consecuencias), b) covariación de causas y efectos (los cambios de una vayan acompañados de cambios en la otra)⁵⁵ y c) se eliminen otras explicaciones alternativas a las relaciones causa-efecto.

Estas tres condiciones, necesarias pero no suficientes, constituyen una definición operativa de la causación. Sin embargo, la última condición es la más difícil de verificar, por cuanto exige control de variables alternativas (extrañas). Al respecto, la obra original de Campbell y Stanley (1966), fue la primera que ofreció un marco general para guiar la consecución del cumplimiento de esta tercera condición. Por ello, su estudio se torna obligatorio para cualquier gerente social. De lo dicho, a continuación se revisan los principales modelos de causación, desde el modelo original de Campbell y Stanley (1966) hasta sus modificaciones y modelos alternativos (Cronbach, 1982).

⁵⁵ Esta condición es necesaria, pero no suficiente, por lo que las conclusiones causales de los diseños correlacionales, por sí solos, no son válidas.

4.2.1.2.1. Modelo original de Campbell y Stanley (1966)

La obra original de Campbell y Stanley (1966) supuso un impulso notable a la investigación causal, quizás la más importante desde las aportaciones de Fisher (creador de los diseños experimentales estadísticos) en la década de 1920. Antes de Campbell, se asumía de modo general que la mera asignación aleatoria de los casos a las distintas situaciones experimentales (según proponía el modelo fisheriano del análisis de varianza) permitía controlar los errores de la investigación y llegar a conclusiones "sólidas". Por tanto, se suponía que el único control posible a la hora de diseñar una investigación era el aleatorio (Campbell, 1986).

Sin embargo, Campbell y Stanley identificaron dos grandes grupos de sesgos posibles (Campbell, 1986), algunos de los cuales podían ser controlados mediante la asignación aleatoria, denominados *Amenazas a la validez Interna* mientras que otro grupo de amenazas o sesgos no quedaban controladas aunque se utilizara este procedimiento azaroso, las *Amenazas a la Validez Externa*⁵⁶. Así, Campbell y Stanley diferencian entre amenazas de validez interna y de validez externa, recomendando que en cualquier investigación se deberían separar ambos conjuntos de amenazas, de tal modo que se pueda identificar si cada uno de ellos está controlado o no y, por tanto, precisar si la investigación es válida o no.

Además de este aporte conceptual, el mayor crédito de Campbell y Stanley fue la creación de los "diseños cuasi-experimentales", para referirse a aquellas investigaciones en las que, por diversas razones, las amenazas a la validez interna no se pueden controlar mediante la aleatorización (Coyle, Boruch & Turner, 1991). Los diseños cuasi-experimentales son entendidos como una aproximación imperfecta (por lo que hace referencia a su validez interna) a los diseños experimentales, pero ambos diseños, experimentales y cuasi experimentales, carecen, en principio, de validez externa, que habría que planificar separadamente.

La *Validez Interna* de una investigación causal representa el nivel de seguridad de que el programa aplicado provoca algún cambio significativo en la variable dependiente (Campbell, 1957; Campbell & Stanley, 1966). Dicho de otro modo, y basándonos tanto en el principio de falsación popperiano como en el tercer criterio de causación expuesto por Stuart Mill, puede entenderse a la *Validez Interna* como la inversa de la "*admisibilidad relativa de las teorías rivales*" (Campbell & Stanley, 1966, p. 74), de tal modo que cuanto menos probables sean las otras explicaciones alternativas, mayor será la validez interna del estudio. Así, de ningún modo, se puede llegar a alcanzar una validez interna plena, es decir, no se podría confirmar (o demostrar) absolutamente que un determinado programa es causa de los cambios registrados. Por ello, los procedimientos cuasi-experimentales han de prestar particular atención a la eliminación de las explicaciones alternativas. Así, precisamente porque se carece de control experimental total, "*...es imprescindible que el investigador tenga un conocimiento a fondo de cuales son las variables específicas que su diseño particular no controla*" (Campbell y Stanley, 1966, Pág. 71).

⁵⁶ Campbell (1986)... "*el fin principal que se persigue con la aleatorización (...) es la validez interna, no la externa*" p. 52.

Tabla N° 4.1. Tipos de Validez y sus amenazas según Campbell y Stanley (1966)

	Amenazas	Definición
Validez Interna	Historia	Acontecimiento específico entre el pretest y el post-test
	Maduración	Cambios debidos al mero paso del tiempo e independientes del tratamiento
	Administración de test	Efecto de la administración de un test sobre los resultados en otro posterior
	Instrumentación	Cambios en los criterios de evaluación o atribuibles a las características del instrumento
	Regresión estadística	Tendencia a aproximarse a la media que ocurre en los grupos seleccionados por sus valores extremos
	Selección	Diferencias entre los grupos provocadas por una asignación sesgada de los sujetos a los tratamientos
	Mortalidad experimental	Pérdida diferencial de participantes entre los grupos de comparación
	Selección	Interacción entre selección y otras amenazas
Validez Externa	Interacción entre las pruebas y el tratamiento	Cambio en la sensibilización de los sujetos experimentales provocado por el pretest
	Interacción entre selección y tratamiento	Se refiere a aquellos casos en los que los resultados observados pudieran ser específicos de la situación investigada
	Otras interacciones con el tratamiento	Referidas a los factores que podían afectar a la Validez Interna, tales como maduración, historia, mortalidad, etc.
	Efectos reactivos de los dispositivos experimentales	Cambios en la reacción de los sujetos como consecuencia de las condiciones experimentales
	Interferencia entre tratamientos múltiples	Efectos que pueden ejercer antiguos tratamientos aplicados al grupo experimental cuando se le aplican varias intervenciones

El segundo concepto fundamental que proponen Campbell y Stanley es el de validez externa, término que hace referencia a la posibilidad de generalización de los resultados obtenidos en un estudio determinado a otras condiciones (poblaciones, situaciones, variables de tratamiento y variables de medición). Puesto que esta generalización siempre parte de datos concretos, se basa en un proceso inductivo⁵⁷. Como indican Campbell y Stanley (1966), *"Desde el punto de vista lógico no podemos generalizar más allá de dichos límites (las características específicas de la investigación original); es decir, no podemos generalizar en absoluto"* (Pág. 39). Sin embargo, la generalización se ha basado, generalmente, en la experiencia y el sentido común, y en lo que los autores denominaron una "ley general empírica": *"...cuanto más cercanos se hallan dos acontecimientos en tiempo, espacio y valor (...) más tienden a ajustarse a las mismas leyes"* (Pág. 40). Téngase en cuenta que para resolver los problemas de validez externa no siempre es posible acudir a procedimientos probabilísticos, ya que los mismos sólo son aplicables a algunos de sus componentes (por ejemplo, a la muestra de casos), pero no a todos (por ejemplo, el instrumento de medida o el tiempo).

La concepción de validez externa planteada por Campbell y Stanley es menos homogénea que la de validez interna, lo que ha llevado a muchos autores a subdividirla en diversos componentes independientes entre sí. Así, la validez externa ha sido dividida en validez de población (representatividad de casos o sujetos) y validez ecológica (representatividad de la situación). Para Snow (1979), la validez externa es el resultado conjunto de la validez de población, de la validez

⁵⁷ Procedimiento denunciado desde hace siglos como particularmente poco fiable.

ecológica y de la "generalidad referente". Para Kratochwill (1978) los componentes de la validez externa serían la validez de población, la generalización de grupos a individuos y la validez ecológica. Por su parte, Cone (1977) y Wiggin (1966), consideran como validez externa a la generalización de las puntuaciones, de los elementos, temporal, de las situaciones, del método y de las dimensiones. Otros autores han hablado de validez de variables, generalización del instrumento de medida, validez de constructo, validez de tratamiento, etc. En fin, se han propuesto diversas subdivisiones, sin embargo, no se ha llegado a un consenso al respecto.

Para verificar la generalización de las conclusiones causales, Campbell y Stanley proponen dos vías. Por un lado la representatividad de las muestras y, por el otro, la replicación de los estudios. Otro aspecto de gran relevancia en el modelo de validez de Campbell y Stanley se refiere a la prioridad de la validez interna (considerada como una condición *sine qua non*) sobre la validez externa, ya que sin seguridad en la existencia de una determinada relación causal no tiene sentido considerar su posible generalización. Debe advertirse que esta concepción ha sido duramente criticada por Cronbach (1982).

4.2.1.2.2. Modelo de Cook y Campbell (1979)

Una década después, Cook & Campbell (1979) reformularon el modelo planteado por Campbell y Stanley (1966). En este caso, se hizo una clasificación más minuciosa de la validez causal. Cook & Campbell (1979) dividieron cada uno de los dos tipos de validez en otros dos nuevos subtipos, produciendo una nueva tipología de cuatro tipos de validez: a) validez de la conclusión estadística, b) validez interna, c) validez de constructo de las causas y efectos y d) validez externa. La siguiente figura esquematiza la evolución conceptual de estos tipos de validez.

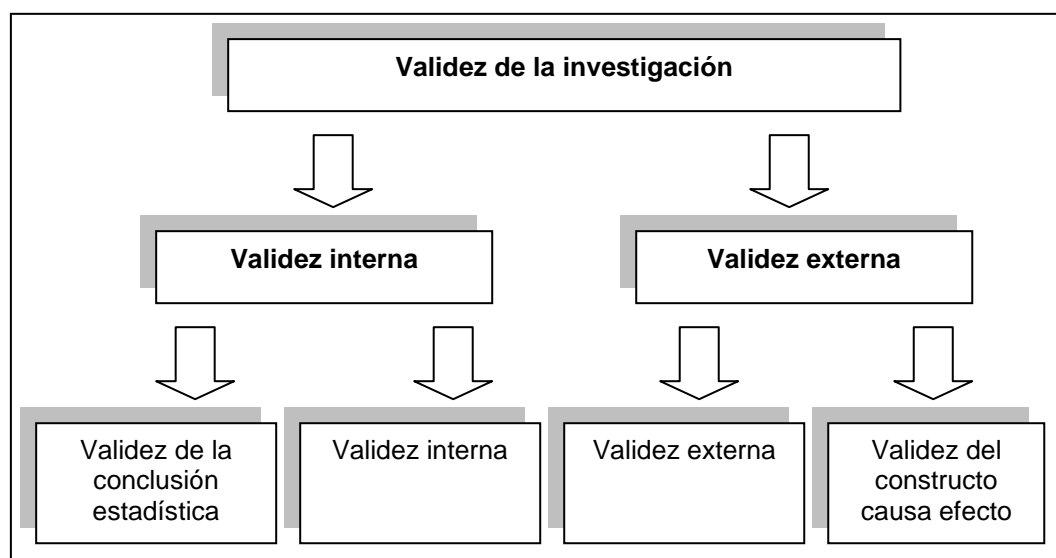


Figura N° 4.4. Evolución desde Campbell & Stanley (1966) a Cook & Campbell (1979) (Fuente: Basado en Aliaga, 2000).

Las amenazas específicas contra los tipos de validez son presentadas en la tabla siguiente. Posteriormente, se analizará cada uno de estos tipos de validez.

Tabla N° 4.2. Amenazas a la validez en Cook y Campbell, 1979. (Fuente: Anguera et al, 1995).

Tipo de validez	Amenazas contra la validez
Validez de la conclusión Estadística	Baja potencia estadística Violación de los supuestos de las pruebas estadísticas Violación del error tipo I Fiabilidad de la medición de la respuesta Fiabilidad de la administración de los tratamientos Irrelevancias aleatorias del contexto de investigación Heterogeneidad aleatoria de las unidades de respuesta
Validez Interna	Historia Maduración Administración de pruebas Instrumentación Regresión estadística Selección Mortalidad Selectiva Interacciones con la selección Ambigüedad en la dirección de la relación causa-efecto Difusión e imitación de tratamientos Igualación compensatoria de tratamientos Rivalidad compensatoria de los sujetos Desmoralización de los sujetos
Validez de constructo	Explicación pre operacional inadecuada. Sesgos derivados del uso de una operación única Sesgos derivados del uso de un solo método Adivinación de hipótesis Recelo de evaluación Expectativas del experimentador Confusión de constructos y niveles de constructo. Interacción de tratamientos intra sujeto Interacción de administración de pruebas y tratamiento Generalidad restringida entre constructor
Validez externa	Interacción selección-tratamiento (validez de población) Interacción contexto-tratamiento (validez ecológica) Interacción historia-tratamiento (validez histórica)

a) Validez de la conclusión estadística:

Puesto que la covariación es una condición necesaria para poder determinar una relación causal, el nivel de seguridad de que tal variación existe, es de considerable importancia, lo que justifica que este concepto se desligue del resto del concepto de "validez interna", tal y como había sido definido originalmente por Campbell y Stanley (1966).

De hecho, los problemas de la validez de la conclusión estadística se habían considerado previamente como amenazas a la validez interna. Sin embargo, la covariación no siempre implica relación causal y los procedimientos para determinar la existencia y magnitud de tal covariación, suelen ser estadísticos independientes del diseño. Por ello, Cook y Campbell (1979) decidieron añadir

esta categoría como diferenciada del resto del concepto de la validez interna, dándole un significado más específico. Así, las amenazas a la validez de la conclusión estadística están relacionadas fundamentalmente con las fases de medición y análisis de datos.

Tabla N° 4.3. Amenazas a la validez de la conclusión estadística según Cook & Campbell, 1979 (Fuente: Elaboración propia).

Amenazas	Descripción
Baja potencia estadística	El número muestral y el cálculo de la probabilidad son inadecuados, disminuyendo la potencia del contraste estadístico.
Violación de los supuestos de las pruebas estadísticas	Se aplican estadísticos sin verificar el cumplimiento de sus supuestos.
Violación del error tipo I	Concluir que existen efectos del programa cuando de hecho no existe.
Fiabilidad de la medición de la respuesta	Las medidas pretest y posttest no son fiables, no tienen consistencia interna, homogeneidad o estabilidad temporal.
Fiabilidad de la administración de los tratamientos	Los tratamientos no se aplican estandarizadamente. Se viola el principio de homogeneidad de situaciones.
Irrelevancias aleatorias del contexto de investigación	Diferencias aleatorias ambientales no controladas entre grupos, producidos posteriormente a la selección. Puede aumentar la varianza artificialmente en algún grupo.
Heterogeneidad aleatoria de las unidades de respuesta.	Diferencias aleatorias intrasujeto no controladas entre grupos, producidos posteriormente a la selección. Puede aumentar la varianza artificialmente en algún grupo.

b) Validez interna:

El concepto de validez interna en la reformulación de Cook y Campbell (1979) se centra, específicamente, en determinar si la relación entre variables (suponiendo que tal relación tenga una buena validez de conclusión estadística) es de tipo causal, y en tal caso, cuál es la dirección de dicha relación causal. Ha de tenerse presente que las conclusiones sobre esta posible relación causal se establece exclusivamente, como ya mantenía el modelo anterior, entre las variables tal y como han sido operativizadas en la práctica de la investigación (Campbell y Stanley, 1966).

A las amenazas de validez interna ya conocidas en el modelo anterior, se suman algunas más, tales como la ambigüedad sobre la dirección de la relación causal, la difusión o imitación de tratamientos, la igualdad compensatoria de tratamientos (cuando dejar "sin nada" al grupo control parece inadecuado y se les concede algún bien o servicio), la rivalidad compensatoria de los sujetos (esto es, el efecto competencia que se puede desatar en el grupo control) o la desmoralización o resentimiento de los sujetos que reciben menor cantidad de tratamiento deseable⁵⁸.

⁵⁸ Curiosamente, algunas de estas nuevas amenazas introducidas por Cook, con la aprobación de Campbell, en el nuevo modelo - particularmente la desmoralización- no cumplen con la característica original que había definido los sesgos de la denominada validez interna, es decir, que no pueden ser controlados mediante la asignación aleatoria. Esto hace que posteriormente Campbell (1986) considere que alguna de estas amenazas (incluso puede que todas) no sean específicamente de la validez interna, sino de la validez de constructo.

Tabla N° 4.4. Amenazas a la validez interna según Cook & Campbell, 1979. (Fuente: Elaboración propia).

Amenazas	Descripción
Historia	Eventos específicos que ocurren al mismo tiempo que el tratamiento y que se asocian a los resultados. La ocurrencia de acontecimientos externos que afecten al comportamiento del individuo, es decir, cualquier tipo de hecho que le ocurra en su vida diaria podría afectar a la Variable dependiente (VD). Si esos acontecimientos ocurren, y el gerente social no los detecta, podría atribuir la mejoría al programa, cuando en realidad podría ser efecto de otras variables.
Maduración	Procesos internos que operan a corto plazo (hambre, fatiga, etc.) o largo plazo (envejecimiento) y que afectan los resultados.
Administración de pruebas	Efectos de administrar una prueba sobre las puntuaciones obtenidas en una administración posterior. Los datos obtenidos con una primera prueba pueden cambiar al repetir la aplicación de esa misma prueba, por efecto de la práctica o el recuerdo. Aunque no se produzca ninguna intervención, la repetición de las pruebas de evaluación puede mostrar cambios.
Instrumentación	Cambios en la calibración de un instrumento de medida (humano o mecánico). Se asocia a la fiabilidad. También se incluyen en este tipo de amenazas los errores de instrumentación, o errores en los propios procedimientos de evaluación. Un aparato no calibrado, con errores no sistemáticos, o procedimientos con poca fiabilidad, no validados, etc., impide llegar a conclusiones seguras sobre los resultados.
Regresión estadística	Tendencia hacia la centralidad producida en grupos con unidades directa o indirectamente seleccionadas sobre la base de sus puntuaciones extremas. Es un concepto matemático y se refiere al efecto de las puntuaciones extremas que tienden a "regresar a la media" de grupo. Cuando en la VD parecen existir efectos de techo o suelo, es decir, las puntuaciones son extremas y poco podrían aumentar o disminuir, respectivamente, entonces la tendencia en la siguiente medición es a obtener datos más cercanos a media, o al menos más estables, y no reflejarían el posible cambio de la Variable Independiente (VI).
Selección	Efectos producidos por la no equivalencia inicial de los grupos tratamiento y control. Se refiere a las diferencias sistemáticas en la comparación de grupos que pueden ser resultado de una selección sesgada, no completamente aleatoria, en la que persiste alguna diferencia específica entre los grupos que podría explicar también los resultados.
Mortalidad selectiva o diferencial	Reducción del tamaño y/o naturaleza original de los grupos durante el proceso de medición. Es la pérdida de sujetos a lo largo de una experimentación con grupos, que hace cambiar la composición o características de los grupos comparados.
Interacciones con selección	Interacción de la selección con otras fuentes de sesgo anteriormente mencionadas produciendo efectos espurios. Las más comunes son la selección-maduración, que ocurre cuando grupos no equivalentes maduran a diferente ritmo; la selección por historia cuando los acontecimientos ocurridos en los dos grupos son diferentes.
Ambigüedad en la dirección de la relación causa-efecto	Efecto producido cuando la presencia temporal de la causa no es clara. Suele aparecer en diseños no experimentales de tipo correlacional.
Difusión e imitación de tratamientos	Las condiciones del grupo experimental son extendidas al grupo control por acción directa de los sujetos, por imitación o por solidaridad. Es posible que ocurra intercambio de información sobre los procedimientos, condiciones u objetivos de una investigación a través de sujetos en varios grupos. La validez dejaría bastante que desear pues no habría independencia entre los grupos a comparar.
Rivalidad compensatoria de los sujetos	Los sujetos del grupo control se enteran de las diferencias de trato con el grupo tratamiento e ingresan en una rivalidad que compensan las diferencias, los sabotean o generan obstáculos posteriores.
Desmoralización de los sujetos	Los sujetos desertan por razones de motivación personal. Cuando los sujetos conocen su pertenencia a un determinado grupo, que recibirá tratamientos con resultados menos deseables, esta información puede actuar para cambiar su comportamiento. Por lo que los sujetos podrían actuar por "rivalidad" o "dejadez" si conocen los resultados esperados de ese experimento. Estos efectos de reactividad impedirían conocer el efecto específico de un determinado tratamiento, independiente de esa reactividad.

c) Validez de constructo de las causas y los efectos:

Cuando Campbell y Stanley (1966) se refieren a la validez externa afirman que ésta se ocupa de la posibilidad de generalizar a "otras poblaciones, situaciones, variables de tratamiento y variables de medición" (p. 16). En la reformulación planteada por Cook y Campbell (1979) las dos últimas generalizaciones se independizan y toman entidad propia. En efecto, ya Campbell & Stanley habían insistido repetidamente en que las relaciones causales encontradas en una investigación determinada se limitaban, exclusivamente, a las operacionalizaciones específicas de la variable dependiente y la variable independiente "en este caso experimental concreto" (Campbell y Stanley, 1966, Pág. 16). Por tanto, la generalización desde esas operacionalizaciones concretas a otras formas de entender (operacionalizar) dichas variables era un proceso que no debía asumirse sin más de modo automático. Debe resaltarse que mientras que la:

"validez interna tiene que ver con la eliminación de explicaciones alternativas de la supuesta relación causal entre X tal y como es manipulada y O tal y como es medida, la Validez de Constructo se ocupa de eliminar las explicaciones alternativas de las supuestas categorías de X y O" (Cook, Campbell & Peracchio, 1990, Pág. 495).

La validez de constructo está íntimamente relacionada con el concepto de "confundido", que hace referencia a la circunstancia de que un mismo hecho o una misma operacionalización, puede ser entendida como perteneciente a dos categorías o constructos teóricos diferentes, de tal modo que las variables que realmente mantienen una relación causal pueden estar enmascaradas o confundidas con otras que no tienen ese tipo de relación, pero que son coherentes con las operacionalizaciones utilizadas en la investigación.

Shadish, Cook & Houts (1986) informan, por ejemplo, de los resultados obtenidos por Coleman (1972) en su análisis sobre la igualdad de oportunidades educativas. Dependiendo del modo en que se hubiera operacionalizado en cada investigación la variable "calidad escolar" (como tamaño de la clase, gastos por alumno, antigüedad de las instalaciones y el material, aislamiento acústico, etc.) encontraba resultados diferentes de dicha variable. Otros ejemplos del efecto "confundido" son el efecto Hawthorne, el efecto Pigmalión o el efecto placebo (Ball, 1997).

Para solventar este problema de generalización desde las causas o las consecuencias tal y como se ha medido y/o manipulado a los conceptos, Cook & Campbell (1979) proponen un doble proceso, basado en la matriz multirrasgo-multimétodo de Campbell & Fiske (1959):

- a) Comprobar la convergencia a través de diferentes medidas o manipulaciones del mismo constructo (Cook & Campbell, 1979: 61).
- b) Comprobar la divergencia respecto de los resultados obtenidos con medidas o manipulaciones de constructo parecidos pero distintos.

La validez de constructo tiene un amplio listado de amenazas identificadas (ver tabla anterior) que Ato (en Anguera et al., 1995) ha dividido en dos grandes grupos⁵⁹. En la siguiente tabla se resume estas amenazas.

Tabla N° 4.5. Amenazas a la validez de constructo según Cook & Campbell, 1979 (Fuente: Elaboración propia).

	Amenazas	Descripción
Sesgos de operacionalización de las causas y efectos	Explicación preoperacional inadecuada	Puede estar presente cuando la elección de la definición operativa de una variable no se realiza a partir de un cuidadoso análisis conceptual de la misma, y por tanto no se adecua exactamente a la definición constitutiva (Kerlinger, 1975) del constructo.
	Sesgos por el empleo de una única operación	Puesto que definir operativamente un constructo o variable puede ser inadecuado (no tener validez de contenido o ser incompleta, estar sesgada, ser incorrecta) siempre es preferible utilizar diversos procedimientos, lo que nos permitirá triangular los datos y captar mejor el significado de la variable. Esta es una amenaza derivada de no comprobar la convergencia a la que hacíamos referencia anteriormente.
	Sesgos derivados de utilizar un solo método	Esta amenaza puede presentarse cuando, independientemente de que se operacionalice de una o varias formas, todas ellas utilizan un mismo procedimiento (sólo técnicas de lápiz y papel o sólo procedimientos para evaluar la reactividad fisiológica). En tales casos, el propio método puede conllevar un sesgo sistemático que nos impida aprehender adecuadamente el constructo que nos interesa.
	Confusión entre constructos y niveles de constructo	Ocurre en aquellos casos en que un determinado constructo sólo se relaciona con otro cuando alcanza un determinado nivel, pero el investigador no es capaz de descubrirlo porque no llega a sobrepasar dicho umbral. Un ejemplo simplificado podría ser el de un presunto investigador que concluyese que la temperatura no afecta al estado físico del agua, porque lo ha modificado desde los 10° centígrados hasta los 90° sin obtener resultados. En el mismo sentido, puede existir problemas para generalizar los efectos de la utilización de un sistema de educación compensatoria (o la ausencia de los mismos) si éstos consistían exclusivamente en una hora semanal de apoyo, cuando hubiera sido necesaria una mayor intensidad del tratamiento para conseguir algún resultado apreciable.
	Interacción entre diferentes tratamientos	Se refiere a la posibilidad de que los efectos de un determinado tratamiento aplicado con anterioridad al grupo experimental interfieran con los del tratamiento actual, por lo que malinterpretamos los efectos producidos por este último, que no podrán generalizarse a otras condiciones.
	Interacción entre la administración de pruebas y el tratamiento	Se refiere a los sesgos que la aplicación de un pretest puede provocar en los sujetos (sensibilización), consiguiendo efectos diferentes a los que se conseguirían en una situación natural, diferente de la investigación en que carece de esos controles previos.
	Generalizabilidad a través de los constructos	Se refiere a los posibles efectos, no considerados en el diseño previo, que un tratamiento puede tener entre las variables o constructos distintos a la variable dependiente.
Sesgos de reactividad de las unidades de respuestas (sujetos)	Adivinación de la hipótesis	Los sujetos que participan en una investigación con frecuencia pretenden adivinar qué es lo que el investigador pretende de ellos, y en un proceso que la psicología social ha denominado deseabilidad social, adaptarse a dichas expectativas, más que a responder

⁵⁹ Ato sólo clasifica siete de las diez amenazas citadas por Cook y Campbell en 1989.

	Amenazas	Descripción
o casos)		directamente al tratamiento que, realmente, se le administra.
	Aprensión a la evaluación (o efectos reactivos de la observación)	Indica la propensión de los sujetos a cambiar su comportamiento cuando saben que están siendo evaluados u observados (es la aplicación blanda en las Ciencias Humanas del Principio de Incertidumbre de Heisenberg en la Física).
	Expectativas del experimentador	También conocido con el nombre de efecto Rosenthal (por su descubridor), efecto Pigmalión o, en sentido inverso, efecto Golem (Ball, 1997). Aunque su auténtico alcance ha sido puesto en duda, está relacionado con los sesgos que de modo más o menos inconsciente puede intentar auto-cumplir sus predicciones.

d) Validez externa:

En la reformulación de Cook & Campbell (1979) la validez externa está referida a la generalización de los resultados a través de diversos componentes, básicamente los sujetos y las situaciones, a los que se ha añadido la dimensión temporal. Las amenazas de validez externa son:

- **Interacción de selección y tratamiento:** Posible limitación de la generalización de resultados a sujetos o poblaciones no directamente incluidas en la investigación, es decir, mantienen presente la posibilidad de que las relaciones causales identificadas sean específicas sólo para los sujetos directamente analizados.
- **Interacción de situaciones y tratamiento:** En este caso las limitaciones a la generalización de los resultados de la investigación puede venir de la especificidad de la relación causal en la situación específicamente considerada.
- **Interacción de historia y tratamiento:** En este caso se considera que las limitaciones a la generalización de las relaciones causales pueden estar directamente vinculadas con el momento temporal, de modo que suponer que tal relación causal existió en el pasado o se mantendrá en el futuro, al ir más allá de lo expresamente descubierto, tiene una alta dosis de suposición no contrastada.

Tabla N° 4.6. Amenazas a la validez externa según Cook & Campbell de 1979. (Fuente: Elaboración propia).

Amenazas	Descripción
Interacción selección-tratamiento	Se conoce también por validez de población. Se refiere a la generalización de la muestra a la población. Supone que haya algunos efectos de la selección de los grupos, de forma que impida la generalización a otros grupos en que no haya ocurrido esa selección inicial, los sujetos de uno de los grupos podría mostrar efectos de reacción diferenciales a la manipulación de la VI.
Interacción contexto-tratamiento	Corresponde a la denominada validez ecológica. Se refiere a la generalización del contexto de aplicación a otros contextos. Los sujetos suelen cambiar su comportamiento cuando se encuentran ante una situación estructurada, que exige un rendimiento determinado, con una gran directividad; frente a situaciones más naturales, donde no hay altas exigencias de la situación. Se han definido distintos tipos de reactividad a la situación experimental, por un lado, la reactividad observacional, que se produce cuando el sujeto cambia su comportamiento por el hecho de ser

Amenazas	Descripción
	actor de una investigación y tiende a reaccionar más a las características generales de la investigación que a las variables independientes efectivamente manipuladas. Por otro, el efecto de predisposición o sensibilización del sujeto a tratamientos posteriores, es decir, si el sujeto no es novato en las tareas experimentales puede que reaccione de forma diferente a otro que sí lo sea, y ello dificulte la generalización. Y, además, el denominado efecto de novedad, la reacción del sujeto a cambiar inmediatamente tras la introducción del tratamiento, sea cual sea éste, lo que impediría la generalización de esos resultados a largo plazo.
Interacción historia-tratamiento	La inferencia se ve limitada por acontecimientos específicos ocurridos en alguno de los grupos y que generan diferencias no controladas.
Interacción administración de pruebas-tratamiento	Generalización del resultado a otras condiciones de instrumentación diferentes de la utilizada.
Interacción instrumentación-tratamiento	Generalización del resultado a otras condiciones de instrumentación diferentes a la utilizada.
Interacción de tratamientos intrasujeto	Generalización del resultado de la situación en que los sujetos reciben más de un tratamiento a la situación (hipotética) en que sólo reciben uno.

Son varios los procedimientos que Cook & Campbell (1979) analizan para procurar aumentar la validez externa de una investigación. En primer lugar se consideran diversos tipos de muestreo, que incumben tanto a sujetos como a situaciones, pero que no considera, obviamente, la dimensión temporal:

- a) **Muestreo aleatorio:** este muestreo aleatorio es el que mejor procura la representatividad de los resultados, y por tanto, el que nos ofrece mayores garantías para generalizar los resultados a otras situaciones o sujetos pertenecientes a la población global. Sin embargo, plantea serios problemas prácticos, fundamentalmente debido a que las circunstancias no siempre permiten hacer un muestreo de este tipo, o a que cuando se realiza, las muestras suelen estar muy dispersas.
- b) **Muestreo heterogéneo deliberado:** consiste en incluir en el estudio un amplio rango de sujetos y circunstancias de tal modo que abarquen la mayor amplitud posible. Esta idea no debe entenderse como selección exclusiva de los casos extremos, sino que han de incluirse también casos "típicos". Este procedimiento es mucho más asequible que el aleatorio, y aunque técnicamente no se puede generalizar a partir de sus resultados, la amplia gama considerada nos permite hacernos una idea global de las relaciones estudiadas en las diversas modalidades posibles.
- c) **Muestreo impresionista de categorías o modos:** Consiste en determinar las clases o categorías a las que queremos generalizar y elegir elementos de cada una de esas modalidades, a fin de obtener una impresión general. Este tipo de procedimiento es útil en aquellos casos en los que las pretensiones de generalización son limitadas, y particularmente en investigación aplicada, ya que no es útil para establecer un conocimiento teórico ampliamente generalizable. Es el procedimiento menos poderoso de los tres tipos de muestreo, pero tiene como ventajas principales su simplicidad y economía.

Hay, sin embargo, un último procedimiento para asegurar la generalización de los resultados: **la replicación**. Mediante este método se consigue establecer si los datos avalan la generalización de los resultados cuando algunas de las condiciones (sujetos, situaciones o tiempo) es modificada. Por tanto, es un procedimiento con respaldo empírico, en lugar de acudir a la justificación teórica o racional más propia de los tipos de muestreo.

4.2.1.2.3. Modelos alternativos: UTOS de Cronbach

Aunque los modelos más influyentes sobre validez causal han sido los creados por Campbell & colaboradores, existen otros modelos alternativos. De ellos el que ha tenido un mayor eco en la comunidad científica ha sido el de Cronbach, un modelo de notable complejidad de análisis. Otras aportaciones interesantes son las de Krathwohl (1985), Kruglanski & Kroy (1976) o el de Reichardt (1986, citado en Mark, 1986), que es una variación del modelo de Cronbach (Mark, 1986).

El modelo de Cronbach (1982), es probablemente el modelo sobre validez más elaborado y de mayor interés entre los que se han construido como alternativa a los de Campbell & colaboradores (Campbell & Stanley, 1966; Cook & Campbell, 1979). Cronbach (1982) distingue cuatro tipos de elementos diferentes en una investigación:

- **Unidades:** Se refiere a los sujetos experimentales, ya sean estos individuos, grupos, centros escolares, etc.
- **Tratamiento:** Es la presunta causa, es decir, aquello que manipula (o selecciona) el investigador.
- **Observación:** Se refiere al presunto efecto en la relación causal, que no se manipula sino que únicamente se mide.
- **Situación:** Se refiere al "*contexto social, entendido en un sentido amplio, en el que el estudio tiene lugar*" (Cronbach, 1982, p. 79). Por tanto, incluye desde el clima social, económico, ideológico o político hasta el momento temporal específico⁶⁰.

Cronbach utiliza las iniciales de estos elementos de distintas maneras según el grado de especificidad. Así, por ejemplo, utiliza el término UTOS, en mayúsculas, para referirse a los constructos que son objeto de investigación, bien entendido que en un nivel abstracto al añadir el momento temporal en uno de los elementos relevantes de la investigación. Cook, Campbell & Peracchio (1990) convierte esas iniciales en UTOSCH, al representar el tiempo como CH, del griego

⁶⁰ Algunos autores (por ejemplo, Cook, Campbell y Peracchio, 1990, Ato en Anguera et al., 1995) han dividido este concepto en dos diferentes, uno sobre las condiciones en las que se realiza propiamente la investigación (el contexto o situación) y otro referido específicamente al momento temporal, en el sentido dado por Cook y Campbell (1979) a este concepto, aspecto que comparto, ya que puede darse el caso de que una misma investigación se realice en un mismo contexto (por ejemplo, en un mismo colegio público entre alumnos de la misma asignatura) pero que los resultados sean distintos debido a los cambios más intangibles que puedan deberse al paso del tiempo (cambios en la ideología, la legislación, etc.).

“*chornos*”. Manuel Ato, (Anguera et al., 1995) ha traducido ingeniosamente los términos al castellano dando lugar al acrónimo TRUCO -Tratamiento, Respuesta, Unidades, Contexto y Ocasión o momento temporal.

Cook & Campbell (1979) resaltaron las dificultades de interpretación que pueden presentarse al operativizar, en la práctica, los distintos constructos. Cronbach (1982) profundiza en esta idea y utiliza el término UTOS para referirse a los elementos concretos incluidos en la investigación. Así, la **u** hace referencia a la muestra de unidades (generalmente se refiere a los sujetos concretos) que son directamente analizados en la investigación; la **t** se refiere al modo específico en que se ha operacionalizado la variable tratamiento y la **o** a la forma en que se ha medido la variable dependiente. Cronbach (1982) siempre utiliza la **S** de la situación en mayúsculas ya que, atendiendo a la clasificación fisheriana de los términos de un análisis de varianza, la situación (y habría que incluir también el tiempo) es un factor fijo, en el sentido de que sólo hay uno en cada estudio, mientras que los demás términos pueden ser factores aleatorios.

En un tercer nivel de especificidad, Cronbach utiliza los *UTOS para referirse a aquellas condiciones donde se puede generalizar los resultados obtenidos al investigar los utoS específicos de una investigación, pero que difieren de las poblaciones de referencia (los UTOS) en alguna o algunas características (por ejemplo, otras poblaciones, contextos diferentes, efectos distintos, o tratamientos parecidos).

Puesto que existen tres niveles diferentes de concreción-generalización, Cronbach (1982) plantea dos tipos deferentes de inferencia, según las relaciones que se desee establecer a partir de las posibles relaciones causales identificadas en una investigación concreta. En primer lugar, el establecimiento de relaciones que puedan existir entre los utos y los UTOS (es decir, las operaciones o muestras con los constructos o poblaciones que pretenden representar) son denominadas por Cronbach inferencias internas, y la seguridad que se puede tener en tal relación es denominada validez interna. En segundo lugar, la relación entre los utoS y los *UTOS (o sea, entre los casos específicos y los constructos o poblaciones ligeramente diferentes a aquellos que pretendían representar) recibe el nombre de inferencias externas, y el grado de credibilidad o seguridad en tal relación es denominada validez externa.

Cronbach (1982) concede una importancia prioritaria a estos dos tipos de inferencias, particularmente en las inferencias externas, ya que como señala *"normalmente sólo una pequeña fracción de la audiencia está interesada en los UTOS que define el estudio"* (p. 99). En efecto, para ser aplicado, los lectores de la investigación normalmente habrán de preguntarse cosas como *¿funcionará este programa de lecto-escritura en nuestro idioma (*S), con nuestros alumnos (*U), tal y como nosotros evaluamos el rendimiento (*O) y adaptado para que lo apliquen nuestros profesores (*T)?* Es decir, las aplicaciones de los resultados han de hacerse con frecuencia a condiciones diferentes de las estudiadas y, en el caso de la investigación aplicada, este proceso es el prioritario, ya que los responsables de tales tomas de decisión no suelen ser tan minuciosos respecto a la seguridad con que se haya establecido la relación causal, sino que están más bien interesados en la aplicabilidad de los descubrimientos.

Cronbach utiliza algunos conceptos acuñados por Campbell & sus colaboradores (fundamentalmente el término de validez interna) en un sentido distinto, lo que ha producido cierta confusión, que añadido a la complejidad del modelo, ha propiciado su poca difusión. Otra de las críticas que se han realizado contra este modelo se refiere al excesivo esquematismo dado al tema de la generalización (Mark, 1986). Es bien cierto que Cronbach tiene razón al plantear que las aplicaciones prácticas de las investigaciones con frecuencia requieren su generalización a situaciones o condiciones que varían, en mayor o menor grado, de las poblaciones o condiciones analizadas inicialmente. Sin embargo, no hay razón alguna para mantener que esa mayor diversidad deba plantearse en sólo dos niveles (UTOS y *UTOS), sino que, más bien, debe ser concebida como un continuo (Mark, 1986).

4.2.1.2.4. Evolución posterior de las tipologías de validez

A pesar de la complejidad del modelo de Cronbach (1982) y de su limitada difusión, su contribución a la aclaración del tema fue importante, básicamente porque obligó a reflexionar sobre los modelos existentes y a introducir algunos cambios que permitiesen aclarar el campo.

Entre las reacciones más importantes que produjo el modelo de Cronbach (1982), aunque hubo algunas otras razones que justificaron el cambio (Cook & Shadish 1994), cabe citar la reformulación que hizo Campbell (1986) de algunos de los conceptos incluidos en sus propios modelos. Así, como una reacción al hecho de que muchos autores solían considerar, erróneamente, (a pesar de lo indicado por él de modo explícito) que el término de validez interna se refería al nivel de seguridad que podíamos tener en que exista una relación causal entre dos conceptos, Campbell propone cambiar el nombre de validez interna por otro que especifique, de modo más detallado, las características que tiene dicho concepto. La alternativa que ofrece es la de pasar a denominarla “validez Causal local molar”. Con esta nueva denominación pretende hacer hincapié en algunos aspectos esenciales, representados por cada uno de los términos empleados:

- **Causal:** se refiere a que este tipo de validez ha de centrarse fundamentalmente en que las relaciones analizadas sean realmente de tipo causal.
- **Local:** hace referencia exclusivamente a que el nivel de relación causal se establece exclusivamente entre la t y la o del modelo de Cronbach (la t y la r en el modelo traducido del esquema truco). Ya en Campbell & Stanley (1966, p. 16), se referían a la validez interna como los cambios que introducían *"los tratamientos empíricos en este caso experimental concreto"*. Existe en esta nueva conceptualización, por tanto, un rechazo implícito a la postura de Cronbach (1982) que no sólo incluye en el concepto de validez interna los constructos T y O, sino incluso los de U y S.
- **Molar:** implica que las relaciones causales identificadas en el nivel más bajo de operacionalización han de entenderse como el resultado de la interacción global entre todos los elementos implicados

(sujetos, contexto, forma de manipular el tratamiento, modo de observar los resultados, etc.) más que como el efecto de alguno o algunos de dichos microelementos aisladamente. Dicho de otro modo, lo que se determina no es si esta causa provoca esta consecuencia, sino si tal relación acontece para estos sujetos determinados en estas circunstancias concretas y en este preciso momento temporal.

Campbell (1986) no sólo propuso cambiar el nombre de la validez interna, sino que su afán clarificador de conceptos le hizo proponer igualmente un cambio de denominación para la validez externa. Como respuesta a los dos niveles de inferencia (interna y externa) planteados por Cronbach (1982), Campbell propone modificar el concepto de validez externa y sustituirlo por el de “gradiente de semejanza” (*Proximal Similarity*), como parte de un proceso de reflexión que le lleva a alejarse cada vez más de los procedimientos basados en procurar la representatividad a través del muestreo. Por un lado debido a que tal tipo de muestreo es impracticable para algunos componentes de la investigación, como el tiempo. Por otro porque esa misma dimensión temporal, aparte de otras, afecta a la representatividad del resto de componentes: ¿un muestreo aleatorio de sujetos - o de escuelas, o de aulas- puede seguir considerándose representativo de cómo es la población de referencia diez o quince años después de realizado el estudio? Por último, razones de tipo práctico (económicas, geográficas, etc.) recomiendan la utilización de muestreos ilustrativos (impresionistas, tal y como los denominaban Cook y Campbell, 1979) orientados por alguna teoría específica en lugar de los muestreos aleatorios de poblaciones amplias, cuyo abandono formalmente recomienda Campbell (1986). Mark (1986) y, de modo especial, Shadish, Cook y Houts (1986) abogan por una planteamiento semejante, en el que las inferencias no se construyan a través de procedimientos de estadística inferencial sino a través de la integración de la información proveniente de diversas fuentes.

Por tanto, y como resumen, Campbell reformula en 1986 su tipología de la validez de la investigación causal de tal modo que decanta su antigua Validez Interna (ahora denominada Validez Causal Local Molar) hacia una concepción más empírica y menos teórica, mientras que impulsa su antigua Validez Externa (ahora Gradiente de Semejanza) hacia un modelo más guiado por la teoría específica de cada estudio y menos por procedimientos aleatorios más o menos automatizados.

Otros planteamientos han puesto de manifiesto algunos problemas con los que se enfrenta la investigación causal. Por ejemplo, Mark (1986) en su excelente trabajo integrador de las diversas tipologías de validez, señala que el intentar asegurar una alta validez en el establecimiento de las relaciones causales puede carecer de utilidad práctica, tal y como ya había señalado anteriormente la postura de Cronbach, mucho más interesado por la trascendencia práctica de las investigaciones que por el nivel de seguridad en las conclusiones.

Cook y Shadish (1994) señalan que las tipologías de validez no están (ni probablemente puedan llegar a estar) completas. Siguiendo a Dunn (1982) reconocen que la cantidad de tipos de validez es ilimitada: ¿Por qué separar la validez de constructo de la validez externa, como hacen Cook y Campbell en 1979,

y no hacer lo propio con la validez de población, o la ecológica o la temporal, por ejemplo? Cook, Campbell y Peracchio (1994) presentan un listado de amenazas a la validez de investigación que, aunque basada en las propuestas anteriores de Campbell y Cook (1979), que a su vez se basaban en la de Campbell y Stanley (1966), presenta ciertas modificaciones, fundamentalmente respecto a la categoría a la que se asigna cada tipo de amenaza, lo que nos sugiere es que éste tampoco es un campo completamente cerrado.

Tabla N° 4.7. Evolución de las clasificaciones de amenazas a la validez en los diversos modelos de Campbell & Cols (Basado en Aliaga, 2000).

Tipo de validez	Campbell & Stanley (1966)	Cook & Campbell (1979)	Cook, Campbell & Peracchio (1990)
Validez Interna	Historia Maduración Administración de tests Instrumentación Regresión Selección Mortalidad experimental Interacción entre selección y otras amenazas	Historia Maduración Administración de pruebas Instrumentación Regresión estadística Selección Mortalidad Interacciones con la selección Ambigüedad en la dirección de la relación causa-efecto Difusión o imitación de tratamientos Igualación compensatoria de tratamientos Rivalidad compensatoria de los sujetos Desmoralización de los sujetos	Historia Maduración Administración de pruebas Instrumentación Regresión estadística Selección Mortalidad Interacciones con la selección Ambigüedad en la dirección de la relación causa-efecto
Validez de Conclusión Estadística		Baja potencia estadística Violación de los supuestos de las pruebas estadísticas Violación del error tipo I Fiabilidad de la medición de la respuesta Fiabilidad de la administración de los tratamientos Irrelevancias aleatorias del contexto de investigación Heterogeneidad aleatoria de las unidades de respuesta	Baja potencia estadística Violación del error tipo I Fiabilidad de las medidas Fiabilidad de la administración de los tratamientos Irrelevancias aleatorias del contexto Heterogeneidad aleatoria de las unidades de respuesta
Validez Externa	Efecto reactivo o de interacción de los tests Interacción selección-tratamiento Efectos reactivos de dispositivos experimentales Interacción entre tratamientos múltiples	Interacción selección-tratamiento Interacción contexto-tratamiento Interacción historia-tratamiento	Interacción entre tratamientos múltiples Interacción medición-tratamiento Interacción selección-tratamiento Interacción contexto-tratamiento Interacción historia-tratamiento
Validez de Constructo de Causas y Efectos		Explicación preoperacional inadecuada. Sesgos derivados del uso de una operación única Sesgos derivados del uso de un solo método Adivinación de hipótesis Recelo de evaluación Expectativas del	Explicación preoperacional inadecuada. Sesgos derivados del uso de una operación única Sesgos derivados del uso de un solo método Interacción entre el procedimiento y el tratamiento Difusión o imitación de

Tipo de validez	Campbell & Stanley (1966)	Cook & Campbell (1979)	Cook, Campbell & Peracchio (1990)
		experimentador Confusión de constructos y niveles de constructo. Interacción de tratamientos intrasujeto Interacción de administración de pruebas y tratamiento Generalidad restringida entre constructos	tratamientos Igualación compensatoria de tratamientos Rivalidad compensatoria de los sujetos Desmoralización de los sujetos Adivinación de hipótesis en de las condiciones experimentales. Recelo de evaluación Expectativas del experimentador Confusión de constructos y niveles de constructo.

La utilidad de los modelos de validez ha sido indudable y ha permitido a los investigadores sociales centrarse en aspectos que pasaban desapercibidos. Sin embargo, hay una creciente complejidad y fragmentación de los modelos, así como la confusión con respecto al vocabulario empleado (con distinto significado según los autores, y con diversidad terminología para un mismo concepto). Adicionalmente, ya Cronbach ha señalado que muchas investigaciones se han visto innecesariamente complicadas al plagarse de controles para algunas amenazas. Reichardt (1985), por su parte, señala algunos ejemplos en los que algunas amenazas no se han detectado por acogerse demasiado rígidamente a los esquemas de validez planteados. Estos excesos no son, sin embargo, achacables a los modelos planteados, sino más bien a la utilización ciega (o excesivamente rígida) que se ha realizado de los mismos. Al respecto, Campbell y Stanley (1966) previenen sobre la utilización rígida de los modelos o tablas por ellos planteados:

"...más bien habría que estimular una actitud abierta hacia la indagación de los nuevos mecanismos de obtención de datos, y un nuevo análisis acerca de algunas de las imperfecciones que acompañan a la aplicación rutinaria de los (diseños) tradicionales (...) hemos llamado la atención sobre la posibilidad de utilizar creativamente las características peculiares de cualquier situación concreta de investigación al diseñar pruebas específicas de hipótesis causales" (pp. 138-139).

4.2.1.3. La validez causal en ARYSIS

El modelo ARYSIS utiliza las diversas modalidades de validez, preocupándose por determinar, previamente, cuál es el nivel de inferencia de la relación causal entre el programa y el impacto. Basado en una serie de aspectos algorítmicos, la evaluación de cada una de las 33 amenazas contra la validez no tiene otro objeto que contribuir a que el investigador (o, posteriormente, el evaluador o gerente social) descarte todas las fuentes de sesgo que pueden afectar

la investigación, y en el caso de no asegurar el descarte, utilizar los procedimientos estadísticos necesarios para que controlen las fuentes de sesgo sospechosas.

Tabla N° 4.8. Amenazas a la validez y tratamiento en ARYSIS (Fuente: Elaboración propia).

	Amenazas	Tratamiento en ARYSIS
Validez de la conclusión Estadística	Baja potencia estadística	Calcular adecuadamente el tamaño de la muestra. En su defecto, utilizar remuestreo con métodos Bootstrap. El análisis del poder estadístico es un modo útil de asegurar que la muestra propuesta sea lo suficientemente grande para los fines del análisis.
	Violación de los supuestos de las pruebas estadísticas	Utilizar el estadístico posteriormente al análisis del cumplimiento de sus supuestos.
	Violación del error tipo I	Fijar un valor alfa de acuerdo a la probabilidad de la presencia del evento.
	Fiabilidad de la medición de la respuesta	Realizar un estudio piloto para determinar la fiabilidad por homogeneidad, consistencia interna y estabilidad temporal de la medición.
	Fiabilidad de la administración de los tratamientos	Establecer un protocolo estandarizado de aplicación del tratamiento y supervisar su ejecución previa capacitación de los administradores.
	Irrelevancias aleatorias del contexto de investigación	Controlar con covariantes. Establecer una lista de registro de covariantes posibles que se registrarán en el monitoreo.
	Heterogeneidad aleatoria de las unidades de respuesta	Evitar la aquiescencia del entrevistado y enfatizar en la consigna consensuada.
Validez Interna	Historia	Realizar un pretest comprensivo. Controlar con covariantes. Establecer una lista de registro de covariantes ambientales posibles que se registrarán en el monitoreo.
	Maduración	Realizar mediciones de control en el tiempo. Controlar con covariantes. Establecer una lista de registro de covariantes intrasujeto posibles que se registrarán en el monitoreo.
	Administración de pruebas	Establecer un procedimiento estandarizado de administración de pruebas.
	Instrumentación	Utilizar pruebas de medida paralelas y con fiabilidad y validez probadas.
	Regresión estadística	Eliminar las puntuaciones extremas o controlar su efecto mediante ecuaciones de regresión.
	Selección	Utilizar procedimientos de emparejamiento "matching".
	Mortalidad Selectiva	Considerar un porcentaje adicional de "mortalidad selectiva" y determinar su efecto en los resultados.
	Interacciones con la selección	Utilizar procedimientos de selección paralelos.
	Ambigüedad en la dirección de la relación causa-efecto	Desarrollar un marco teórico previo. Utilizar el path análisis.
	Difusión e imitación de tratamientos	Evitar la cercanía socio-geográfica de los grupos control. Establecer estados contrafactuales utilizando criterios de propensity score.
	Igualación compensatoria de tratamientos	Evitar la cercanía socio-geográfica de los grupos control. Establecer estados contrafactuales utilizando criterios de propensity score.
	Rivalidad compensatoria de los sujetos	Evitar la cercanía socio-geográfica de los grupos control. Establecer estados contrafactuales utilizando criterios de propensity score.
	Desmoralización de los sujetos	Diseñar mecanismos de motivación incluidos en los programas. Evitar la cercanía socio-geográfica de los grupos control. Establecer estados contrafactuales

	Amenazas	Tratamiento en ARYSIS
		utilizando criterios de propensity score.
Validez de constructo	Explicación pre operacional inadecuada.	Diseñar el marco teórico contextualizado a la realidad y acorde al estado del arte. Utilizar estudios de evaluación rápida.
	Sesgos derivados del uso de una operación única.	Utilizar operaciones paralelas. Utilizar métodos mixtos.
	Sesgos derivados del uso de un solo método.	Utilizar el enfoque multimétodo y multirasgo.
	Adivinación de hipótesis.	Utilizar instrumentos (cuestionarios) paralelos. Controlar la aquiescencia.
	Recelo de evaluación	Establecer una consigna concertada.
	Expectativas del experimentador	Utilizar un evaluador externo. Incluir a los administradores como sujetos de estudio.
	Confusión de constructos y niveles de constructo.	Analizar previamente las dimensiones teóricas del constructo. Determinar la validez de contenido y constructo.
	Interacción de tratamientos intra sujeto	Aplicar cuestionarios de control de estímulos paralelos.
	Interacción de administración de pruebas y tratamiento	Entrenar cuidadosamente a los encuestadores y evitar que sean los mismo promotores del tratamiento.
	Generalidad restringida entre constructos	Determinar la validez de contexto. Analizar la validez de constructo de los indicadores.
Validez externa	Interacción selección-tratamiento (validez de población).	Garantizar la equivalencia de los grupos con técnicas de emparejamiento.
	Interacción contexto-tratamiento (validez ecológica)	Contextualizar la selección de indicadores para ponderar los impactos. Utilizar análisis de variables múltiples.
	Interacción historia-tratamiento (validez histórica)	Utilizar marcadores de experiencia previa en encuestas de control.

Estas alternativas algorítmicas se fundamentan procedimentalmente sobre la base de tres conceptos que son aceptados en la comunidad científica de evaluadores de impacto. Estos conceptos son los siguientes: a) estado de la naturaleza, b) potenciales resultados y c) decisión de participar.

1. El primer concepto, **“estado de la naturaleza”**⁶¹, se usa típicamente en la evaluación de programas sociales para describir dos estados posibles: el “estado de tratamiento” y “el estado de no-tratamiento”: El primero hace referencia a una realidad en que un individuo cualquiera ha participado efectivamente de un programa, es decir si ha sido “tratado” por el programa. El segundo estado hace referencia a una realidad en la que un individuo cualquiera no ha participado efectivamente de un programa, es decir que no ha sido “tratado” por el programa. Dentro de este concepto ingresa la preocupación por la validez de la conclusión estadística, el control del proceso de inferencia y otras amenazas contra la validez.
2. El segundo concepto, **“potenciales resultados”** que puede experimentar un individuo cualquiera durante su vida, está referido

⁶¹ El concepto de “estados de la naturaleza” es ampliamente utilizado en teoría de Finanzas. Así, a través del “precio” de cada uno los “estados de la naturaleza” es posible valorar cualquier activo financiero.

a distintas variables o aspectos de su vida y durante varios momentos en el tiempo que pudieran verse afectados por el programa. En la evaluación de impacto interesa concentrarse en aquellos aspectos en los cuales se espera que el programa tenga alguna implicancia. Así, por ejemplo, si se evalúa un programa de desayunos escolares posiblemente se definirá los “resultados” en función del estado nutricional y del rendimiento académico. Dentro de este concepto ingresa la preocupación por la validez de constructo, la fiabilidad de las medidas y la validez de contexto.

3. El tercer concepto se refiere a la “**decisión de participar**” en un programa social. La mayoría de los programas sociales suelen ser voluntarios en el sentido de que todos aquellos individuos que quieren participar (y cumplen con los requisitos para hacerlo) pueden potencialmente hacerlo (Ej. Comedores populares o programas de empleo juvenil). En algunos otros programas la decisión de participar en el programa no está en manos de los participantes sino de terceros (Ej. Construcción de caminos rurales o electrificación). Dentro de este concepto ingresa la preocupación por la validez de selección, el problema de los grupos control, las amenazas de la historia, maduración e interacción entre grupos, tratamientos y selección.

Estos conceptos previos, que resume en gran medida el aparato fundamental del Modelo ARYSIS, son claves para definir los procedimientos estandarizados de la evaluación de impacto. Tal como se desarrolló en el marco teórico (véase el punto 2.2.3.7), el procedimiento general de evaluación de impacto es el siguiente:

1. Determinar si realizar o no una evaluación.
2. Aclarar los objetivos de la evaluación.
3. Examinar la disponibilidad de datos.
4. Diseñar la evaluación.
5. Formar el equipo de evaluación.
6. Si se recopilan datos: diseñar y seleccionar muestras, elaborar instrumentos de recopilación de datos, reunir y capacitar personal para trabajo en terreno, realizar pruebas piloto, recopilar datos y administrar datos y acceder a ellos.
7. Recopilar datos actuales y analizar los datos.
8. Redactar los resultados y analizarlos con las autoridades responsables y otras partes interesadas.
9. Incorporar los resultados en el diseño del proyecto.

A este procedimiento general, ya presentado y explicado, conviene agregarle algunos aspectos –característicos de ARYSIS- con suficiente desarrollo. Con fines didácticos, se agruparán los nuevos procedimientos estandarizados propuestos en tres grupos: a) exploración inicial: contextualización, b) elaboración del diseño teórico propio, c) elaboración del diseño instrumental y estadístico.

En primer lugar, toda evaluación de impacto exige, en primer lugar, **contextualizar el procedimiento**. Puede resultar materialmente contraproducente aplicar el mismo procedimiento de evaluación de impacto –así sea el mismo programa- cuando se aplica en dos o más contextos socioculturales distintos (Youker, 2005; Bedi, Bhatti, Gine, Galasso, Goldstein & Legovini, 2006). La contextualización es fundamental por cuanto, si se deja de tomar en cuenta, afectaría significativamente la validez de los resultados y produciría sesgo en las comparaciones inter-muestra.

La contextualización es un criterio de aplicación mundial en las ciencias sociales. En psicología se usa para “adaptar” instrumentos de medición; tanto en la estructura de la prueba (para estandarizar su significativo) como en los baremos de calificación y comparación. En educación se usa el término de “diversificación”, entendiendo que los contenidos educativos deben adaptarse a la diversidad sociocultural del país.

La contextualización implica considerar la disponibilidad de los datos, la ponderación de la calificación de las mediciones, la determinación de covariantes y variables “no-observables” que pueden generar efecto confundido y, finalmente, la adaptación de los instrumentos de medida.

En cuanto a la elaboración del diseño teórico único, éste es de vital importancia por cuanto genera el marco para la comprensión de los constructos, la interpretación de los resultados y el control de la validez de las medidas. Recuérdese que el marco teórico es fundamental para la definición de los indicadores, para la sustentación de la relación causal entre variables y para la identificación de covariantes importantes que, si no se controlan, pueden afectar la validez de los resultados.

En cuanto a la elaboración del diseño estadístico apropiado, este se desarrollará con detenimiento en el subtítulo 4.3. Sin embargo, es importante dejar claro cuál es el procedimiento general que se usa para determinar los efectos de impacto y el procedimiento más común para disminuir los sesgos de selección.

4.2.1.3.1. Estimación de los efectos de impacto

En términos analíticos, la evaluación de impacto busca conocer como cambia el “flujo de resultados” de un individuo cualquiera cuando éste pasa de un “estado de no-tratamiento” a un “estado de tratamiento”. Sin embargo, tal como se ha mencionado con anterioridad, será inviable conocer ese cambio con certeza pues sólo uno de esos estados será el que se materializa, el otro no será nada más que un estado hipotético sobre el cual uno puede elucubrar “cómo hubiese sido”.

En términos formales el objetivo de la evaluación de impacto de un programa es determinar el cambio esperado en los flujos de resultados después de pasar por el programa que evidencia un beneficiario con determinadas características. Dado que al grupo de beneficiarios sólo se les observa en el estado de “tratamiento”, el principal problema consiste en estimar el estado contrafactual. Los distintos métodos de evaluación suelen resolver este problema utilizando un grupo de comparación o control, con individuos que no han participado en el programa y que poseen similares características. Así, dada la imposibilidad de

conocer el verdadero impacto de un programa para un individuo cualquiera, se tiene que recurrir a un “estimador” de ese efecto, necesitando un grupo de control.

Al respecto, existen tres estimadores ampliamente usados en la literatura sobre evaluación para medir el impacto de un programa: estimador antes-después, estimador de diferencia en diferencias (también conocido como doble diferencia) y estimador de corte transversal (Ñopo & Robles, 2002). Para un mejor entendimiento de estos tres estimadores se usa la figura siguiente. En ella se representa el “flujo de resultados” para el grupo de beneficiarios y el grupo de control en dos momentos del tiempo: antes del programa y después del programa (puede pensarse que ambos momentos abarcan uno o más períodos). Asimismo se representa el flujo de resultados en dos estados de la naturaleza (para el momento después del programa): estado de tratamiento y estado de no-tratamiento. La nomenclatura utilizada es la siguiente:

- A: “flujo de resultados” de los beneficiarios después del programa en estado de tratamiento.
- B: “flujo de resultados” de los beneficiarios después del programa en estado de no-tratamiento.
- C: “flujo de resultados” de los controles después del programa en estado de tratamiento.
- D: “flujo de resultados” de los controles después del programa en estado de no-tratamiento.
- E: “flujo de resultados” de los beneficiarios antes del programa.
- F: “flujo de resultados” de los controles antes del programa.

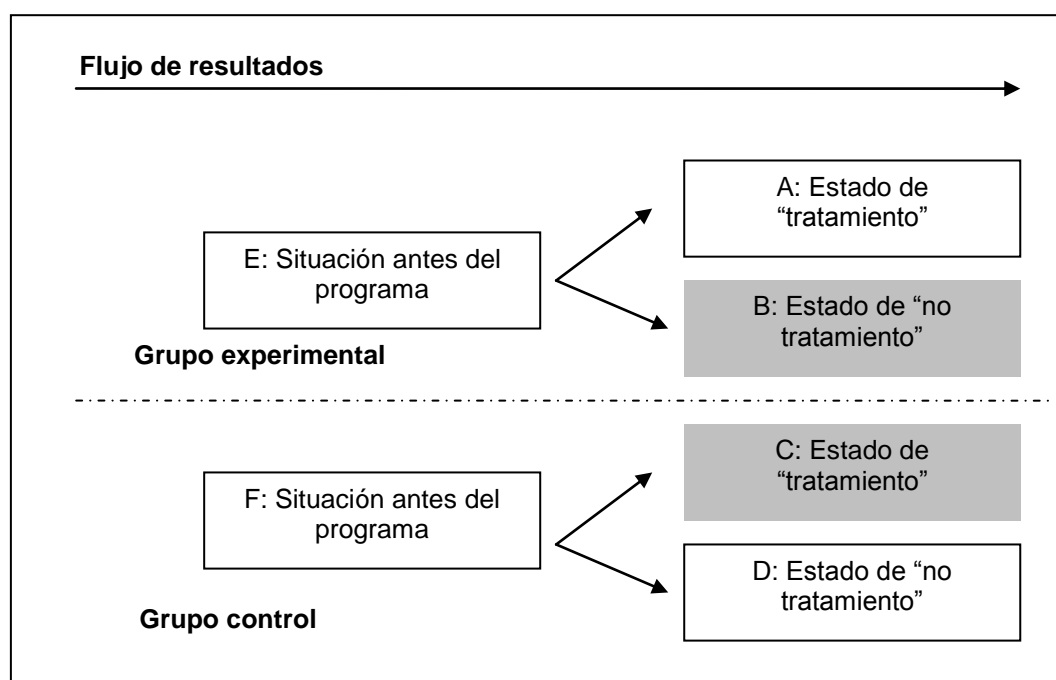


Figura N° 4.5. Estados posibles para beneficiarios y controles (Fuente: Ñopo & Robles, 2002, con modificaciones)

De esta representación debe quedar claro que el impacto sobre el grupo de beneficiarios es igual a $(A - B)$. Si lo que se quiere medir es el impacto sobre los controles entonces se compara C y D.

El **estimador antes-después** resulta de comparar la situación de los beneficiarios del programa antes del programa (E) con su situación después de haber pasado por el programa (A). En este caso se usa como grupo de control a los mismos beneficiarios en su situación anterior al programa, siendo el estimador de impacto el equivalente a $A - E$. El supuesto detrás de la utilización correcta de este estimador es que (E) logra aproximar bien la situación de los beneficiarios en caso no hubieran participado en el programa (B). El principal problema de este estimador es que la simple comparación antes y después puede llevar a atribuir erróneamente al programa cambios que se hubieran dado en ese grupo de beneficiarios independientemente de su participación en el programa, tal como se ha demostrado en los fundamentos teóricos de la validez causal.

Dado que se trabaja con medias y se asume que el promedio del error de aproximación tiende a cero, este estimador, como se mencionó líneas arriba, también se puede construir a partir de información sobre una misma población (no necesariamente conformada por los mismos individuos) en distintos momentos en el tiempo. En este sentido, la ventaja de utilizar este estimador es que sólo se requiere de información de los participantes antes y después del programa. Sin embargo, la principal desventaja es el supuesto de que el error de aproximación tiende a promediar cero. Bajo este supuesto, el resultado promedio de los participantes en el estado de “no tratamiento” es el mismo en t y t' , lo cual requeriría que no se hubiesen dado cambios en el entorno económico o cambios en la posición en el ciclo de vida del cohorte de participantes durante ese lapso, algo difícil de imaginar (Ñopo & Robles, 2002).

El **estimador de diferencia de diferencias** resulta de comparar las situaciones antes-después de los beneficiarios con aquella de los controles. Así, el impacto del programa se estima mediante la siguiente fórmula $(A - E) - (D - F)$. El supuesto detrás de este estimador es que el cambio en la situación de los controles (D-F) entre el momento previo al programa y el momento posterior al programa es una buena aproximación del cambio que hubiesen experimentado los beneficiarios durante ese mismo período de no haber pasado por el programa (B-E). Una ventaja que se atribuye a este estimador es que si existe un “sesgo de selección” en no observables y éste es constante en el tiempo, entonces la doble diferencia permite limpiarlo.

Adviértase que dado que no se puede observar el paso de un “estado de no-tratamiento” a un “estado de tratamiento” será imposible observar el cambio para cualquier individuo. Pero sí se puede obtener ya sea uno o el otro de estos términos dependiendo de si el individuo es un beneficiario o un control. Ello implica que se requieren supuestos adicionales para obtener una estimación de toda la distribución del impacto a través de este estimador. Con el estimador de doble diferencias sólo se podrá estimar la media de esa distribución.

El **estimador de corte transversal** sólo toma en cuenta la situación de beneficiarios y controles después del programa. Directamente el impacto se estima a través de $A - D$. Aquí, el supuesto es que D es una buena representación de B. Esta

metodología se utiliza en los casos en los que no hay información acerca del programa al inicio del mismo, es decir no se cuenta con una línea de base, y sólo se puede recoger información luego del programa. Nótese que si las situaciones de los beneficiarios y controles son similares en el período anterior al programa, el estimador de diferencia en diferencias es equivalente al estimador de corte transversal.

Es importante señalar que tanto el estimador de diferencia en diferencias como el de corte transversal utilizan como grupo de control a individuos distintos a los beneficiarios. En general, el grupo de control está conformado por individuos que por alguna razón no participan del programa o que si quisieron entrar lo hicieron tarde, o es un grupo explícitamente identificado. Lo crucial es lograr que el grupo de control y el de beneficiarios sean lo más similares posibles. Para ello existen diversos procedimientos, siendo el más popular el “*matching*”, seguido de la “selección econométrica”.

Habiendo estimado el impacto del programa por cualquiera de los enfoques previamente mencionados, es posible desagregar el mismo en dos componentes: uno atribuible a cambios en las características individuales de los beneficiarios y el otro a cambios en los retornos que estas características tienen. Así por ejemplo, en un programa de capacitación laboral, uno de los principales cambios en las características observables de los beneficiarios se observaría en la capacitación ocupacional que ellos obtienen al participar del programa, pero a su vez, esta capacitación ocupacional adicional podría generar una mejora en los retornos a la experiencia que ellos tengan.

Para esto, se hace necesario estimar por separado “ecuaciones de cambio” en la variable de impacto tanto para los beneficiarios como para sus respectivos controles. Esto es, se necesita estimar un modelo econométrico en el que la variable explicada sea el cambio en la variable resultado del individuo, explicada en función de una serie de características individuales relevantes.

De esta manera se obtiene un estimador de “retornos” para los beneficiarios y otro para los controles, así como estimadores de las características medias de ambos grupos. Así, el estimador de diferencias en diferencias para el impacto del programa es desagregado en dos componentes aditivos, uno explicando cambios en las características individuales de los participantes, y el otro, cambios en los retornos a dichas características.

La metodología de evaluación propuesta puede ser considerada como una metodología de “análisis de equilibrio parcial”. Ello se debe al hecho de que se concentra en la evaluación de los beneficiarios y los correspondientes controles (contrafactuales) sin tomar en consideración la manera en que ellos interactúan con el entorno socio-económico.

En general, todo programa social tiene un efecto directo sobre los beneficiarios del mismo, pero también tendrá efectos indirectos sobre otros agentes de los mercados donde tales beneficiarios operan, generando así reacciones que afectarán a su vez las relaciones económicas de los mismos en el futuro. Analizar los efectos de un programa social en un contexto en el que se presume que existen efectos indirectos no despreciables requiere un enfoque de equilibrio general, para el cual es necesario afinar y complementar las herramientas de medición.

La descomposición de los efectos del programa, en términos de cambios en las características individuales y cambios en los retornos a estas características, podría ser considerada como un primer paso hacia una estimación de equilibrio general, en el sentido que los cambios a los retornos a las características implican una interacción en el ámbito social. Sin embargo, un verdadero enfoque de equilibrio general requiere de un nivel de sofisticación aun mayor en la modelación de esos cambios en los retornos, pues en realidad corresponden a cambios en las interacciones sociales (Ñopo & Robles, 2002).

Finalmente, es conveniente analizar –en detalle– algunos procedimientos modernos para disminuir el sesgo de selección, una de las amenazas más poderosas contra la validez de los resultados del estudio de impacto.

4.2.1.3.2. Procedimientos para disminuir el sesgo de selección

Un concepto recurrente en la literatura sobre evaluación de impacto es el “sesgo de selección”. Esta idea hace referencia al hecho que quienes deciden participar en un programa lo hacen porque poseen algunas características particulares en comparación a quienes deciden no participar. Supóngase que se inaugura un comedor popular en un distrito de la ciudad y que puede acudir todo aquel que quiera hacerlo, a condición de que resida en el mismo distrito. Lo que vamos a observar es que entre los adultos que acuden a ese comedor popular la proporción de mujeres es mayor que la de hombres, cuando se compara con la proporción de mujeres adultas al nivel de todo el distrito. Supóngase que la explicación a ese fenómeno radica en el hecho de que la mayor parte de los hombres del distrito trabajan lo suficientemente lejos de ese comedor popular como para que sea costo-eficiente asistir a él. En este caso, la población que decide “participar” en el comedor popular presenta un “sesgo” a ser mayoritariamente población que trabaja lo suficientemente cerca del lugar o que no trabaja. Ello es más común entre las mujeres.

Como regla general, puede pensarse que quienes deciden participar en un programa son aquellos para quienes luego de una evaluación costo-beneficio el programa representa su mejor opción, quienes deciden no hacerlo es porque tienen mejores opciones de acción disponibles. La pregunta que surge es por qué para algunos individuos el programa puede ser una buena opción y para otros no. Ello se explica por el hecho de que para un individuo cualquiera su “flujo de resultados” en un determinado “estado de la naturaleza” depende de sus propias características o las de su entorno. Por tanto, el “impacto” del programa sobre ese individuo dependerá también de esas características.

Un ejemplo de lo anterior sería pensar cómo será el impacto del programa de vaso de leche si éste se lleva a colegios privados. Sin duda que el impacto sería menor que el potencial impacto que se puede alcanzar en colegios públicos, ya que los ingresos familiares, acceso a agua, educación de los padres, etc. es mayor entre este último grupo de niños y por tanto es mucho más probable que cuenten con un mejor nivel nutricional. El programa vaso de leche poco les aportará a estos niños, pero sí mucho a los niños de colegios públicos. Una distinción adicional que vale la pena establecer es el “sesgo de selección” sobre variables observables del “sesgo

de selección” sobre variables no observables. La primera se da cuando los que deciden participar en un programa lo hacen sobre la base de contar con características que pueden ser observadas por el evaluador o investigador. En el ejemplo del comedor popular sería muy sencillo preguntar a una muestra de personas del distrito si asisten o no al comedor popular y cuán lejos se encuentra su centro de trabajo. En este caso podremos observar que los asistentes al comedor se “seleccionan” de entre quienes trabajan cerca y los que no asisten se “seleccionan” de entre quienes trabajan lejos.

En otros casos no será posible observar la característica que produce el “sesgo de selección”. Por ejemplo, si los individuos que deciden participar en un programa son aquellos que cuentan con una cualidad de liderazgo difícilmente el evaluador dispondrá de una variable que mida esa cualidad. Buena parte de la literatura sobre evaluación de impacto conviene en usar el término “sesgo de selección” para referirse únicamente al sesgo de selección en variables no observables.

Además, toda técnica de emparejamiento va acompañada de un nivel de tolerancia en la mínima distancia a considerar. Según esto, si para cierto beneficiario sucede que el control que minimiza la distancia (Euclídeana, Mahalanobis, Propensity o cualquier otra) aún se encuentra muy alejado, es decir, la distancia entre ambos es mayor al nivel de tolerancia, el emparejamiento no se realiza. Así, el beneficiario a evaluar deja de ser considerado en el análisis por falta de un adecuado control.

A medida que el nivel de tolerancia se hace más exigente, dos efectos interesantes (y correlacionados) son observados: por un lado la probabilidad de encontrar beneficiarios no emparejados aumenta, pero por el otro, el “grado de semejanza” entre la muestra de beneficiarios y controles también aumenta. Se conoce a esta coexistencia de efectos como el balance entre el emparejamiento incompleto y el imperfecto. El emparejamiento incompleto es no deseado porque eventualmente implica sesgos (o falta de representatividad) en la muestra de beneficiarios emparejados. El emparejamiento imperfecto es no deseado porque implica diferencias en características observables entre los beneficiarios y sus respectivos. La elección del nivel de tolerancia requiere de un adecuado balance entre estos dos efectos.

Un problema en el emparejamiento sobre el cual no se tiene control es el de la existencia de diferencias en características no-observables. Un supuesto modelístico en el que se basa la literatura de evaluación es que la minimización de las diferencias en características observables estará ligada a mínimas diferencias en características no-observables. Este supuesto, aunque razonable, es imposible de validar empíricamente. En la medida que se pueda trabajar con un vasto conjunto de características observables y bajos niveles de tolerancia en las diferencias en dichas características entre beneficiarios y controles, se podrá **suponer** que las diferencias en características no-observables están bajo control. Este punto será tratado posteriormente.

4.2.1.3.2.1. El matching

El método del “*matched comparisons*” consiste en asignar para cada beneficiario un respectivo individuo-control. Dicha asignación se realiza sobre la base de emparejar aquel beneficiario con aquel control que sean lo “más parecidos”. Una vez establecidas las parejas de beneficiarios y controles “emparejados” se podrá estimar el impacto promedio del programa como el promedio del impacto para cada pareja. La idea de encontrar un “*match*” o pareja para cada beneficiario apunta a reducir potenciales “sesgos de selección”, especialmente en variables observables. Existen diversas maneras de realizar el proceso de emparejamiento y de entender la noción del “más parecido”⁶². A continuación se revisan algunos métodos particulares de “*matching*”.

1. **Minimización de Distancia Euclideana.** Se asigna a cada beneficiario aquel control que tenga la mínima distancia euclideana entre sus características y las del control. Operativamente, primero se estandarizan las variables que representan aquellas características en las cuales queremos que beneficiarios y controles sean parecidos; ello evita el problema de trabajar con variables que se encuentren en distintas unidades de medida. En segundo lugar se computa la distancia entre un beneficiario y cada uno de los controles disponibles. Se asigna como control para ese beneficiario al control que arroje la menor distancia. Se repite el mismo procedimiento para cada uno de los beneficiarios.
2. **Minimización de Distancia Mahalanobis.** Una variante a la medición de distancias euclidianas consiste en utilizar como métrica de distancia la propuesta por Mahalanobis, que es similar al de la distancia euclideana pero además de corregir por las varianzas, lo hace también por la covarianzas entre las variables. Es superior al método de las distancias euclidianas en un doble sentido: en primer lugar no hace falta tomar en cuenta el problema de las distintas unidades de medida porque la corrección es hecha automáticamente y, por otro lado, corrige por la posible correlación entre las variables consideradas. Ahora bien, como sucede muy a menudo, dichas ventajas traen un costo, en este caso computacional: es necesario computar la matriz de varianzas-covarianzas de las variables consideradas en el emparejamiento y luego realizar transformaciones lineales sobre las variables originales.
3. El método del “*propensity score*”. El “*propensity score*” es una medida que trata de establecer el grado en que un individuo cualquiera hubiese tendido a participar en el programa que se está evaluando. Así, se utiliza un modelo de regresión del tipo “*probit*” o “*logit*” para estimar la propensión a participar en el programa utilizando la muestra disponible de beneficiarios y controles. Como variable dependiente se usa una variable dicotómica (1 si se trata de beneficiarios y 0 si se trata de control) y como variables predictoras de la regresión se utilizan todas aquellas variables en las que se desea similitud entre beneficiarios y controles y aquellas que puedan afectar la decisión de participar o no en el programa. Una vez estimada la regresión se computa para cada

⁶² Para aplicar este procedimiento es muy importante determinar las potenciales fuentes de sesgo de un estimador de impacto y cómo el método de “*matching*” logra evitarlos.

individuo la “propensión” predicha por el modelo, es decir el llamado “*propensity score*”. Luego se asigna como pareja de un beneficiario a aquel control que posea el “*propensity score*” más cercano al de ese beneficiario. Originalmente desarrollado por Rosenbaum y Rubin (1983), éste es un método de emparejamiento que se basa en las probabilidades de participación (*propensity scores*) en vez de variables específicas. Este tipo de metodología ha sido utilizada intensivamente en los últimos años a nivel internacional y, por tanto, ha estado en el centro del debate metodológico reciente⁶³. Sin embargo, a pesar de sus ventajas aparentes, existen algunas desventajas serias que deben ser tomadas en cuenta para evitar confusiones y nuevos errores (Yanovitzky, Zanutto & Hornik, 2005).

Tabla N° 4.9. Ventajas y desventajas del *Propensity Score Matching* (Basado en Cueto & Mato, 2004, con modificaciones).

Ventajas	Inconvenientes
<ul style="list-style-type: none"> - Menor costo que los diseños experimentales, ya que generalmente las muestras proceden de otras fuentes diferentes a las de la administración del programa. - Puede dar respuesta a todos los parámetros de interés. En particular puede dar los indicadores de distribución del impacto. - Implica la utilización de modelos de comportamiento social en la modelización, lo que deja de lado el enfoque únicamente estadístico prevaleciente en los diseños experimentales. - No hay limitaciones éticas como en la experimentación social. - No se incurre en el sesgo de aleatorización, ni hay efectos del tratamiento pues los individuos no saben si están en el grupo de tratamiento o control ya que son datos que se observan y no se inducen. - Generalmente las muestras que se utilizan son grandes. 	<ul style="list-style-type: none"> - Si no se logra conformar un grupo de tratamiento que se comporte como si hubiera sido asignado aleatoriamente, el método fracasaría, puesto que toda evaluación de impacto tendría estimaciones inconsistentes. - Se necesita una elaboración estadística muy afinada para evitar dar resultados diferentes utilizando el mismo conjunto de datos. El desarrollo de estos instrumentos ha sido muy importante en la última década pero la proliferación de métodos han vuelto confuso contar con un resultado fiable, ya que los resultados diferentes pueden dar lugar a que la utilización de los métodos se haga en función de querer mostrar mejores o peores resultados de un programa. - El conjunto de datos disponible, no forzosamente llena los requisitos que el evaluador necesita que cumpla. Como se necesita que los individuos se comporten como si fueran participantes o grupo de control, a veces los datos no se adecuan a los perfiles de uno u otro grupo. - Se necesita ineludiblemente abordar con rigurosidad la estimación del sesgo de selección mediante diferentes estrategias que pueden dar diversidad de resultados. - Es posible que exista “sesgo de contaminación”, en el sentido de que en los datos como grupo de control puede haber individuos que hayan pasado por capacitación similar a la del programa a evaluar.

Una variante en los métodos de emparejamiento consiste en construir **controles ficticios**. Así, ya sea que se haya usado el método de distancias euclidianas o “*propensity score*” se sugiere utilizar como control de un

⁶³ Desde las propuesta iniciales de Rosenbaum y Rubin (1983, 1984) se han desarrollado numerosas aplicaciones económicas por autores diversos Heckman, Ichimura y Todd (1997); Dehejia y Wahba (1999); Hotz, Imbens y Mortimer (1999); Lechner (1999); Heckman, LaLonde y Smith (1999); Sianesi (2001); y Bryson, Dorsett y Purdon (2002).

beneficiario el promedio de los 5 o 10 controles más cercanos a él. Así, se puede obtener un promedio para la(s) variable(s) resultado de los controles más cercanos. Otra modificación es construir la pareja de un beneficiario a partir de promedios ponderados de todos los controles disponibles. En este caso se dará mayor ponderación al control más cercano y menor ponderación al control más lejano. El valor de las ponderaciones se puede determinar a partir de lograr estimaciones de la distribución de las distancias entre un beneficiario dado y todos los controles.

Una vez hecho el proceso de emparejamiento o “*matching*”, es decir una vez que se cuente ya con una pareja (verdadera o ficticia) para cada beneficiario, se puede estimar el estimador de impacto de corte transversal o el de doble diferencias.

Todos estos métodos de emparejamiento surgen para disminuir la presencia de un “sesgo de selección” debido a variables observables. Sin embargo, cuando el sesgo de selección proviene por variables no observables, una manera de abordar este problema es a través del uso del “**ratio inverso de Mills**”. Ello supone correr una regresión primero para explicar la participación en el programa, teniendo cuidado de incluir entre los regresores alguna variable que explique la participación pero que no explique la variable “resultado” que se está evaluando. Luego se puede construir el ratio “inverso de Mills” para cada individuo, que es un indicador que aporta información sobre las características no observables de los individuos. Posteriormente, se puede estimar una ecuación sólo para la muestra de no beneficiarios para predecir la variable “resultado” (por ejemplo, ingresos) empleando el ratio inverso de Mills como un regresor adicional. Finalmente, para cada beneficiario se puede estimar el impacto como la resta del resultado efectivamente alcanzado por él, menos una predicción de su resultado utilizando los parámetros estimados mediante el último modelo de regresión.

4.2.1.3.2.1. El uso de variables instrumentales

Existen algunos programas en los cuales los beneficiarios son elegidos bajo ciertos procedimientos o por razones motivacionales. En este caso, la participación es una variable de elección endógena y casi siempre afecta el proceso de elección.

En este caso, el proceso de selección contiene errores que se confunden con los efectos del tratamiento, sobreestimándolo o mitigándolo. El problema entonces ocurre cuando la correlación entre el Tratamiento (T) y los errores (ε) es diferente de cero ($\text{Corr } T, \varepsilon \neq 0$). Tal correlación es casi inevitable en los programas de participación voluntaria (Ej. Programas de Capacitación Laboral, ProJoven, etc.), pues la motivación para participar es diferente, la habilidad es diferente, la información disponible es diferente, el costo de oportunidad de participar es distinto y existe distinto nivel de acceso (Vermeersch, 2006).

Una de las estrategias para controlar o “limpiar” la correlación entre T y ε es mediante el uso de variables instrumentales. El procedimiento busca aislar la variación en T que no está correlacionada con ε . Una variable instrumental es una variable que está correlaciona (teóricamente) con T ($\text{corr } Z, T \neq 0$) y no correlacionada con ε ($\text{corr } Z, \varepsilon = 0$). La primera etapa consiste en identificar la variable instrumental; la segunda etapa consiste en regresionar la variable

endógena sobre la variable instrumental y otros regresores exógenos, calculando el valor predicho para cada observación. Finalmente, se regresa el resultado y sobre la variable predicha (y otras variables exógenas).

Es importante advertir que la variable instrumental elegida debe tener suficiente poder en predecir T. De otro modo, tenemos instrumentos débiles que llevan estimaciones de variables instrumentales sesgadas (Vermeersch, 2006).

4.3. CRITERIOS METODOLÓGICOS

En concordancia con el parámetro teórico de la causación, el Modelo de Evaluación de Impacto ARYSIS se fundamenta también en un criterio metodológico esencial: los diseños cuasi-experimentales. Esto debido a que la historia de la metodología cuasi-experimental es inseparable del desarrollo de la teoría de la validez de la inferencia causal.

Campbell & Stanley (1966) han sido los pioneros en definir y delimitar la naturaleza de estos métodos. Para ellos, los diseños cuasi-experimentales se ubican entre los diseños pre-experimentales y los experimentales. El diseño cuasi-experimental se caracteriza porque emplea escenarios sociales y porque carece de un control experimental completo, resultando imposible controlar ciertas variables extrañas. Los diseños cuasi-experimentales suelen utilizar diversas estrategias como sustitutivos del control experimental aleatorizado. De este modo, se consigue minimizar y, en algunos casos, eliminar los efectos de numerosas fuentes de invalidez interna.

Al igual que los diseños experimentales, en los diseños cuasi-experimentales también existe, al menos, una variable independiente (siempre es el programa social) para observar su efecto y relación con una o más variables dependientes (el impacto esperado). En todo caso, difieren de los experimentos en el grado de seguridad que se tiene sobre la equivalencia inicial de los grupos; pues, en este caso, los participantes no se asignan al azar ni se emparejan, sino que dichos grupos ya estaban formados antes del experimento⁶⁴.

Así, los diseños cuasi-experimentales se utilizan cuando no es posible asignar a los participantes en forma aleatoria, como comúnmente se hace en los diseños experimentales. Por tanto, la falta de aleatorización introduce posibles problemas de validez interna si, por ejemplo, varios elementos pueden influenciar en la formación de los grupos y que no están bajo el control del evaluador. Por eso, dada que la validez interna de estos diseños es menor a la de los experimentos, reciben el nombre de cuasi-experimentos; por eso el evaluador debe intentar establecer semejanzas entre los grupos utilizando diversos procedimientos (Schoeni & Blank, 2001; Trochim, 1986).

⁶⁴ A este tipo de grupos se les llama grupos intactos porque la razón por la que surgen y la manera como se formaron fueron independientes del experimento.

4.3.1. Definición de cuasi-experimento

Shadish, Cook & Campbell (2002) explicitan que los diseños cuasi-experimentales se caracterizan por la manipulación de cuatro elementos básicos, con sus respectivas alternativas:

- Asignación
- Medición u observación de la variable de interés (“O”).
- Grupos de comparación (GE y GC)
- Tratamientos (“X”)

Por asignación se entiende la distribución de los participantes y de los tratamientos (programas) en el grupo experimental (GE) y eventualmente también entre los otros grupos de control (GC)⁶⁵.

En casi todos los cuasi-experimentos la asignación no es controlada por el evaluador. Al contrario, los participantes ya están distribuidos en grupos formados o algún otro agente hace la asignación mediante criterios no probabilísticos. Justamente, tal como se mencionó, la diferencia esencial entre el diseño experimental y el cuasi-experimental es la ausencia de la asignación aleatoria⁶⁶. Por ello, los diseños cuasi-experimentales han sido inventados para suplir a los experimentales cuando no es posible asignar grupos aleatoriamente. Y esto es muy común en las ciencias sociales, donde los estudios no aleatorizados son, a veces, el único medio ético y posible de realizar investigación, pues son usualmente más baratos, políticamente sensibles y se aproximan en mayor medida a las situaciones cotidianas.

Cuando hay varios grupos, la asignación puede frecuentemente controlarse a través de otras formas no aleatorias. El “*matching*” y la estratificación, por ejemplo, pueden incrementar la similaridad entre los grupos. El ocultamiento (“*masking*”) a los evaluadores, a los participantes o a otros integrantes del *staff*, respecto de cuáles son los grupos de tratamiento y cuáles los de comparación suele también disminuir la reactividad y la intrusividad.

Varias amenazas contra la validez que operan mediante la temporalidad (historia, maduración, regresión a la media), el trabajo de registro (administración de test, instrumentación), la selección (efecto selección, mortalidad, interacciones), pueden anularse hasta cierto punto incluyendo un pretest, tanto en la variable dependiente, como en otras variables de interés pero no equivalentes entre sí, o en variables relacionadas entre sí (covariantes). Es de notar que este puede ser efectivamente realizado por el evaluador, pero también en forma retrospectiva (sea

⁶⁵ Existen tres procedimientos de asignación de grupos control y experimental (Judd & Kenny, 1981): a) asignación aleatoria (en el caso de todos los diseños experimentales) la cual garantiza la equivalencia inicial de los grupos, b) no aleatoria pero conocida; es decir, cuando se asigna a uno de los grupos en función de la puntuación obtenida en un pretest de la misma medida utilizada para evaluar el impacto del tratamiento y c) no aleatoria ni conocida, cuando no se conoce exactamente sobre qué variables se constituyen los grupos.

⁶⁶ La aleatorización es la clave para la equivalencia inicial de los grupos experimental y control. La aleatorización implica una distribución similar de las características de un sujeto en cada grupo y facilita la inferencia causal. Elimina, también, los sesgos de selección y proporciona la base segura para la inferencia estadística.

produciendo directamente el dato o relevándola de información secundaria disponible).

Los grupos de comparación (GC) son elementos del diseño destinados a proporcionar evidencia respecto de la plausibilidad de las inferencias contrafactuales (¿hubiera sucedido “O” si “X” no se hubiera verificado? ¿Se pueden descartar las otras hipótesis alternativas Z, para aquellos casos en que “X” no está presente?). Es de precisar, sin embargo, que una parte importante de los diseños cuasi-experimentales carecen de grupo de control, circunstancia que ha generado desarrollos metodológicos muy particulares para contrarrestar esta carencia. Un posible sub-tipo de diseño muy importante y de gran utilidad son las “series temporales interrumpidas”, donde la reiteración de mediciones sobre las “O” antes y después de la “X” sustituyen la carencia de grupo de comparación. Otros diseños, sin llegar a la alternativa de las series, son los diseños pretest y postest, con múltiples variables dependientes sustantivamente interesantes, vinculadas hipotéticamente a la presencia, ausencia o variaciones en la “X”.

En los cuasi-experimentos, los grupos de comparación son del tipo “no equivalentes” en la medida en que los sujetos no han sido asignados aleatoriamente a ellos. En muchos casos se trata de grupos (organizaciones, pequeños grupos, territorios, comunidades) pre-existentes a la investigación.

Una primera forma de elección de los grupos de comparación se puede realizar mediante el proceso de “*matching*” o equiparación. El objetivo es encontrar grupos semejantes entre sí respecto de un conjunto de variables sustantivamente importantes, pero que no forman parte de la explicación. Actualmente este proceso se hace ajustando un modelo de regresión logística o regresión probit, cuyas variables explicativas sean ese conjunto de variables importantes, independientes de las variables “X” (programa). Se terminan eligiendo aquellos casos que tienen similares probabilidades de haber sido grupos de tratamiento. Esto es, que si hubiera existido asignación aleatoria de tratamientos, cualquiera de ellos hubiera sido buen candidato para tal fin (Przeworski & Teune, 1970).

4.3.2. Tipología cuasi-experimental

En este subtítulo se presentará a los diseños cuasi-experimentales más conocidos, así como una breve referencia a las amenazas contra la validez interna, de conclusión estadística, de constructo y externa más comunes en cada uno de ellos (Cook, Campbell & Peracchio, 1990). Para cada diseño se presenta, también, una tabla donde se identifican varios aspectos de interés para su valoración, entre ellos:

1. El número de grupos existentes, si es único o múltiple y, en este caso, la especificación de los grupos. La situación más común distingue entre grupo experimental (GE) en donde se aplica el programa social y grupo de control (GC) en donde no se aplica.
2. La regla o variable de asignación que se utiliza: si no existe, porque se trata de un grupo único (NE), si es aleatoria (A), si no es aleatoria ni conocida (NA), o si no es aleatoria pero sí conocida (NAC).

3. Finalmente, la secuencia de tratamiento seguida, donde las **Y** se refieren a las medidas registradas, medidas pretest si se efectúan antes de la administración de programa y medidas posttest si se efectúan con posterioridad. El orden de registro de las medidas y el grupo al que pertenecen la observación determina el subíndice que acompaña a cada **Y**. La administración del programa a un grupo se simboliza con **X**, la no administración con guión continuo (--) y su retirada, tras haber sido administrada, con $_{no}X$. Una gradación del programa se simboliza numéricamente (X_1, X_2, X_3 , etc.) y un programa invertido (positivo para unos y negativo para otros) con los símbolos apropiados ($X+$ y $X-$).

En este contexto, es posible distinguir dos grupos de diseño cuasi-experimentales clásicos, que se refieren entre sí por la secuencia de registros de observaciones previa y posterior a la introducción del tratamiento o intervención.

- Los **diseños de grupo pretest-postest** se caracterizan por la existencia de una medida (o unas pocas medidas) obtenidas antes del programa (Pretest), una medida post-intervención (Posttest) que refleja el efecto del programa, y una variable de asignación que determina la pertenencia a (al menos) uno de los grupos: un grupo experimental (GE) que recibe el programa y un grupo de control (GC) que no lo recibe. Dependiendo de la naturaleza, conocida o desconocida, pero no aleatoria de la variable de asignación, cabe a su vez designar dos tipos de diseños de grupos pretest-postest: a) los diseños con grupo de control no equivalente forman una amplia clase de diseños caracterizados por el uso de una variable de asignación desconocida frente a b) los diseños de discontinuidad de la regresión, caracterizados por el uso de una variable de asignación conocida.
- Los **diseños de series temporales interrumpidas** constituyen una clase alternativa cuya característica fundamental es la existencia de múltiples medidas previas y múltiples medidas posteriores a la introducción del programa social, el cual también puede ser múltiple.

En la siguiente figura se presenta la tipología general de los diseños cuasi-experimentales.

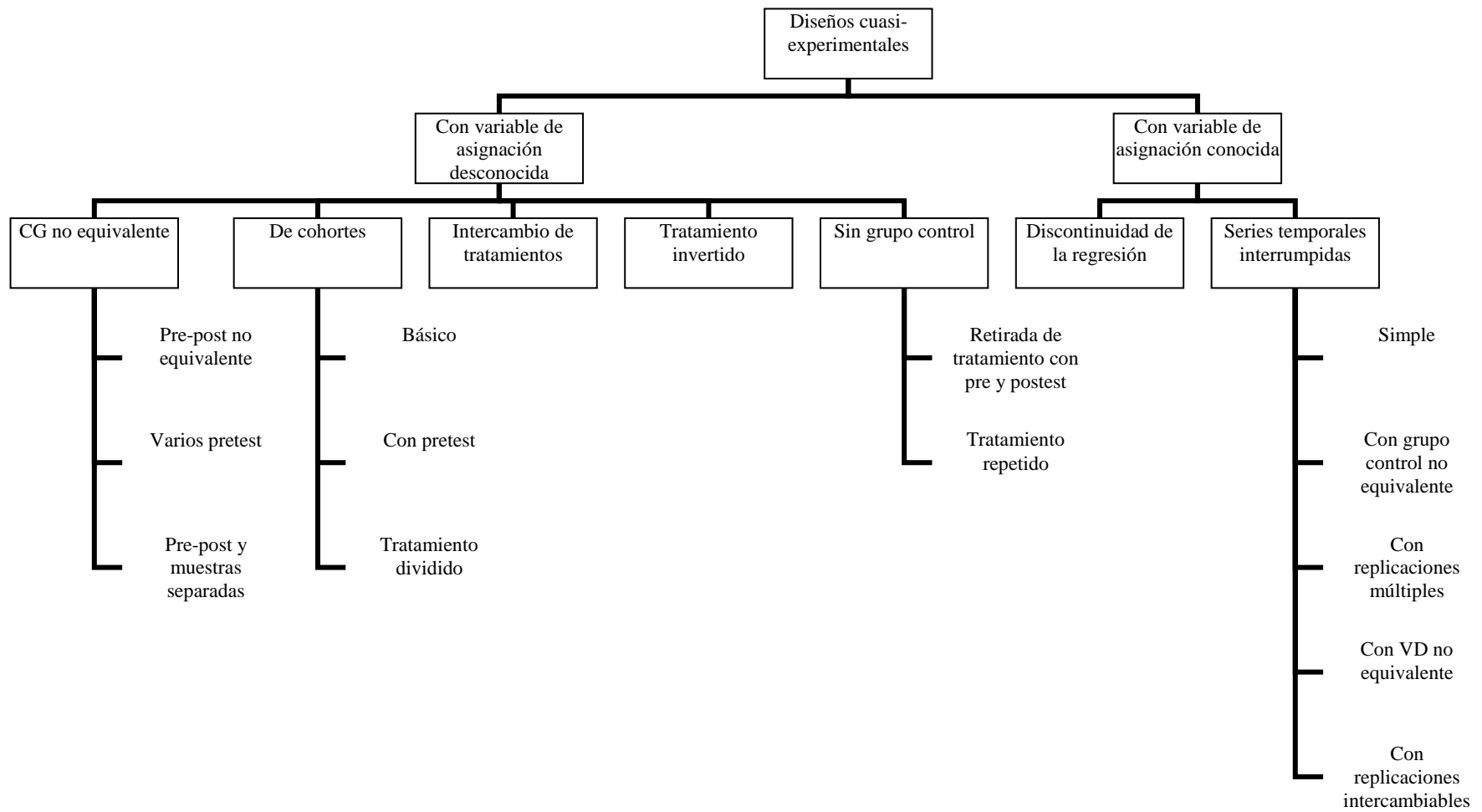


Figura N° 4.6. Tipología del diseño cuasi-experimental (Fuente: Elaboración propia).

4.3.2.1. Los módulos básicos más empleados en la EvPro tradicionales.

Los diseños pretest-postest y solamente postest, son los módulos más empleados en la evaluación de impacto de los programas sociales en el Perú. Estos dos diseños son pre-experimentales y, aunque no permiten establecer con suficientes garantías la relación causa-efecto, sin embargo presentan módulos básicos a partir de los cuales se configuran diseños más complejos.

4.3.2.1.1. Diseños pretest-postest sin grupo control

El primero es el diseño pretest-postest sin grupo control (Cook & Campbell, 1979; Kish, 1987; Cook, Cambell & Peracchio, 1990; Mohr, 1992) el cual requiere una observación antes de la aplicación del programa (X) registrada en un grupo único de individuos (Y_1); y otra observación post-intervención (Y_2). Puesto que solo existe un grupo de participantes, la asignación de grupos no existe (NE). La siguiente tabla resume el diseño:

Tabla 4.10. Diseño pretest-postest sin grupo control

Grupos	Asignación	Secuencia de registros		
		Pretest	Tratamiento	Postest
Único	No Existe	Y1	X	Y2

Con este diseño, la inferencia causal se obtiene a través de la comparación entre pretest y postest, que al implicar los mismos sujetos es de carácter intrasujeto. A pesar que este modelo es aún uno de los más usados en la evaluación de los programas sociales, la carencia de grupo de control hace difícil establecer la causalidad del impacto del programa (X).

Por ejemplo, un programa de asistencia alimentaria ha tomado la medida de “índice nutricional” antes y después de distribuir los alimentos en una población de 200 familias de un pueblo joven. Intenta evaluar el impacto del programa y tiene los siguientes resultados:

Tabla 4.11. Representación de resultados del diseño pretest-postest sin grupo control

Grupo	N	Pretest	Postest
Único	200	5.28 (D.E.=1.79)	7.41 (D.E.=2.73)

En este caso, el efecto del programa alimentario es la diferencia entre las medidas de postest y pretest ($7.41 - 5.28 = 2.13$). La hipótesis nula que se somete a comprobación es que ambos promedios son iguales y la diferencia encontrada es debido al azar. La prueba estadística que se usa frecuentemente para evaluar el efecto del programa es la “t de Student para grupos relacionados”. Generalmente aquí termina el asunto y se asume que el programa ha sido efectivo, sin embargo, existen demasiadas amenazas que no han sido consideradas.

En efecto, aunque aparentemente ha aumentado el índice nutricional de la población asistida, no se tiene total certeza de que esto haya ocurrido por causa del

programa asistencial. Tal como se ha explicado en el parámetro teórico, existen explicaciones alternativas que no han sido controladas. Además es posible la existencia de algunas amenazas contra la validez interna involucradas en comparaciones intra-sujeto, tales como:

- La historia (eventos diferentes del programa, ocurridos entre pretest y postest, que afectaron el índice nutricional, por ejemplo, la instalación de un comedor popular),
- la regresión estadística (los índices nutricionales pudieron alcanzar valores extremos, produciendo efectos techo-suelo y un promedio aparente),
- la maduración (los índices nutricionales pudieron estar en aumento por razones biológicas temporales),
- otras amenazas tales como la administración de pruebas y la instrumentación.

Todas estas amenazas producen sesgos significativos en los resultados y aumentan la incertidumbre sobre la relación causal entre el programa y su impacto en la población. En las simulaciones se demostrará los efectos perniciosos de estas amenazas en este tipo de diseño.

En general, el diseño pretest-postest no permite argumentos válidos de causalidad, a menos que se trabaje con cortos o muy cortos intervalos de tiempo entre pretest y postest y, por tanto, no resulta recomendable en situaciones donde se evalúa un programa social a mediano o largo plazo. Solo se aconseja si se utiliza con propósito puramente exploratorio o ampliado con un grupo de control.

4.3.2.1.2. Diseños solamente postest con grupo control

El diseño solamente postest utiliza dos grupos: un grupo experimental (GE) que recibe el programa social (X) y un grupo de control (GC) que no lo recibe (Cook & Campbell, 1979; Judd & Kenny, 1981; Kish, 1987; Cook, Campbell & Peracchio, 1990; Mohr, 1992). La asignación a los grupos se establece como en todos los diseños cuasi-experimentales, mediante el empleo de una regla de asignación no aleatoria (NA). La siguiente tabla resume el diseño:

Tabla 4.12. Diseño solamente postest

Grupos	Asignación	Secuencia de registros		
		Pretest	Tratamiento	Postest
Experimental (GE)	NA	-	X	Y ₁
Control (GC)	NA	-	-	Y ₂

En este diseño, la inferencia causal se obtiene de la comparación entre grupo experimental y grupo control, que al implicar a diferentes sujetos es de carácter intersujeto. Sin embargo, este diseño también adolece de la ausencia de una regla de asignación aleatoria y, además, carece de medidas pretest, imposibilitando el uso de procedimientos de ajuste estadístico para compensar la no equivalencia inicial de ambos grupos. Como consecuencia, las amenazas contra

la validez interna limitan la inferencia casual (amenazas involucradas en comparaciones intrasujeto y, particularmente, la selección). Pese a todo, este diseño se utiliza profusamente en contextos aplicados, bajo la denominación de investigación *ex post facto*.

A modo de ejemplo, un programa de educación sexual ha sido desarrollado en una comunidad educativa. El programa ha sido dirigido a 1300 adolescentes escolares de diversos centros educativos (grupo experimental), mientras que otros dos centros educativos cercanos a los primeros han sido omitidos de la capacitación (grupo control con 1,300 alumnos). El efecto del programa es ahora la diferencia entre las medidas de ambos grupos ($16.45 - 11.60 = 4.85$) y la hipótesis nula es que ésta diferencia no es estadísticamente significativa. La prueba estadística aplicada, comúnmente, es la “t de Student para grupos independientes”.

Tabla 4.13. Representación de resultados en diseño solamente posttest

Grupo	N	Postest
GE	1300	11.60 (3.19)
GC	1300	16.45 (5.07)

En el diseño solamente posttest, la ausencia de pretest suele compensarse utilizando algún tipo de medida previa, bien sea procedente de archivos existentes o directamente de los sujetos, por uno de los procedimientos siguientes (Cook, Campbell y Peracchio, 1990):

- Utilizando pretests retrospectivos (Hutton & McNeil, 1981). Es preciso suponer en este caso que el programa no posee ninguna influencia en la memoria de los sujetos afectando así a las medidas retrospectivas.
- Formando los grupos experimental y control a posteriori mediante emparejamiento (Anderson et al., 1980; Ato, 1991; Levy et al., 1985).
- Empleando variables delegadas (“Proxy variables”), que son variables que sustituyen a las variables originales y correlacionan con el posttest dentro de cada uno de los grupos pero no se miden en la misma escala (Rao & Miller, 1971). Las más populares variables delegadas son las demográficas "edad", "sexo", "clase social", "raza", "lugar de nacimiento", etc.

4.3.2.2. Diseños con variable de asignación no conocida

4.3.2.2.1. Diseño con grupo control no equivalente

El prototipo de todos los diseños cuasi-experimentales de grupo pretest-postest con variable de asignación no conocida es el diseño con grupo control no equivalente. Básicamente, consta de un grupo experimental (GE) que recibe algún tipo de tratamiento, y un grupo de control (GC) no tratado y sus unidades son evaluadas con el mismo instrumento de medida, al menos en dos ocasiones en el

tiempo: un pretest, antes de administrar el tratamiento y, un postest, con posterioridad a su administración.

Tabla N° 4.14. Diseño con grupo de control no equivalente.

Grupos	Asignación	Secuencia de registro		
		Pretest	Tratamiento	Postest
Experimental	NA	Y_{E1}	X	Y_{E2}
Control	NA	Y_{C1}	-	Y_{C2}

Como consecuencia de no utilizar una regla de asignación aleatoria, el problema esencial del diseño con grupo de control no equivalente es que, en ausencia de efectos de tratamiento, GE y GC no obtienen puntuaciones equivalentes en el postest. Y con la finalidad de evaluar la no equivalencia y permitir la utilización de algún tipo de ajuste estadístico que equilibre los grupos, se toman medidas pretest.

Tabla N° 4.15. Presentación de resultados del diseño con grupo de control no equivalente.

Grupo	N	Pretest	Postest
GE	51	4.16 (1.11)	5.42 (1.84)
GC	60	3.09 (0.95)	3.95 (1.57)

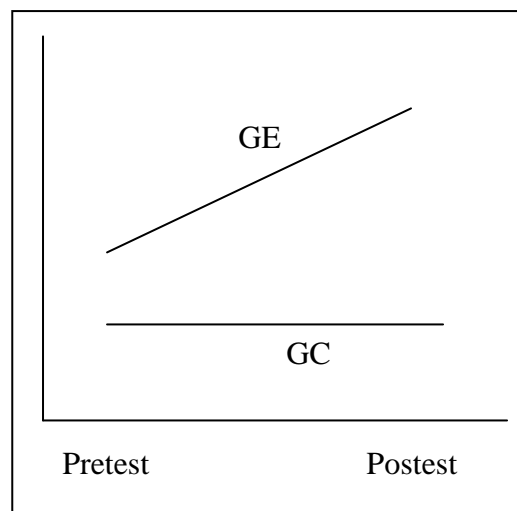
La hipótesis nula plantea que la diferencia de medias existente entre grupos en el postest ($5.42 - 3.95 = 1.47$) es fruto del azar. Sin embargo, en este caso se puede tener en cuenta la no equivalencia inicial ya que las diferencias de grupo en el pretest no son nulas ($4.16 - 3.09 = 1.07$). En este caso, el análisis estadístico es más complicado y requiere del “Análisis de Covarianza”.

Judd y Kenny (1981) han destacado varios procedimientos para configurar los dos grupos:

- El más sencillo consiste en utilizar grupos naturales ya formados antes de comenzar la investigación. La formulación original de Campbell y Stanley (1966) contempló dos grupos intactos que se asignaban al azar a las dos condiciones de tratamiento y control.
- Más usual es hacer un esfuerzo para asegurar que el GC sea aproximadamente equivalente al GE, por lo general mediante alguna forma de emparejamiento sobre variables relevantes, en la medida de lo posible relacionadas con la variable de asignación (Anderson et al., 1980; Cochran, 1983; Ato, 1991). Otra fórmula consiste en disponer de un amplio número de sujetos para participar en un programa, de donde se selecciona un grupo de voluntarios (GE) y el resto forma el GC.
- Una última fórmula consiste en examinar únicamente los sujetos sometidos al tratamiento y comparar sujetos con alta exposición al tratamiento con los que reciben baja exposición. Este enfoque es necesario en investigaciones donde, tarde o temprano, todos los sujetos terminan recibiendo el tratamiento.

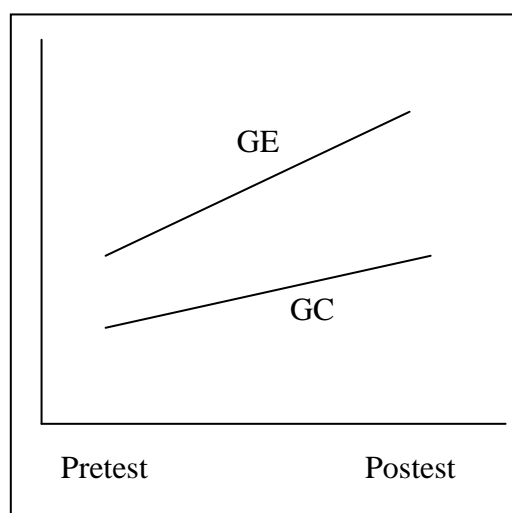
Cook & Campbell (1979) y Cook, Campbell & Peracchio (1990) han destacado cinco diferentes pautas de resultados de las cuales depende la calidad de la inferencia causal, a saber:

1. No hay cambio en GC. Mientras que el GE alcanza un notable aumento en el postest, el GC no obtiene ningún cambio. Cuando aparece esta pauta de resultados, la atribución causal al tratamiento (X) es difícil, y pasa primero por determinar las tasas de crecimiento de GE y de GC durante el período temporal pretest-postest. Las amenazas contra la validez interna más comunes en todas las pautas son las siguientes:

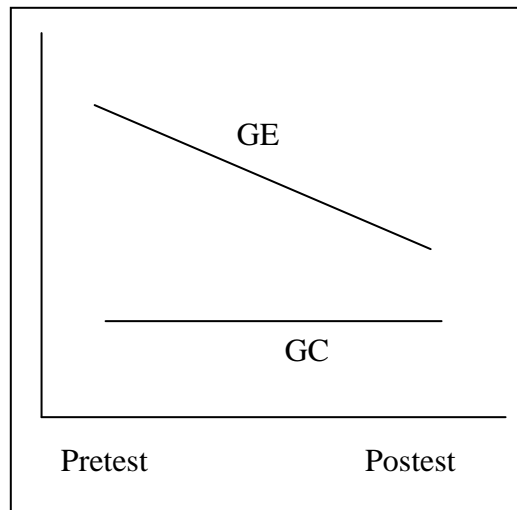


- Interacción selección-maduración: el GE, como consecuencia de la no equivalencia en el pretest con el GC, madura o cambia a diferente ritmo como consecuencia del tratamiento.
- Instrumentación: GE y GC parten de puntos diferentes de la escala de medida y sus diferencias se agudizan en función de la distancia inicial (pretest), de la distancia a cubrir entre pre y postest y de la proximidad a los valores extremos (efectos techo y suelo).
- Regresión estadística: el GC representa en el pretest un rango extremo de valores de la escala de medida respecto al GE que se aproxima al centro de la escala en el postest.
- Historia: eventos locales que tuvieron lugar diferencialmente en uno de los dos grupos han podido cambiar su trayectoria respecto del otro.

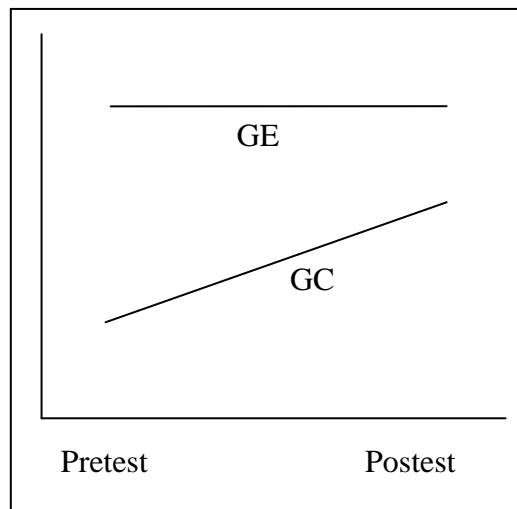
2. Grupo Experimental (GE) y Grupo Control (GC) cambian en la misma dirección pero a diferente ritmo. Esta pauta tiene más apariencia de normal que la anterior: GE y GC obtienen sendas mejoras en el postest, pero el aumento producido en el GE es mayor que en el GC. Las amenazas contra la validez interna son las mismas que en el caso anterior.



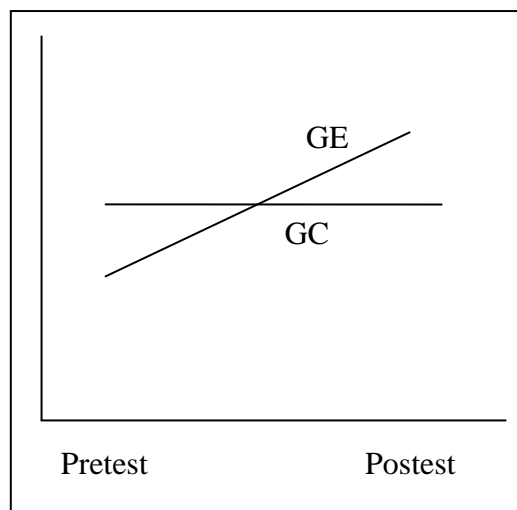
3. Las diferencias entre Grupo Experimental y Grupo Control en el pretest disminuyen en el postest. Esta pauta presenta una situación similar a la primera y es típica de contextos de investigación donde se persigue la igualación del Grupo Experimental al Grupo Control en el postest sobre la base de alguna característica específica. Presenta las mismas amenazas contra la validez interna que las pautas anteriores.



4. GE con tratamiento compensatorio con interacción ordinal. Esta pauta se presenta en contextos de investigación sobre educación compensatoria y se caracteriza porque los grupos exhiben grandes diferencias en el pretest. Una característica de este caso es la presencia de una interacción ordinal (las pautas para ambos grupos tienden a la convergencia pero no se cruzan en ningún punto) entre tratamiento y administración de pruebas. Las amenazas contra la validez interna son las mismas que en casos anteriores.



5. GE con tratamiento compensatorio con interacción no ordinal. Esta pauta, también típica de contextos de investigación sobre educación compensatoria, difiere de la anterior únicamente por el hecho de que el GE obtiene puntuaciones medias por encima del GC en el postest. Sin embargo, el cambio de tendencia, o sea, la presencia de una interacción no ordinal (las pautas se cruzan en algún punto), la hace más susceptible de interpretación en términos causa-efecto que las



anteriores, porque en gran medida se disipan las amenazas de instrumentación, interacción selección-maduración y regresión estadística, aunque todavía persiste la amenaza de la historia.

El diseño con grupo de control no equivalente se utiliza habitualmente en su forma básica, es decir, con dos grupos y dos registros de medición. Sin embargo, técnicamente ya no es aconsejable emplearlo bajo esta forma (Moffitt, 1991). Hay dos opciones, no excluyentes, en que el diseño puede mejorar sus condiciones para la inferencia causal. La primera consiste en ampliar el número de grupos, o bien mediante una gradación de tratamientos, o bien por la incorporación de uno o más grupos de control. La segunda consiste en ampliar el número de registros de medición, usualmente incorporando más pretests. Esta última solución origina el diseño con doble pretest, que se abordara a continuación.

4.3.2.2.2. Diseño con doble pretest

Este diseño es una variante más conveniente del diseño con grupo de control no equivalente, del que se diferencia porque se registra una observación más en el pretest para cada uno de los dos grupos (Cook & Campbell, 1979; Peracchio & Cook, 1988; Cook, Campbell & Peracchio, 1990). El diagrama del diseño aparece en la tabla siguiente.

Tabla N° 4.16. Diseño con doble pretest.

<i>Grupos</i>	<i>Asignación</i>	<i>Secuencia de registro</i>			
		<i>Pretest1</i>	<i>Pretest2</i>	<i>Tratamiento</i>	<i>Postest</i>
Experimental	NA	Y_{E1}	Y_{E2}	X	Y_{E3}
Control	NA	Y_{C1}	Y_{C2}	-	Y_{C3}

El tratamiento estadístico de este diseño debe considerar el ajuste del postest respecto de las puntuaciones obtenidas en el pretest y será objeto de un epígrafe posterior. La presentación de resultados es muy similar al diseño con grupo de control no equivalente. Utilizando nuevamente datos hipotéticos, con los mismos promedios de grupo y desviaciones típicas entre paréntesis del diseño anterior, un resultado típico de este diseño es el que aparece en la tabla.

Tabla N° 4.17. Presentación de resultados del diseño con doble pretest.

Grupo	N	Pretest	Pretest2	Postest
GE	51	3.85 (1.16)	4.16 (1.11)	5.42 (1.84)
GC	60	3.04 (1.20)	3.09 (0.95)	3.95 (1.57)

El diseño con doble pretest fue empleado por Wortman, Reichard y St Pierre (1976) para evaluar el efecto del programa de crédito educativo “Alumn Rock” sobre la lectura. Bajo este programa, los padres seleccionaban una escuela para su hijo y ésta recibía un crédito equivalente al coste de la educación del niño. El objetivo del programa era fomentar la competición entre escuelas para mejorar así la educación recibida. Y aunque las evaluaciones iniciales habían encontrado que el programa disminuía las puntuaciones en lectura, Wortman y colegas

registraron el rendimiento lector de un grupo de alumnos desde primero hasta tercer grado, en escuelas sin sistema de crédito (control) y con sistema de crédito (experimental). Dentro de éstos, examinaron un sistema de crédito tradicional y un sistema no tradicional. Los resultados mostraron que la disminución de las puntuaciones era debida al grupo entrenado con el sistema no tradicional con crédito, que era afectado negativamente por el programa. Los otros grupos, tradicional con crédito y sin crédito presentaban tasas de desarrollo similares.

Una de las ventajas de este diseño es que permite la evaluación de la amenaza de la interacción selección-maduración bajo el supuesto de que la tasa de cambio entre Y_2 y Y_3 se encontrará también entre Y_1 y Y_3 . Sin embargo, conviene ser cauteloso para calcular diferencias entre tasas de desarrollo, ya que dichas tasas de desarrollo se estiman con error de medida y, por otra parte, ciertos cambios de escala pueden hacer que el crecimiento entre Y_1 e Y_2 no sea representativo de lo que se espera entre Y_2 e Y_3 .

Otra ventaja deriva del hecho de que es posible detectar la presencia de valores extremos en la escala en alguna de las observaciones pretest y por tanto también puede evaluarse con rigor si la regresión estadística es una amenaza auténtica. En consecuencia, la inclusión de dos observaciones pretest puede contribuir a interpretar relaciones causa-efecto. Sin embargo, este diseño no se emplea con frecuencia, quizá debido a que no se conoce suficientemente, pero probablemente la razón más sólida es la dificultad de obtener pretests. Si casi siempre se tropieza con problemas institucionales para obtener un pretest, ¿qué decir cuando se programa obtener dos?

4.3.2.3. Diseños de cohorte

El término cohorte se emplea para indicar un grupo de individuos que experimentan un determinado evento en un mismo período (Visser, 1985) y son muy útiles en investigación (sobre todo, en investigación educativa) porque:

- Determinadas cohortes experimentan un tratamiento particular (por ejemplo, una reforma del sistema educativo), mientras que otras no lo experimentan.
- Existe, por lo general, abundante información sobre características de los sujetos de una cohorte en archivos institucionales.
- Es razonable suponer que una cohorte difiera en pequeños aspectos de matiz de cohortes previas o posteriores.

Una característica crucial que hace estos diseños particularmente útiles es que, con frecuencia, puede suponerse que existe cuasi-comparabilidad (una forma de sesgo de selección atenuado) entre, una segunda cohorte, o cohorte experimental (CE), sometida a tratamiento, y una primera cohorte, o cohorte de control (CC), no sometida a tratamiento. Aunque no cabe nunca esperar la equivalencia que se consigue con diseños aleatorizados, el grado de comparabilidad se evalúa usualmente analizando las variables que puedan ser mediadoras de los efectos de tratamiento.

4.3.2.3.1. Diseño básico de cohortes

Minton (1975) intentó probar si la primera serie del conocido programa de televisión “Plaza Sésamo” afectaba a las puntuaciones en el MRT (Metropolitan Readiness Test) de una muestra de niños de jardín de infancia. Utilizó como cohorte experimental los niños de un jardín de infancia que siguió “Plaza Sésamo”, a quienes administró el MRT al final del primer año. Como cohorte de control utilizó las puntuaciones obtenidas por los hermanos mayores de los niños cuando tenían la misma edad y no habían visto “Plaza Sésamo”. Este diseño se representa en la siguiente tabla.

Tabla N° 4.18. Diseño de cohortes básico.

<i>Cohortes</i>	<i>Asignación</i>	<i>Secuencia de registro</i>		
		<i>Pretest</i>	<i>Tratamiento</i>	<i>Posttest</i>
Primera (Control)	NA	Y ₁	--	--
Segunda (Experimental)	NA	--	X	Y ₂

En este diseño, Y₁, y Y₂ son los dos momentos en que se registran las medidas y la línea punteada indica un grado restringido de no equivalencia. Por lo demás, el diseño básico de cohortes representa, en esencia, la misma idea que el diseño con grupo de control no equivalente.

Aunque muchas de las amenazas contra la validez interna quedan eliminadas en este diseño (entre otras, maduración, regresión estadística, mortalidad e interacciones con selección, en particular la interacción entre selección y maduración), todavía pueden ser importantes algunas amenazas. Dos de las más importantes son la selección y la historia.

- La selección es la amenaza más severa, puesto que entre los hermanos mayores existe una mayor proporción de primogénitos, que se ha demostrado que alcanzan mejores puntuaciones en desarrollo intelectual que sus hermanos menores (Zajonc & Marcus, 1975). Una forma de superar este problema sería analizar los datos considerando únicamente pares contiguos en el orden de nacimiento (por ejemplo, nacidos en primero y segundo lugar por un lado, nacidos en segundo y tercer lugar por otro, etc.).
- La historia es también otra amenaza importante, puesto que la cohorte de control puede haber tenido experiencias alternativas a “Plaza Sésamo” cuya influencia se hace imposible evaluar.

4.3.2.3.2. Variaciones del diseño básico de cohortes

Diseño de cohortes con pretests. En un estudio que comparaba la efectividad relativa de profesores ordinarios de un centro con la de educadores externos contratados para estimular el rendimiento de los niños, Saretsky (1972) observó que los profesores ordinarios realizaban esfuerzos especiales y alcanzaban un rendimiento mayor de lo esperado en comparación con el de años anteriores. Aunque no está claro en su informe cómo comprobó Saretsky esta hipótesis, lo más lógico es comparar la ganancia media de la clase objeto de experiencia con la

ganancia obtenida por los mismos profesores en años anteriores. El diagrama de este diseño sería el que aparece en la tabla.

Y_1 Y_2 representan las observaciones pre y posttest de la cohorte anterior a la experiencia y Y_3 e Y_4 son observaciones pre y posttest de la cohorte que fue objeto de la experiencia. Y_1 Y_2 e Y_3 sirven en realidad en este contexto como puntuaciones pretest. Obsérvese que Y_2 e Y_3 no son normalmente registradas en el mismo momento temporal: la primera puede registrarse al final de un curso escolar y la segunda al comienzo del siguiente, pero ambas en el mismo año natural. La hipótesis nula que se somete a prueba es que las puntuaciones de cambio $Y_2 - Y_1$ $Y_4 - Y_3$, son iguales. Este diseño se denomina por ello **diseño de ciclos institucionales** (Cook, Campbell & Peracchio, 1990) y su similitud con el diseño con grupo de control no equivalente es evidente. Además, incrementa su potencia estadística cuando incluye cohortes previas (o sea, incorpora mayor número de pretests) en la misma línea que el diseño con grupo de control no equivalente con doble pretest.

Tabla N° 4.19. Diseño de cohortes con pretest.

Cohortes	Asignación	Secuencia de tratamiento					
		Pret	Trat.	Post	Pret	Trat.	Post
Primera (Control)	NA	Y_1	-	Y_2			
Segunda (Experimental)	NA				Y_3	X	Y_4

Una presentación típica de resultados (con datos hipotéticos) es la que aparece en la tabla siguiente.

Tabla N° 4.20. Presentación de resultados del diseño de cohortes con pretest.

Cohorte	N	Pretest(1982)	Postest(1983)	Pretest(1983)	Postest(1984)
Cohorte 1982	100	6.10 (1.51)	6.72 (1.89)	--	--
Cohorte 1983	124	--	--	6.34 (1.46)	8.45 (2.09)

En cualquier caso, la principal amenaza del diseño es la historia, como en el diseño básico, aunque la incorporación de mayor número de pretests puede servir de gran ayuda. Si las puntuaciones de cambio entre todos los pretests son comparables (en este caso, las diferencias 6.72 - 6.10, 6.34 - 6.10 y 6.72 - 6.34), la amenaza de historia se reduce sensiblemente. Sin embargo, la amenaza de la historia puede también ser examinada si se puede programar un estudio con tres cohortes, tal y como se presenta en la tabla siguiente.

En el **diseño con tres cohortes**, Y_1 e Y_2 no se registran simultáneamente porque pertenecen al final de un curso/ciclo y comienzo del siguiente, respectivamente. Se trata, como el diseño anterior, de un ciclo institucional que se repite con la observación Y_3 e Y_4 y por esta razón se denomina diseño de ciclo institucional recurrente o diseño de retazos (Cook & Campbell, 1979; Judd & Kenny, 1981; Cook, Campbell & Peracchio, 1990). Como en el diseño con grupo de control no equivalente, el tratamiento puede demostrarse efectivo si Y_1 e Y_3 son mayores que Y_2 e Y_4 y si Y_2 e Y_4 (y también, por la misma razón, Y_1 e Y_3) no difieren entre sí.

Tabla N° 4.21. Diseño con tres cohortes.

Cohortes	Asignación	Secuencia de registros				
		Cohorte 1		Cohorte 2	Cohorte 3	
		Postest	Pretest	Tratamiento	Postest	Pretest
Primera	NA	(X)Y ₁				
Segunda	NA		Y ₂	X	Y ₃	
Tercera	NA					Y ₄ (X)

El control (parcial) de la amenaza de historia se obtiene si se cumple que $Y_3 > Y_2$, $Y_3 > Y_4$ y $Y_3 > Y_4$, puesto que con ello se demuestra que el tratamiento ha sido efectivo en dos momentos diferentes y que además ha tenido que ocurrir dos veces el evento histórico causante de la amenaza para explicar que $Y_1 > Y_2$ e $Y_3 > Y_4$.

Por supuesto, la amenaza de selección también puede descartarse en este diseño, porque las mismas personas están involucradas en algunas de las comparaciones (en particular, $Y_3 - Y_2$).

Sin embargo, la amenaza de la administración de pruebas no puede descartarse porque todas las comparaciones contrastan puntuaciones pretest (Y_2 e Y_4) con puntuaciones postest (Y_1 e Y_3). Por esta razón, en el trabajo original de Campbell & Stanley (1966) se recomendaba extender este diseño dividiendo la segunda cohorte en dos mitades aleatorias, una de las cuales recibe la secuencia pretest-tratamiento-postest mientras la otra recibe la misma secuencia sin pretest. El diagrama de este diseño aparece en la siguiente tabla.

Tabla N° 4.22. Extensión del diseño de tres cohortes.

Cohortes	Asignación	Secuencia de registro				
		Cohorte 1		Cohorte 2	Cohorte 3	
		Postest	Pretest	Tratamiento	Postest	Pretest
Primera	NA	(X)Y ₁				
Segunda (A)	A		Y ₂	X	Y ₃	
Segunda (B)	A			X	Y ₄	
Tercera	NA					Y ₅ (X)

En este caso, cualquier diferencia entre Y_3 e Y_4 sólo puede ser atribuida a la administración de las pruebas. Si la diferencia es nula, entonces la validez de la inferencia causal en este diseño es alta, y sólo puede quedar oscurecida si se emplean medidas poco fiables (o tamaños muestrales escasos) que directamente afecten a Y_2 , una medida involucrada en varias comparaciones.

4.3.2.3.3. Diseño de cohortes con tratamiento dividido.

En un estudio original de Minton (1975), una formulación alternativa para superar la amenaza de historia y fortalecer el diseño de cohortes básico podría ser dividir a los niños de la segunda cohorte (cohorte experimental) en función del grado de exposición a la experiencia objeto de evaluación (por ejemplo, si se trataba de asistente asiduo o asistente no asiduo) y efectuar la misma división en la primera cohorte (cohorte de control). Así, en ausencia de un efecto de tratamiento, no sería admisible pensar que existieran diferencias entre niños asiduos y no asiduos ya que ambos experimentan la misma historia.

Este diseño se denomina diseño de cohortes con tratamiento dividido y, suponiendo tres niveles de exposición al tratamiento, exposición fuerte (X_1), débil (X_2) y nula (X_3), se presenta en la tabla siguiente.

La división de los sujetos en grupos de tratamiento aporta otra ventaja a la validez interna. En el diseño básico, si las condiciones de administración de pruebas difieren entre la primera y sucesivas cohortes, la administración puede resultar una amenaza por sí misma. Dividir a los sujetos en función de la exposición al tratamiento contribuye a eliminar esta amenaza.

En general, los diseños de cohorte son una alternativa interesante en contextos institucionales (sobre todo, educativos) donde existen intervenciones cíclicas de algún tipo de tratamiento. Pueden alcanzar un alto nivel inferencial en la medida en que se acompañen de características tales como la incorporación de tratamientos divididos (y la consiguiente partición de las cohortes en subgrupos) y/o la ampliación con cohortes de control no equivalente.

Tabla N° 4.23. Diseño de cohortes con tratamiento dividido.

Cohortes	Asignación	Secuencia de registro		
		Pretest	Tratamiento	Postest
Subgrupo 1 Cohorte 1 (Control) Cohorte 2 (Experimental)	NA NA	Yc1	X ₁	Y _{E1}
Subgrupo 2 Cohorte 1 (Control) Cohorte 2 (Experimental)	NA NA	Yc2	X ₂	Y _{E2}
Subgrupo 3 Cohorte 1 (Control) Cohorte 2 (Experimental)	NA NA	Yc3	X ₃	Y _{E3}

4.3.2.4. Diseños de intercambio de tratamiento

El diseño cuasi-experimental de intercambio de tratamiento es una combinación adosada de un diseño con grupo de control no equivalente y un diseño solamente postest con el tratamiento permutado (Basadur, Graen y Scandura, 1986), tal y como se presenta en la tabla siguiente.

Tabla N° 4.24. Diseño de intercambio de tratamiento.

Grupos	Asignación	Secuencia de registro				
		Pretest	Tratamiento	Post/Pret	Tratamiento	Postest
Experimental	NA	Y _{E1}	X	Y _{E2} /Y _{C2}	X	Y _{C3}
Control	NA	Y _{C1}		Y _{C2} /Y _{E2}		Y _{E3}

Este diseño (*switching replications design*) parece una combinación porque, entre el primer y el segundo registro de medición un grupo sirve como experimental y el otro como control (de ahí su semejanza con el diseño con grupo de control no equivalente) mientras que entre el segundo y el tercer registro los papeles de experimental y control se invierten (y de ahí su semejanza con el diseño solamente postest).

4.3.2.5. Diseños con tratamiento invertido

Hackman, Pearce & Wolfe (1978) investigaron los efectos que los cambios en las propiedades motivacionales del trabajo tienen sobre las actitudes y la conducta en el trabajo. Como resultado de la innovación tecnológica, se transformó el trabajo administrativo típico de la banca para convertirlo en un trabajo más complejo y exigente para algunos sujetos (trabajo enriquecido) y menos complejo para otros sujetos (trabajo empobrecido) sin informar al personal de las posibles consecuencias motivacionales. Las puntuaciones pretest del grupo con trabajo enriquecido fueron inferiores a las del otro grupo, indicando una diferencia inicial en selección. En este trabajo se utilizó un diseño denominado diseño pretest-postest con tratamiento invertido (*reversed-treatment control group design with pretest and posttest*) y se ha representado en la siguiente tabla.

Tabla N° 4.25. Diseño con tratamiento invertido

Grupos	Asignación	Secuencia de registro		
		Pretest	Tratamiento	Postest
Experimental	NA	Y ₁	X+	Y ₂
Control	NA	Y ₁	X-	Y ₂

En este caso, X+ representa un tratamiento que se espera que produzca un efecto en dirección positiva y X- un tratamiento conceptualmente opuesto que se espera que invierta la pauta del efecto.

Este diseño tiene una ventaja especial de validez de constructo. El constructo causa tiene que ser rigurosamente operacionalizado para crear una prueba sensible que dependa de una versión de la causa (en el ejemplo, trabajo enriquecido) afectando a un grupo en un sentido, mientras su opuesto conceptual (trabajo empobrecido) afecta al otro grupo en sentido invertido.

Además de los problemas éticos que en ocasiones plantea la administración de un tratamiento invertido, un serio problema interpretativo aparece cuando la pauta de cambio es diferente entre tratamientos, pero en la misma dirección.

4.3.2.6 Diseños sin grupo control

Hay situaciones cuando resulta imposible contar con un grupo de control. Los de mayor interés son los diseños de grupo único y los diseños de línea-base no causal.

4.3.2.6.1. Diseños de grupo único

Se caracterizan por disponer de un solo grupo de unidades de respuesta. Como contrapartida, el investigador prolonga la secuencia de registro con la intención de encontrar algún equivalente funcional que compense la ausencia de un grupo comparativo. Los dos diseños de grupo único más importantes son el diseño de un solo grupo con tratamiento y control y el diseño de tratamiento repetido.

Diseño de un solo grupo con tratamiento y control.

Este diseño se utiliza cuando no es factible obtener un grupo de control no equivalente y el investigador desea obtener un equivalente funcional. Una forma típica es la que aparece en la tabla.

Tabla N° 4.26. Diseño de grupo con tratamiento y control más típico.

Grupos	Asignación	Secuencia de registro						
		Pret	Trat	Pret/Post	Trat	Pre/Post	Trat	Post
Único	NE	Y,	X	Y2	X	Y3	X	Y4

Se caracteriza porque utiliza un solo grupo (muestra) pre-post, repetido en dos ocasiones (Y, - Y2 e Y2 - Y3) tras el cual se convierte en un grupo de control (Y3 - Y4). En general, este diseño plantea diferentes problemas a los que es preciso atender para producir inferencias válidas:

Se requiere utilizar medidas con alta fiabilidad y grandes tamaños muestrales. Muchos tratamientos son intervenciones sociales de carácter benéfico y su eliminación puede plantear problemas éticos. Además, suele plantear una tasa de mortalidad elevada.

Es conveniente que las observaciones se registren a intervalos igualmente espaciados. Se consigue así un control para cambios lineales espontáneos que tienen lugar en un determinado espacio de tiempo. Una simple comparación de las diferencias Y3 - Y2 e Y4 - Y3 no tendría sentido si el intervalo temporal entre Y3 e Y2 fuera diferente del intervalo entre Y4 e Y3.

Diseño de tratamiento repetido.

Cuando el investigador sólo tiene acceso a una población, será posible en ocasiones introducir, eliminar y después volver a introducir el tratamiento en fechas diferentes. Esta situación es únicamente viable cuando el efecto inicial de tratamiento tiene carácter transitorio. Un diseño típico de esta naturaleza es muy similar al anterior y se aborda en el la tabla siguiente.

El resultado más interpretable se produce cuando Y, difiere de Y₂, Y₃ difiere de Y₄ y la diferencia Y₃ - Y₄ se presenta en la misma dirección que la diferencia Y, - Y₂. El diseño tiene muchas características en común con los diseños de N = 1 del análisis experimental de la conducta.

Tabla N° 4.27. Diseño de tratamiento repetido.

Grupos	Asignación	Secuencia de registro						
		Pret	Trat	Pret/Post	Trat	Pre/Post	Trat	Pos
Único	NE	Y,	X	Y2	-	Y3	X	Y4

Entre los problemas de validez interna destacan las amenazas de maduración e historia. Pero los problemas fundamentales proceden de la sensibilidad (porque suele ser muy escaso el número de sujetos empleado) y de la validez de constructo (derivadas de la introducción y retirada del tratamiento).

4.3.2.6.2. Diseños de línea-base no causal

Cuando no se dispone de grupo control y se desea desarrollar algún equivalente funcional que sirva de línea-base no causal, es posible elegir una de dos alternativas para evaluar los efectos de un tratamiento: el diseño de extrapolación de la regresión y el diseño de control baremado.

Diseño de extrapolación de la regresión.

Cook et al. (1975) evaluaron las mejoras en aprendizaje de niños que veían “Plaza Sésamo” en varias áreas de EEUU. Tras administrar un pretest, se examinó el diagrama de dispersión de la edad (en meses) contra las puntuaciones obtenidas en el pretest. Las rectas de regresión por áreas eran fundamentalmente lineales y los investigadores calcularon una ecuación de regresión para determinar cuánta mejora podía esperarse por mes de edad en cada área.

El diseño resultante se denomina por ello diseño de extrapolación de la regresión y consiste en comparar la puntuación de un grupo de tratamiento en el posttest con su puntuación pronosticada teniendo en cuenta las amenazas contra, la validez interna la más importante de tales amenazas, aunque no la única, es la maduración.

El diseño puede utilizarse también en una situación en la que existan múltiples predictores, en lugar de un único predictor, como en el ejemplo. En uno y en otro caso, los datos y el fenómeno objeto de estudio han de presentar cierta estabilidad a lo largo del tiempo para que la predicción de regresión sea lo más exacta posible.

Diseño de control baremado.

Este diseño se utiliza también, particularmente en contextos educativos, cuando no se dispone de un grupo de control. Siguiendo a Tallmadge (1982), uno de sus principales proponentes, la ejecución obtenida en el/los grupos experimental/es en pretest y posttest se expresa en función de baremos de una población similar a la estudiada. Por ejemplo, supóngase que la puntuación media en un pretest de lenguaje para un grupo de tratamiento equivale al percentil 50 de un test estandarizado. Después de ser sometido a un programa de mejora de vocabulario, la puntuación que se obtiene en el posttest equivale al percentil 55 del mismo test.

La hipótesis es que la diferencia de 5 puntos de percentil se debe al programa de mejora que, en esencia, ha sido contrastado contra los baremos nacionales durante el período pretest-posttest, y constituye el efecto de tratamiento.

El diseño de control baremado parte de dos supuestos cruciales. El primero es que, cuando no hay efecto de tratamiento, los percentiles de pretest y posttest son iguales. El segundo es que debe disponerse de datos de una población baremada que sea similar a la población bajo estudio y cuyas edades cubran el mismo rango, un requisito que se cumple con frecuencia en el campo educativo.

Los principales problemas que plantea este diseño se refieren a instrumentación, regresión estadística y selección.

Respecto a la primera, el modelo supone que un cambio en percentil se relaciona sistemáticamente con un cambio en las puntuaciones directas, y por tanto el error aleatorio puede tener consecuencias diferentes en puntos distintos de la escala original.

La regresión estadística puede sesgar el diseño de control baremado cuando la prueba de selección para entrada al programa se acerca a las puntuaciones pretest o postest. Sin embargo, es la selección la amenaza más problemática, porque se supone que la población bajo estudio y la población baremada son equivalentes en composición y tasa de cambio. Pero muy probablemente las poblaciones utilizadas para la baremación no son equivalentes.

Cuando no se dispone de normas publicadas, los investigadores pueden incluso construir grupos de control *ad hoc* con datos de archivo que se registraron con otros propósitos. Jackson y Mohr (1986), en su análisis de los efectos de un programa experimental de subvención de viviendas, utilizaron un grupo de control a partir de los datos registrados durante los dos años anteriores de la encuesta anual de vivienda que se administró a toda la población.

4.3.2.7. Diseños con variables dependientes no equivalentes

Un diseño de dudosa interpretabilidad, pero utilizado en ocasiones, es un diseño con grupo de control no equivalente que utiliza diferentes medidas para grupo experimental y grupo de control. Llamando A y B a las medidas diferentes para cada grupo, el diagrama de este diseño es el que aparece en la tabla siguiente.

Tabla N° 4.28. Diseño con grupo control no equivalente.

Grupos	Asignación	Secuencia de registro		
		Pretest	Tratamiento	Postest
Experimental	NA	Y _{1A}	X	Y _{2A}
Control	NA	Y _{1B}	-	Y _{2B}

Este diseño precisa para su consideración variables medidas con alta fiabilidad y una potente teoría sustantiva que justifique la utilización de los constructos. Las amenazas más importantes suelen especificarse en términos de diferencias relativas a los constructos en tasas de maduración, historia o instrumentación.

4.3.2.8. Diseños con variable de asignación conocida

Cuando la variable de asignación no es aleatoria, pero sí conocida, la rigurosidad del diseño aumenta considerablemente y se aproxima a la de un diseño experimental auténtico. El diseño típico es el diseño de **discontinuidad de la regresión**, un diseño que goza de gran prestigio, pero que es escasamente utilizado en la práctica (Trochim & Cappelleri, 1992; Cook & Shadish, 1994).

La característica básica del diseño de discontinuidad de la regresión es que la asignación, aunque no es aleatoria, sí es al menos conocida. Suelen haber dos

grupos, uno experimental (GE) y el otro control (GC). El investigador conoce exactamente bajo qué condiciones los sujetos se asignan a los dos grupos.

La estructura de este diseño es, desde este punto de vista, muy similar a la del diseño con grupo de control no equivalente. En la notación que se presenta aquí, la distinción esencial es que el investigador no sabe a qué grupo será asignado un sujeto hasta no conocer su puntuación en el pretest, que se utiliza – generalmente- como variable de asignación. Por ello, la única diferencia con el grupo de control no equivalente es la naturaleza de la variable de asignación.

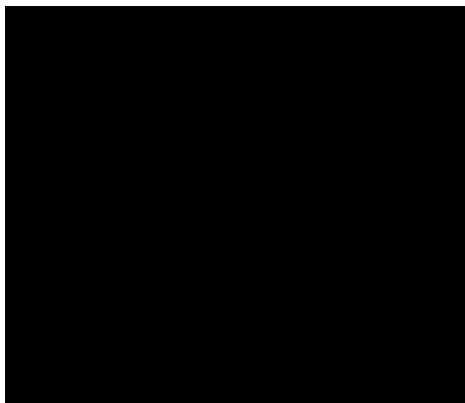
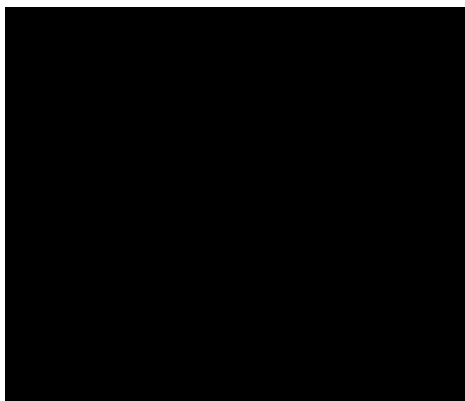
Imagine una situación en la que cada sujeto de una amplia muestra de escolares de enseñanza primaria pueda ser clasificado de acuerdo con su puntuación en un continuo generado por la administración de una prueba de rendimiento en solución de problemas (pretest) y que la institución ha determinado la existencia de un punto de corte por encima del cual los sujetos pueden mejorar notablemente con ayuda de un determinado programa y por debajo del cual la mejora que se produce es prácticamente insignificante, no mereciendo la pena el esfuerzo económico que se precisa.

Tabla N° 4.29. Diseño con variable de asignación conocida.

Grupos	Asignación	Secuencia de registro		
		Pretest	Tratamiento	Postest
Experimental	NAC	YEi	X	YEZ
Control	NAC	Yci	-	YE2

Sobre estos supuestos, los sujetos que obtienen puntuaciones en la prueba por encima del punto de corte recibirán el tratamiento (por ejemplo, son sometidos a un intenso entrenamiento durante un mes con el objeto de prepararse para participar en unas pruebas de rendimiento a nivel nacional) mientras que los que obtienen puntuaciones en la prueba por debajo del punto de corte no reciben tal tratamiento (por ejemplo, siguen el curso habitual de entrenamiento hasta entonces recibido).

Supóngase que se administra de nuevo la misma prueba de rendimiento una semana antes de participar en las pruebas nacionales a todos los sujetos inicialmente probados (postest). Se construye a continuación un diagrama bidimensional tomando las puntuaciones del pretest en la abscisa, fijando también la puntuación que sirvió de punto de corte, las puntuaciones del postest en la ordenada. Si el tratamiento no resulta efectivo, la nube de puntos aparecería de forma ordenada, y las rectas de regresión para cada grupo no revelarían ninguna



discontinuidad en el punto de corte (Figura derecha superior). Si, por el contrario el tratamiento resulta efectivo, la nube de puntos se presentará de forma desordenada, y las rectas de regresión para cada grupo revelarán la existencia de una discontinuidad sobre el punto de corte (Figura derecha inferior).

El diseño de discontinuidad de la regresión es útil cuando se desea estudiar un programa que se administra sobre la base de alguna necesidad o mérito. Así sucede, por ejemplo, en el caso de la educación compensatoria donde los niños con mayor necesidad de instrucción adicional (medida en algún pretest) reciben servicios compensatorios del Gobierno, o en el caso de la selección de deportistas para su preparación en algún campeonato. La presentación de resultados sigue las pautas anteriormente apuntadas. Un ejemplo aparece en la tabla siguiente.

Tabla N° 4.30. Presentación de resultados del diseño de discontinuidad de regresión.

<i>Grupos</i>	<i>N</i>	<i>Pretest</i>	<i>Postest</i>
GE	246	7.49 (2.36)	9.61 (3.54)
GC	282	6.03 (1.95)	7.58 (3.57)

La historia del diseño de discontinuidad de la regresión ha discurrido en torno a dos tradiciones paralelas (Trochim, 1984). La primera es la tradición académica, la cual describe el desarrollo técnico que ha sufrido el diseño⁶⁷; la segunda es la tradición aplicada y describe el uso del diseño para evaluar programas de intervención de diversa naturaleza⁶⁸.

4.3.2.9. Diseños de series temporales interrumpidas

Los diseños de series temporales interrumpidas consisten en una serie de observaciones y una intervención aplicada en un determinado punto del tiempo. Dicha intervención interrumpe la serie y el objetivo es hallar el patrón de cambio entre el período pre-tratamiento y post-tratamiento. En los últimos años, estos diseños se muestran muy apropiados para evaluar programas sociales o comunitarios (Galster et al, 2004), intervenciones educativas (Bloom, 2003),

⁶⁷ El diseño de regresión discontinua fue inicialmente propuesto por Thistlethwaite y Campbell (1960) más como alternativa analítica del clásico diseño pretest-postest que como un diseño propiamente dicho. Campbell lo llamó por primera vez diseño en un artículo de 1969. Posteriormente fue tratado en profundidad en varios artículos (Rubin, 1977; Boruch y Gómez, 1977; Visser & De Leeuw, 1984) y textos de metodología avanzados (Cook y Campbell, 1979; Judd y Kenny, 1981; Visser, 1985) y de forma monográfica en un texto de Trochim (1984). La aceptación actual de este diseño es elevada: Mósteller (1990) lo acepta como un diseño experimental auténtico y Rubin () proporciona pruebas fehacientes que permite obtener estimaciones insesgadas de los efectos de tratamiento.

⁶⁸ Pese a su excelente consideración académica, el diseño es escasamente utilizado. Cook y Shadish (1994) argumentan entre las razones, además de motivos menores de orden profesional, el que la asignación a tratamientos no siempre puede hacerse de acuerdo con reglas estrictas y precisa de múltiples criterios, y la dificultad añadida que presenta el análisis estadístico. El principal contexto de aplicación del diseño ha sido la educación compensatoria. Una extensa literatura, analizada parcialmente en el texto de Trochim (1984) y en Cook & Shadish (1994), muestra la consideración que en este área ha recibido el diseño. Otras áreas en las que se ha aplicado son la justicia criminal (Berk & Rauma, 1983), los servicios de salud (Lohr, 1972, citado en Cook y Campbell, 1979) y otras temáticas menores (Seaver & Quarton, 1976).

efectos de leyes o regulaciones legales (Sverdrup, 2003; White, 2003) o beneficios terapéuticos (Campbell, 1996).

En la actualidad se considera a los diseños de series temporales interrumpidas como uno de los más poderosos diseños cuasi-experimentales. Sin embargo, de su empleo se derivan tres problemas fundamentales que ningún investigador debiera pasar por alto a la hora extraer inferencias:

1. Al carecer de un principio básico similar al acto físico de la aleatorización no existe un adecuado grado de control sobre algunas de las principales amenazas que atentan contra la validez interna y, muy en especial, de la amenaza “historia”. Consecuentemente, se puede formular la hipótesis rival de que el efecto del tratamiento puede deberse a la acción de otros eventos que han ocurrido al mismo tiempo y que son de hecho los verdaderos responsables de los cambios observados. Como es obvio, la posibilidad de que esta explicación rival prospere dependerá, en buena medida, de lo cuidadoso que sea el evaluador a la hora de verificar la existencia de factores extraños, de lo complejo que sea el diseño que se elija y del número de puntuaciones que efectúe.
2. Los procedimientos estadísticos tradicionales, tales como la prueba “t de Student” o el análisis de la varianza de Fisher, han jugado un importante papel durante bastantes décadas a la hora de estimar y probar cambios entre las medias de diferentes grupos. Sin embargo, estas pruebas solamente son válidas si las observaciones registradas con anterioridad y con posterioridad al evento de interés varían en torno a las medias de las respectivas fases no sólo, normalmente y con varianza constante, sino también independientemente (Box & Tiao, 1975). Ahora bien, por lo general, los datos registrados sucesivamente a lo largo del tiempo carecen de la gracia que habitualmente confiere la aleatorización, y son usualmente dependientes y frecuentemente no estacionarios. Consecuentemente, todos aquellos procedimientos estadísticos, tanto paramétricos como no paramétricos, que requieren para su correcta aplicación el supuesto de independencia no deberían emplearse, pues la presencia de autocorrelación puede distorsionar sustancialmente los resultados de las pruebas que no lo tienen en cuenta⁶⁹. A raíz de las críticas surgidas con la aplicación rutinaria de las

⁶⁹ Durante bastantes años se ha operado como si la presencia de dependencia serial encontrada en los diseños de series temporales interrumpidas sólo tuviera implicaciones negativas para los análisis estadísticos convencionales, sobre todo, a raíz de que Scheffé (1959) pusiera de relieve cómo la presencia de correlación serial positiva convertía a la prueba de F en excesivamente liberal, mientras que la presencia de correlación serial negativa la volvía excesivamente conservadora, pero no para los clásicos análisis visuales, pues existía la creencia de que éstos eran más conservadores que los análisis estadísticos y, por ende, los analistas sólo responden a efectos de gran tamaño. No obstante, Matyas y Greenwood (1990), tras llevar a cabo varios experimentos y una exhaustiva revisión de la literatura existente en torno a la técnica de análisis visual, presentan datos originales en los que demuestran cómo la técnica del trauma ocular, alias acertadamente empleado por Kazdin (1984), además de no ser fiable, es en exceso liberal. En concreto, los autores citados descubrieron que los analistas visuales reivindicaban en numerosas ocasiones intervenciones significativas cuando de hecho no se habían producido; por el contrario, raramente fallaban en detectar efectos

técnicas visuales y con los intentos de aplicación de las pruebas estadísticas convencionales, a lo largo de las dos últimas décadas han aparecido diversos métodos estadísticos tendientes a paliar los problemas reseñados⁷⁰.

3. El problema de la inducción y de la evaluación de la generalidad. Ante la pregunta ¿hasta que punto los resultados experimentales obtenidos con un único sujeto son representativos de los logrados con otros sujetos? Cowles (1989) afirma que la representatividad es un asunto de carácter conductual más que lógico y, por tanto, es un problema de constatar hechos; es decir, se trataría de añadir vigor al rigor mediante algún programa de replicación sistemática. Pues bien, en vez de efectuar replicas adicionales mediante diseños de series temporales simples se puede utilizar diseños transversales, como por ejemplo, la serie temporal interrumpida con replicaciones intercambiables (véase el punto 4.3.2.9.5.)

A continuación se describen los diseños de series temporales interrumpidas más significativos.

4.3.2.9.1. Diseño simple

Las series temporales interrumpidas pueden entenderse como una mejora del diseño pre-experimental con pretest y posttest. La mejora implica efectuar múltiples replicaciones en el tiempo con anterioridad y con posterioridad a la introducción del programa o intervención.

El diseño de series temporales interrumpidas más básico o diseño simple de series temporales interrumpidas puede ser esquematizado así:

verdaderos, aunque el tamaño de tales efectos fuese relativamente modesto. Más aún, si la autocorrelación está presente en la serie temporal, los problemas con los que usualmente se encuentran los partidarios de los análisis visuales no sólo no permanecen estables, sino que la evidencia existente se ha encargado de poner de manifiesto que, generalmente, éstos se suelen acentuar (Matyas y Greenwood, 1990, 1991). Por consiguiente, este descubrimiento debilita enormemente la postura de aquellos investigadores que abogan por utilizar rutinariamente la técnica visual a la hora de evaluar el impacto ocasionado por una intervención planificada.

⁷⁰ Con todo, debemos manifestar que la solución más prometedora y también más practicada en el campo de las ciencias socio-comportamentales ha consistido en la adaptación efectuada por Glass, Willson y Gottman (1975) de la técnica de las series temporales, desarrollada inicialmente por Box y Tiao (1965) y Box y Jenkins (1976). Este enfoque se basa en la adaptación e integración dentro de una teoría comprensiva del análisis espectral utilizado en las ciencias físicas con datos de carácter continuo al análisis de datos de corte longitudinal, pero de carácter discreto. En los trabajos de estos autores, además de presentarse las aportaciones más novedosas en torno al tratamiento estocástico de las series temporales, se propone una metodología que permite llegar a modelar adecuadamente la estructura que sigue la parte sistemática (parte responsable de la dependencia serial) del componente estocástico de la serie bajo estudio. Modelamiento que se encuadra dentro de una clase paramétrica de procesos estocásticos lineales y discretos formados por los denominados autorregresivos, integrados y de medias móviles; procesos que reciben el nombre genérico de modelos ARIMA.

Tabla N° 4.31. Serie de tiempo interrumpida de diseño simple

<i>Grupos</i>	<i>Asignación</i>	<i>Secuencia de registro</i>
Experimental	Natural	O1 O2 O3 O4 O5 X O6 O7 O8 O9 O10

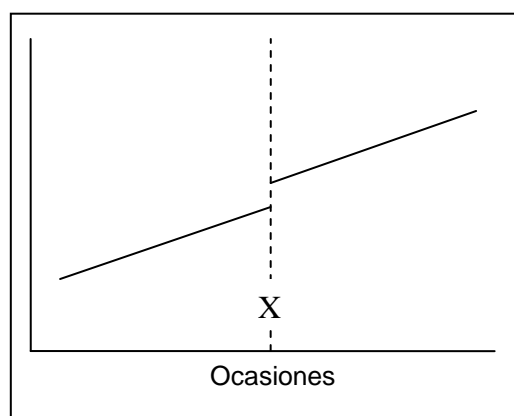
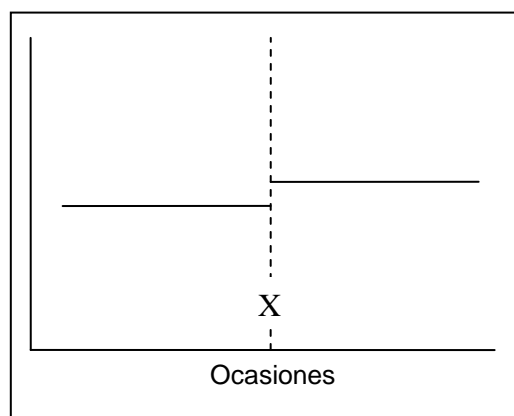
Donde las observaciones anteriores a X son pretest y las observaciones posteriores al tratamiento son el postest. En este diseño la inferencia sobre la efectividad del programa se basa en comprobar si existen cambios en el nivel de la serie entre las fases pre y post programa, cambios en la dirección de las tendencias, o cambios en ambos patrones conjuntos. Este diseño nos permite analizar si los cambios son demorados, temporales o permanentes; además, si son abruptos o graduales.

En efecto, este diseño tiene la ventaja de detectar con mayor facilidad la posible presencia de efectos maduracionales e inclusive evaluar el papel de las variaciones estacionales que en ocasiones se confunde con los efectos del programa. Sin embargo, a pesar de lo dicho, existen algunas amenazas que afectan la validez interna de este diseño: a) la instrumentación, b) la selección, c) la regresión estadística y d) la historia.

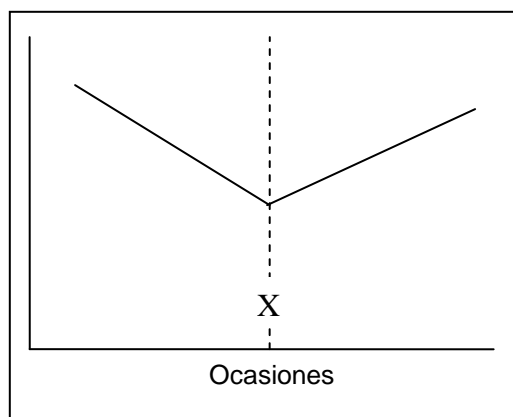
Las tres primeras amenazas se pueden controlar utilizando procedimientos estandarizados de recolección de datos, contando con observaciones muestrales inalteradas y si las medidas son numerosas y no muy espaciadas. Sin embargo, la amenaza más seria es la historia, la cual sólo puede ser mitigada en este diseño si se mantiene un registro de sucesos teóricamente relacionados a los efectos del programa (Cook & Campbell, 1979; Cook et al, 1990).

En el diseño simple es posible encontrar, mediante el análisis visual de los valores, cuatro soluciones típicas (Anguera et al, 1996):

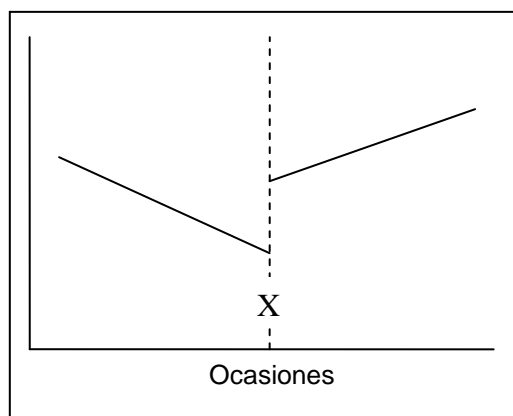
1. Situaciones donde las tendencias están ausentes, si bien existe cambio de nivel. Es decir, a partir del programa se ha incrementado el nivel de la media, pero no existe ninguna tendencia creciente o decreciente.
2. Situaciones en las que existe un cambio de nivel y tendencia, pero que no exhiben cambio alguno entre las fases pre y post programa. En este caso, la cuestión a resolver es si la aplicación del programa produce algún impacto en los beneficiarios, al margen de la curva evolutiva que siguen los datos.



3. Situaciones en las cuales no existe cambio de nivel, la tendencia es distinta de cero y existe un cambio en la orientación de esta. En este caso, a diferencia de lo ocurrido en la situación anterior, no existe interrupción entre las dos fases del diseño, si bien existen patrones de cambio que pueden manifestar una variación en la dirección del mismo, bien sea en el mismo sentido o bien sea en sentido opuesto, con una mayor o menor inflexión en la tendencia dependiendo de la significación del programa.



4. Situaciones en las que existe un cambio de nivel, la tendencia es distinta de cero y se da un cambio en la dirección de ésta a través de las fases pre y post-programa. En este diseño se debe verificar no sólo si la presentación del programa produce una discontinuidad en el punto de intervención de la serie temporal, sino también si como consecuencia de la introducción del impacto se produce interacción programa x tendencia, o si se prefiere algún cambio en el sentido u orientación de las tendencias a través de las fases.



4.3.2.9.2. Diseño con grupo control no equivalente

Este diseño es superior al diseño simple anteriormente analizado, pues incluye un grupo control o un estado contrafactual.

Tabla N° 4.32. Serie de tiempo interrumpida con grupo control

<i>Grupos</i>	<i>Asignación</i>	<i>Secuencia de registro</i>
Experimental	Natural	O ₁ O ₂ O ₃ O ₄ O ₅ X O ₆ O ₇ O ₈ O ₉ O ₁₀
Control	Natural	O ₁ O ₂ O ₃ O ₄ O ₅ X O ₆ O ₇ O ₈ O ₉ O ₁₀

A diferencia del diseño simple, este permite controlar más amenazas contra la validez interna: maduración, instrumentación, medición, regresión a la media y, principalmente, historia (Cook & Campbell, 1979). Los efectos de la historia se controlan gracias a que cualquier evento o circunstancia ajena al programa que pueda alterar los resultados del grupo beneficiado, deberá también existir en el grupo control. Sin embargo, a pesar de lo dicho, este diseño no es invulnerable a la

interacción entre selección con maduración; es decir, que en virtud de la selección de grupos, la maduración pueda ser mayor en un grupo que en otro.

Los análisis pueden realizarse visualmente (de forma exploratoria) y luego estadísticamente. En este caso puede aplicarse el enfoque de la regresión (Manly, 1992), incluyendo a la serie temporal correspondiente al grupo control dentro del modelo regresivo como una variable predictora más.

4.3.2.9.3. Diseño con variables dependientes no equivalentes

Tal como se mencionó previamente, la principal amenaza contra el diseño de series temporales interrumpidas es la historia. Se ha dicho ya que dicha amenaza puede controlarse disminuyendo el tiempo entre los sucesivos registros y, sobre todo, complicando el diseño a través de la incorporación de un grupo control (véase el diseño anterior). Sin embargo, existe otra forma de controlar tal amenaza: controlando el impacto del programa en dos mediciones dependientes no equivalentes.

Tabla N° 4.33. Serie de tiempo interrumpida con variables dependientes no equivalentes

<i>Grupos</i>	<i>Asignación</i>	<i>Secuencia de registro</i>
Experimental	Natural	O _{A1} O _{A2} O _{A3} O _{A4} O _{A5} X O _{A6} O _{A7} O _{A8} O _{A9} O _{A10} O _{B1} O _{B2} O _{B3} O _{B4} O _{B5} X O _{B6} O _{B7} O _{B8} O _{B9} O _{B10}

En este caso, A y B representan medidas diferentes registradas en el mismo grupo y afectadas diferencialmente por el tratamiento. Debe advertirse que B no es una covariante, sino una variable independiente del de los efectos del programa. Así, mientras se espera que A cambie producto del impacto del programa, en B se espera que esta no tenga ningún cambio. Si se encontrase alguna correlación temporal entre A y B, entonces ese sería el efecto de la historia.

4.3.2.9.4. Diseño con replicaciones múltiples

Puede ocurrir que el investigador esté interesado en el análisis de múltiples intervenciones (bien sean estas debidas a la introducción, retirada, reintroducción y así sucesivamente de un tratamiento o bien sean debidas a diversas condiciones de tratamientos consecutivos). En este caso, el diseño consiste en cuatro o más periodos de observación repetidos, de un solo grupo poblacional intacto en donde se aplica un programa (X), se lo retira después (noX), y se implementa posteriormente (X), y así sucesivamente.

Tabla N° 4.34. Serie de tiempo interrumpida con replicaciones múltiples

<i>Grupos</i>	<i>Asignación</i>	<i>Secuencia de registro</i>
Experimental	Natural	O ₁ O ₂ O ₃ O ₄ O ₅ X O ₆ O ₇ O ₈ O ₉ O ₁₀ noX O ₁₁ O ₁₂ O ₁₃ O ₁₄ O ₁₅ X O ₁₆ O ₁₇ O ₁₈ O ₁₉ O ₂₀

Dado que dos fases del diseño coinciden con la presentación del programa, existen tres oportunidades para demostrar su efectividad:

1. Introducción del programa (Observaciones 1 al 10).
2. Retiro del programa (Observaciones 6 al 15).
3. Reintroducción del programa (Observaciones del 11 al 20).

Por tanto, este diseño es superior al diseño simple de series temporales interrumpidas; pues garantiza más la validez interna al controlar los efectos de la historia. Sin embargo, aún persisten algunas limitaciones: En primer lugar, este diseño sólo puede ser aplicado cuando el evaluador puede asumir que los efectos del programa son provisionales y, por tanto, se disipan en el tiempo; así pues, este diseño está condicionado a la reversibilidad de los efectos del programa. En segundo lugar, se requiere cierto grado de control sobre las respuestas de los beneficiarios, para evitar las amenazas de “desmoralización de los sujetos” cuando se les retira el programa (Anguera et al, 1995).

4.3.2.9.5. Diseño con replicaciones intercambiables

En este diseño, existen dos grupos que reciben el mismo programa en dos tiempos diferentes. Así, cuando un grupo recibe tratamiento, el otro sirve de control y viceversa.

Tabla N° 4.35. Serie de tiempo interrumpida con replicaciones intercambiables

<i>Grupos</i>	<i>Asignación</i>	<i>Secuencia de registro</i>
Experimental	Natural	O ₁ O ₂ O ₃ O ₄ O ₅ X O ₆ O ₇ O ₈ O ₉ O ₁₀
Experimental	Natural	O ₁ O ₂ O ₃ X O ₄ O ₅ O ₆ O ₇ O ₈ O ₉ O ₁₀

Existen varias razones para emplear este diseño (Cook et al, 1990; Anguera et al, 1995), entre ellas:

- Porque controla muchas de las amenazas que atentan contra la validez interna, sobre todo, de los efectos de la historia; pues al presentarse el tratamiento a través de los grupos en momentos diferentes es factible rechazar la hipótesis de que los cambios debidos a los efectos de la historia se confunden con los derivados de la introducción del tratamiento (a no ser que dichos efectos operen en distintos ambientes y momentos temporales).
- Porque se ve potenciada la validez interna y reforzada la validez externa, por causa de que el efecto del tratamiento puede observarse en dos muestras distintas en al menos dos lugares y tiempos distintos.
- Porque es útil para detectar impactos, que si bien son de cierta intensidad sin embargo su acción tan sólo se manifiesta tras un periodo de tiempo más o menos prolongado.

Sin embargo, a pesar de estas ventajas, aún es posible que existan amenazas contra la validez como la interacción entre selección con historia.

4.4. SIMULACIONES EXPERIMENTALES

La mejor forma de demostrar la importancia de los criterios de validez en el diseño de la evaluación de impacto de los programas sociales, es mediante la simulación experimental estadística. Utilizando el software estadístico MINITAB 14, el método que se seguirá corresponde a una demostración estadística de los efectos perniciosos de la falta de control de algunas amenazas contra la validez de la inferencia causal. Se utilizará el modelo lineal general (análisis de regresión y derivados), por ser el más flexible para estos casos (Maxwell & Delaney, 1990; Pedhazur, 1982).

4.4.1. El modelo ideal

Se dijo que los diseños cuasi-experimentales son aplicaciones intermedias entre el diseño experimental y el pre-experimental. Al primero se lo definió como un estudio en el cual se tiene, por lo menos, dos grupos (experimental y control), los cuales han sido asignados aleatoriamente, y con medidas pretest y posttest. Al último (pre-experimental) se lo definió como un estudio sin grupo control o sin medidas pre-post. El diseño cuasi-experimental, por tanto, es un estudio que se caracteriza por tener, al menos dos grupos, pero sin asignación aleatoria (que puede ser desconocida o conocida) y con medidas pretest y posttest.

Por tanto, con los diseños experimentales se controla las diferencias entre grupos, producto de variables extrañas al programa, ya que la asignación aleatoria disipa esas diferencias en equivalencias indistintas en el pretest. De lo dicho, cualquier diferencia en las medidas posttest de los grupos es producto del impacto del programa. Este es pues el modelo ideal, el diseño experimental, en donde la aleatorización es la norma y en donde se puede estar seguro de la equivalencia inicial de los grupos control y experimental. Para demostrar lo dicho, se simulará un diseño experimental clásico pretest-posttest con grupo control.

```
MTB > random 1000 c1;
SUBC> normal 50 5.
MTB > random 1000 c2;
SUBC> normal 0 5.
MTB > random 1000 c3;
SUBC> normal 0 5.
MTB > add c1 c2 c4
MTB > add c1 c3 c5
MTB > name c1='puntaje verdadero' c2='error pretest' c3='error
posttest' c4='pretest' c5='posttest'
MTB > set c6
DATA> 1:1000
DATA> end
MTB > code (1:500) 0 c6 c6
MTB > code (501:1000) 1 c6 c6
MTB > tabl c6
Rows: grupo
      Count
```

```

0          500
1          500
All       1000

MTB > sign c6
0 Negative values    500 Zero values    500 Positive values
MTB > tabla c6;
SUBC> means c4 c5.
Rows: grupo
      pretest  postest
      Mean    Mean
0      50.50   50.45
1      49.70   49.48
All    50.10   49.96

```

En este caso, se han creado 1000 observaciones con un promedio teórico de 50 y una desviación estándar de 5. A estas observaciones se les ha agregado un error de medición (siempre existentes en la realidad) con un promedio de 0 y una desviación estándar de 5. La suma de estas variables ha configurado una puntuación de pretest y postest. Las observaciones han sido divididas en dos grupos (control = 0 y experimental=1), ambos grupos son aleatorios y, por tanto no tienen diferencias significativas entre ellos, es decir, son equivalentes. Además, el programa no tiene ningún impacto teórico en el postest. La siguiente figura demuestra lo dicho.

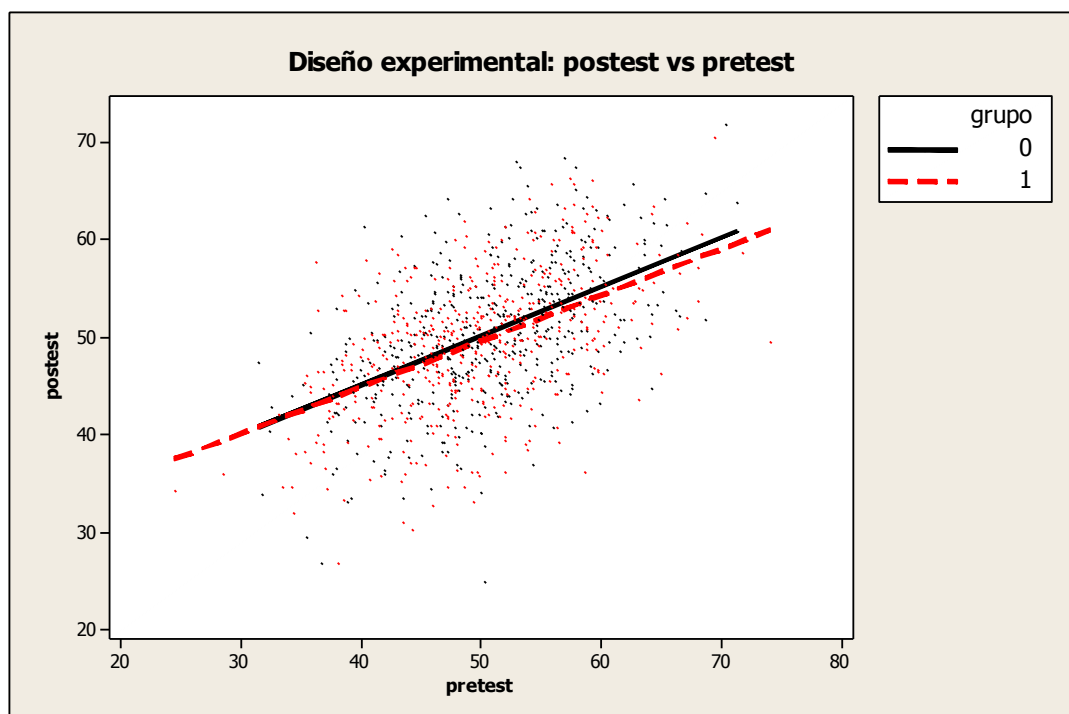


Figura 4.7. Relación entre postest y pretest cuando no existe impacto alguno del programa (Fuente: Elaboración propia).

Tal como se aprecia en la Figura 4.7, tanto el pretest como el postest tienen una distribución aleatoria entre los grupos experimental y control, no existiendo ninguna diferencia entre ellos. Este es el caso teórico en que el programa no ha sido efectivo en absoluto.

Ahora, supongamos que se ha aplicado el programa y después de un año se ha medido el postest, con un impacto teórico de 10 puntos a favor del grupo experimental.

```

MTB > let c7=c5+(10*c6)
MTB > name c7='postest 2'
MTB > table c6;
SUBC> mean c4 c7.
Rows: grupo
      pretest  postest 2
      Mean      Mean
0      50.50     50.45
1      49.70     59.48
All     50.10     54.96
MTB > Plot 'postest 2'*'pretest';
SUBC> Symbol 'grupo';
SUBC> Regress 'grupo';

```

Como consecuencia de la ecuación, la diferencia de 10 puntos a favor del grupo experimental se demuestra tanto en las estadísticas promedio como en la Figura 4.8.

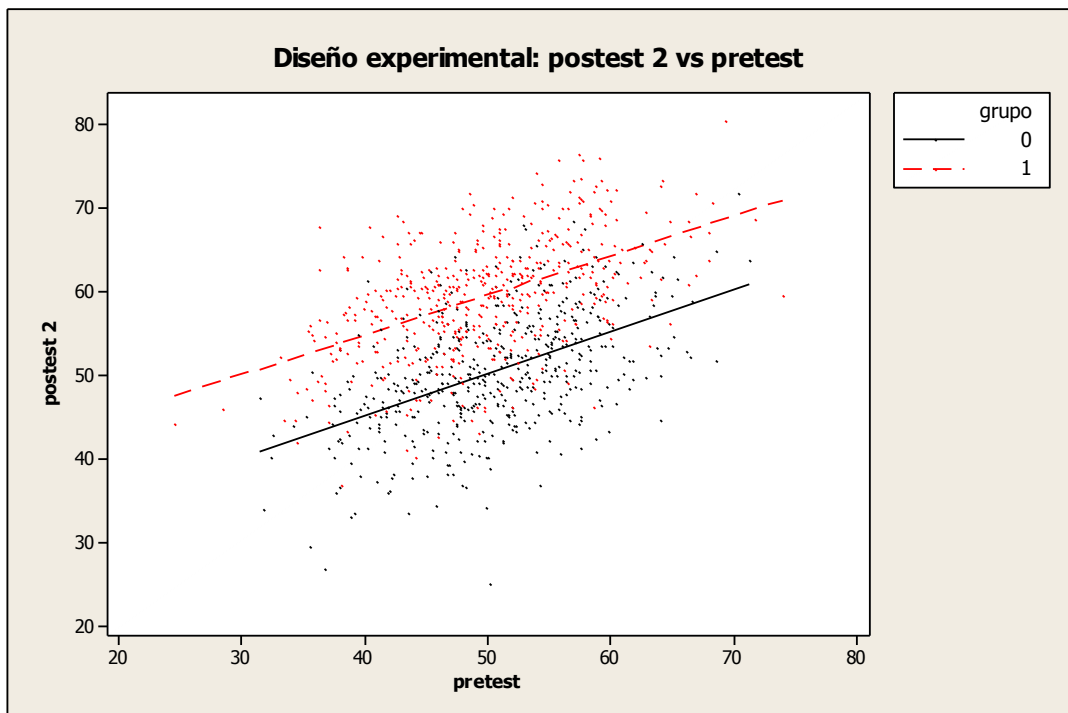


Figura 4.8. Relación entre postest y pretest cuando existe impacto del programa en 10 puntos (Fuente: Elaboración propia).

En este caso, se observa con claridad dos rectas de regresión, en la cual la del grupo experimental se muestra mayor en puntuación posttest. Al ser equivalentes los dos grupos (debido a su asignación aleatoria), el efecto está limpio de sesgos de selección. Para demostrarlo se realizará una ecuación de regresión.

Ecuación a

MTB > regress c7 1 c6

The regression equation is
 postest 2 = 50.5 + 9.02 grupo

Predictor	Coef	SE Coef	T	P
Constant	50.4516	0.3136	160.88	0.000
grupo	9.0239	0.4435	20.35	0.000

S = 7.01242 R-Sq = 29.3% R-Sq(adj) = 29.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	20358	20358	413.99	0.000
Residual Error	998	49076	49		
Total	999	69433			

Ecuación b

MTB > regress c5 1 c6

The regression equation is
 postest = 50.5 - 0.976 grupo

Predictor	Coef	SE Coef	T	P
Constant	50.4516	0.3136	160.88	0.000
grupo	-0.9761	0.4435	-2.20	0.028

S = 7.01242 R-Sq = 0.5% R-Sq(adj) = 0.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	238.19	238.19	4.84	0.028
Residual Error	998	49075.67	49.17		
Total	999	49313.86			

Tal como se aprecia en la primera ecuación de regresión, los 10 puntos teóricos de impacto del programa en el grupo experimental se han convertido, por efectos del azar en 9.02 puntos. El ajuste de la ecuación es alto, tal como lo indica el determinante de la regresión (29.3% de predicción). Caso contrario ha ocurrido con la ecuación de regresión cuando el programa no ha tenido ningún impacto. En este caso, los cero puntos teóricos del impacto del programa en el grupo experimental se han convertido, por efectos del azar, en -0.97 puntos. En este caso, el ajuste de la ecuación es bajo (0.4% de predicción). Debe advertirse que en estos ejercicios el impacto nulo no ha sido cero ni el impacto ha sido 10, por una sencilla razón: error de medida. En efecto, si bien los diseños experimentales controlan

muchas amenazas de validez interna, no controlan en absoluto amenazas de validez de constructo. Eso corresponde al terreno de la psicometría. Recuérdese que al generar los datos se agregó un error de medida de 5 puntos de dispersión. En este caso, para demostrar lo afirmado, se reducirá el error de medida a 2 puntos de dispersión.

```
MTB > random 1000 c9;
SUBC> normal 0 2.
MTB > random 1000 c10;
SUBC> normal 0 2.
MTB > add c1 c9 c11
MTB > add c1 c10 c12
MTB > let c13=c12+(10*c6)
```

Ecuación c

```
MTB > regress c13 1 c6
```

```
The regression equation is
postest 2 2p = 50.5 + 9.20 grupo
```

Predictor	Coef	SE Coef	T	P
Constant	50.5343	0.2456	205.73	0.000
grupo	9.1991	0.3474	26.48	0.000

```
S = 5.49265    R-Sq = 41.3%    R-Sq(adj) = 41.2%
```

```
Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	1	21156	21156	701.24	0.000
Residual Error	998	30109	30		
Total	999	51265			

Obsérvese como se ha reducido el residual de error de 49075 (ecuación b) a 30109 (ecuación c), repercutiendo positivamente en la predicción de la ecuación (el modelo explica ahora el 41.2% de la varianza de los datos y ya no 29.2% del modelo anterior). En este caso, una mejora en la fiabilidad de las mediciones ha repercutido en la sensibilidad de los resultados. Incluso en la representación visual se observa la mejora.

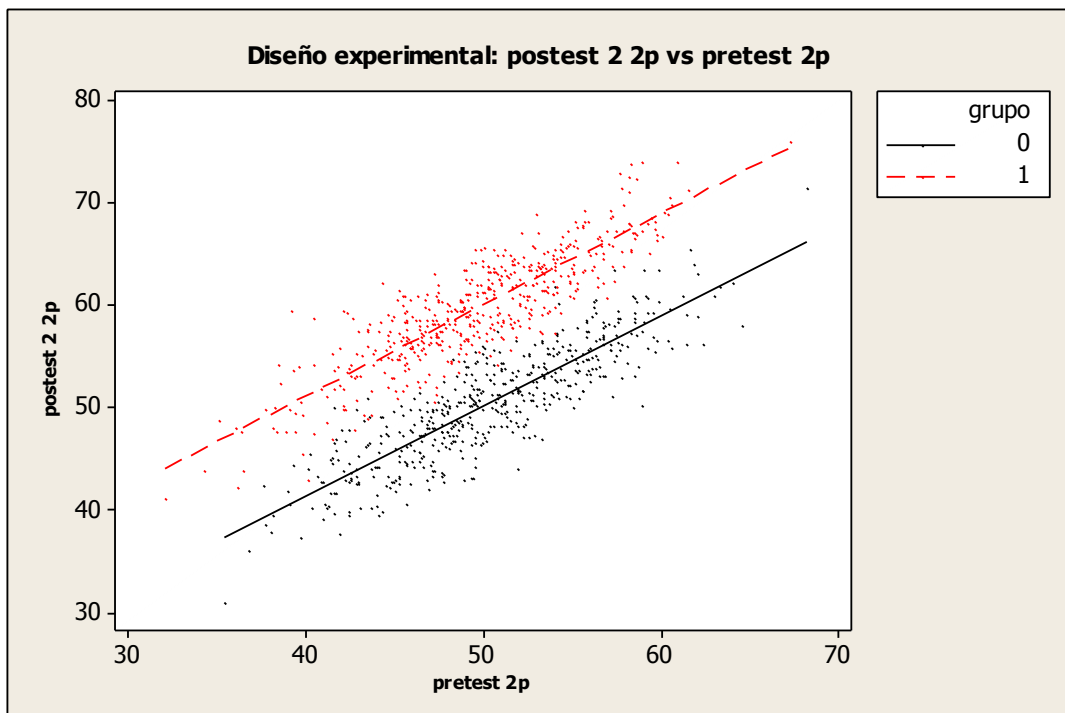


Figura 4.9. Relación entre posttest y pretest cuando existe impacto del programa en 10 puntos y con una reducción del error de medida (Fuente: Elaboración propia).

Frente a lo demostrado, es importante reiterar la necesidad de controlar las amenazas contra la validez según el parámetro teórico del Modelo Arysis. A pesar que se tenga un diseño experimental, no se garantiza con ello la validez de constructo ni la validez de contexto. Recuérdese que los diseños experimentales solamente controlan la validez interna, es decir, la validez de relación causal. Por eso, una revisión de las propiedades psicométricas (fiabilidad y validez de constructo) de los instrumentos de medida del pretest y posttest son totalmente necesarios; porque postular un modelo de análisis que omita variables relevantes o que incorpore variables con baja fiabilidad puede arruinar la estimación del efecto del programa, produciendo conclusiones erróneas.

4.4.2. El modelo común: Grupos no equivalentes

Los diseños cuasi-experimentales se crearon porque en la evaluación de programas es común que los grupos control y experimental no sean equivalentes, es decir, que no sean semejantes. Al no tener una asignación aleatoria, no existe seguridad de equivalencia inicial entre grupos, y la validez interna se ve amenazada. Por eso, se requieren medidas pretest y, de ser posible, ajustes posteriores de covarianza, emparejamiento de grupos o uso de variables instrumentales para corregir la falta de equivalencia entre los grupos control y experimental. Recuérdese que estos ajustes deben ser planificados previamente y no dejados a posteriori. El Modelo Arysis propuesto es un modelo analítico y se centra en la planificación del diseño más que en el ajuste posterior. No debe olvidarse que el diseño es prioritario sobre el análisis estadístico. La validez interna

se fortalece adoptando soluciones basadas en el control experimental y no en el control estadístico. Es preferible descartar amenazas contra la validez a priori, mediante una planificación adecuada del diseño que minimice la actuación de variables extrañas.

En este ejemplo de simulación, se han creado 1000 observaciones con parámetros similares al modelo anterior (promedio teórico de 50 y una desviación estándar de 5; sumado a un error de medición con un promedio de 0 y una desviación estándar de 5). Así, se tiene una puntuación pretest y postest. Las observaciones han sido divididas en dos grupos (control = 0 y experimental=1), ambos grupos no son equivalentes, por tanto tienen diferencias significativas entre ellos, en este caso, de cinco puntos a favor del grupo experimental.

```
MTB > random 1000 c1;
SUBC> normal 50 5.
MTB > random 1000 c2;
SUBC> normal 0 5.
MTB > random 1000 c3;
SUBC> normal 0 5.
MTB > add c1 c2 c4
MTB > add c1 c3 c5
MTB > set c6
DATA> 1:1000
DATA> end
MTB > code (1:500) 0 c6 c6
MTB > code (501:1000) 1 c6 c6
MTB > table c6
Rows: grupos
      Count
0         500
1         500
All      1000

MTB > sign c6
0 Negative values    500  Zero values    500  Positive values
```

```
MTB > table c6;
SUBC> mean c4 c5.
Rows: grupos
      pretest  posttest
      Mean    Mean
0         49.84   50.08
1         50.33   50.03
All       50.08   50.05
```

```
MTB > let c4 = c4 + (5*c6)
MTB > let c5 = c5 + (5*c6)
```

```
MTB > table c6;
SUBC> mean c4 c5;
SUBC> stdev c4 c5.
Rows: grupos
      pretest  posttest  pretest  posttest
      Mean    Mean    StDev    StDev
```

```

0          49.84    50.08    7.399    7.374
1          55.33    55.03    7.327    7.157
All        52.58    52.55    7.857    7.674

```

```
MTB > let c5 = c5 + (10*c6)
```

```
MTB > table c6;
```

```
SUBC> mean c4 c5;
```

```
SUBC> stdev c4 c5.
```

```
Rows: grupos
```

	pretest	postest	pretest	postest
	Mean	Mean	StDev	StDev
0	49.84	50.08	7.399	7.374
1	55.33	65.03	7.327	7.157
All	52.58	57.55	7.857	10.425

```
MTB > Plot 'postest'*'pretest';
```

```
SUBC> Symbol 'grupos';
```

```
SUBC> Regress 'grupos';
```

```
SUBC> Overlay.
```

En este ejemplo de simulación, se ha previsto un impacto del programa de 10 puntos. La siguiente figura grafica el impacto.

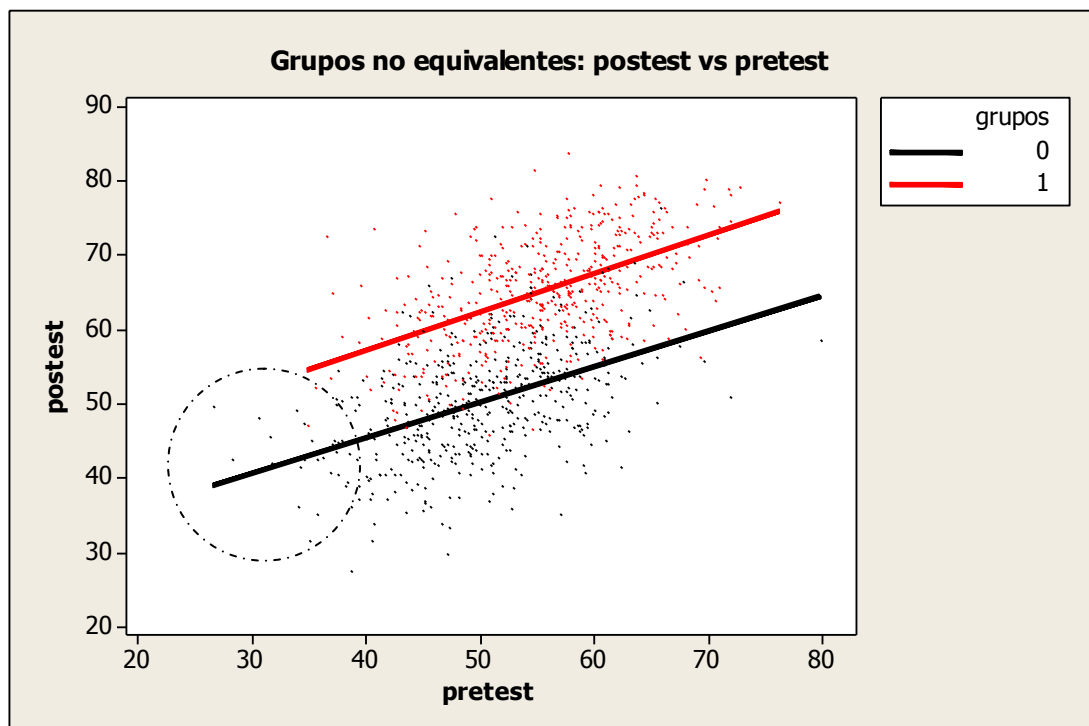


Figura 4.10. Relación entre postest y pretest cuando existe impacto del programa en 10 puntos y cuando los grupos no son equivalentes (Fuente: Elaboración propia).

Tal como se aprecia en la Figura 4.10, el grupo experimental tiene un efecto positivo producto de la aplicación del programa, sin embargo, se observa

también una no-equivalencia inicial en el pretest (el círculo discontinuado), lo que daría a entender que los datos están sesgados positivamente a favor del grupo experimental, incrementando artificialmente el impacto.

Para demostrar el efecto de la no-equivalencia de los grupos en la validez de los resultados, se realizan dos ecuaciones de regresión. En la primera se somete a prueba el efecto del programa sin ajuste alguno. En la segunda, se ajusta la no-equivalencia del grupo.

Ecuación d

MTB > regress c5 1 c6

The regression equation is
 postest = 50.1 + 15.0 grupos

Predictor	Coef	SE Coef	T	P
Constant	50.0786	0.3250	154.11	0.000
grupos	14.9515	0.4596	32.53	0.000

S = 7.26638 R-Sq = 51.5% R-Sq(adj) = 51.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	55887	55887	1058.46	0.000
Residual Error	998	52695	53		
Total	999	108581			

Ecuación e

MTB > regress c5 2 c6 c4

The regression equation is
 postest = 25.2 + 12.2 grupos + 0.499 pretest

Predictor	Coef	SE Coef	T	P
Constant	25.201	1.373	18.36	0.000
grupos	12.2072	0.4234	28.83	0.000
pretest	0.49920	0.02696	18.52	0.000

S = 6.27125 R-Sq = 63.9% R-Sq(adj) = 63.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	69371	34685	881.94	0.000
Residual Error	997	39211	39		
Total	999	108581			

Source	DF	Seq SS
grupos	1	55887
pretest	1	13484

Tal como se aprecia en la primera ecuación de regresión (ecuación d), los 10 puntos teóricos de impacto del programa en el grupo experimental se han convertido, por efectos de la no-equivalencia, en 14.95 puntos (con un ajuste de ecuación del 51.4%). En la segunda ecuación (ecuación e), cuando ya se controla

los efectos de la no-equivalencia de los grupos, se encuentra que el impacto del programa se reduce a 12.20, mejorando el ajuste de la ecuación (63.8% de predicción) en casi el 12%. Esta mejora también se ve reflejada en el residual de error, el cual reduce la suma cuadrática de 52695 a 39211. Téngase en cuenta también que en este caso también se ha usado un error de medición alto, por lo que si se mejora la validez de constructo y la fiabilidad de la medida, la predicción mejora sustancialmente.

En el ejemplo anterior se demuestra como la no-equivalencia inicial de los grupos control y experimental pueden sesgar positivamente los resultados, y hacernos concluir –erróneamente- que el programa ha tenido un fuerte impacto, cuando en realidad el impacto es mucho menor. En el caso contrario, es decir, cuando la no-equivalencia inicial es en detrimento del grupo experimental, puede subestimarse el impacto del programa. Véase el ejemplo de simulación siguiente:

```

MTB > add c1 c2 c10
MTB > add c1 c3 c11
MTB > let c10 = c10 - (5*c6)
MTB > let c11 = c11 - (5*c6)
MTB > table c6;
SUBC> mean c10 c11;
SUBC> stdev c10 c11.
Rows: grupos
      pretest 2  postest 2  pretest 2  postest 2
      Mean      Mean      StDev    StDev
0      49.84     50.08     7.399    7.374
1      45.33     45.03     7.327    7.157
All    47.58     47.55     7.697    7.689

MTB > let c11 = c11 + (10*c6)
MTB > table c6;
SUBC> mean c10 c11;
SUBC> stdev c10 c11.
Rows: grupos
      pretest 2  postest 2  pretest 2  postest 2
      Mean      Mean      StDev    StDev
0      49.84     50.08     7.399    7.374
1      45.33     55.03     7.327    7.157
All    47.58     52.55     7.697    7.674

MTB > Plot 'postest'*'pretest';
SUBC> Symbol 'grupos';
SUBC> Regress 'grupos';
SUBC> Overlay.
MTB > Plot 'postest 2'*'pretest 2';
SUBC> Symbol 'grupos';
SUBC> Regress 'grupos';
SUBC> Overlay.

```

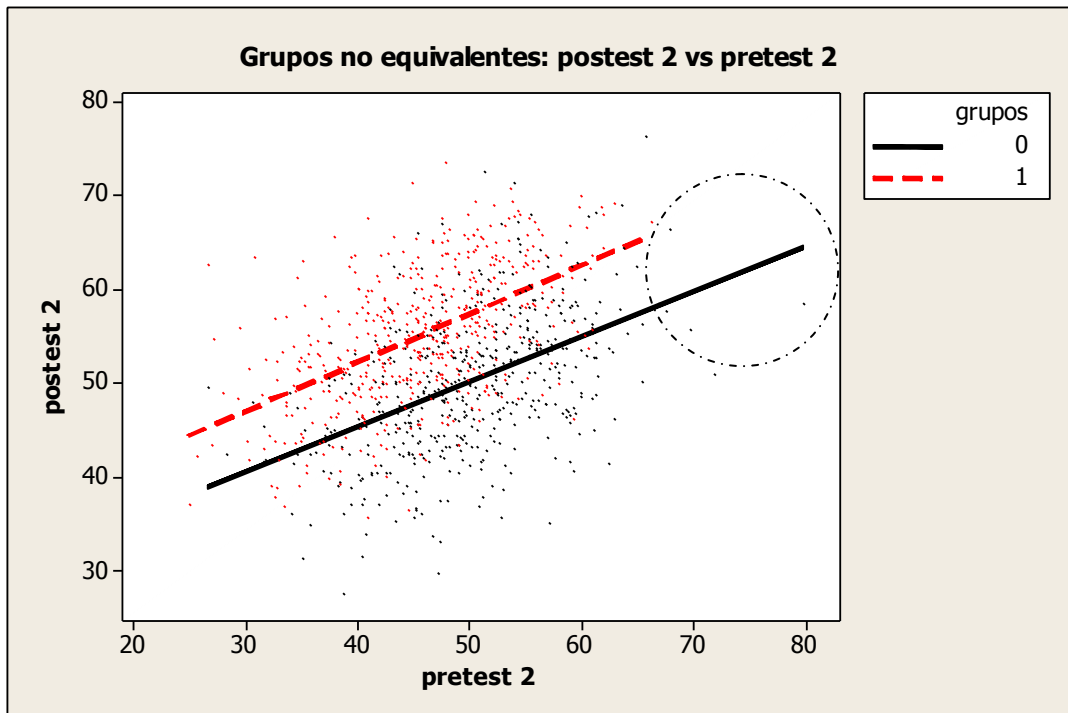


Figura 4.11. Relación entre postest y pretest cuando existe impacto del programa en 10 puntos y cuando los grupos no son equivalentes en perjuicio del GE (Fuente: Elaboración propia).

Tal como se observa en la Figura 4.11, existe una no-equivalencia inicial (pretest) en detrimento del grupo experimental, es decir, el grupo control tenía mayor puntuación que el grupo experimental antes de iniciar el tratamiento, por eso el efecto del tratamiento (de 10 puntos) se ve más pequeño de lo que realmente es, es decir, se subrepresenta.

En efecto, cuando el grupo control tiene una puntuación pretratamiento sesgada positivamente, entonces es muy probable que el impacto del programa se subrepresente. Si se observa la siguiente ecuación (f), se verá que el impacto predicho del programa es de sólo 4.9515 puntos y no 10 como se estipuló teóricamente. Esa reducción de más de 5 puntos es producto de la no-equivalencia inicial. Al hacer el ajuste del sesgo inicial que produce la no-equivalencia, el impacto del programa predicho se aproxima más al real (7.1992). La mejora ocurre en más del 23%; pero siempre está presente el error de medición.

Ecuación f

MTB > regress c11 1 c6

The regression equation is

postest 2 = 50.1 + 4.95 grupos

Predictor	Coef	SE Coef	T	P
Constant	50.0786	0.3250	154.11	0.000
grupos	4.9515	0.4596	10.77	0.000

S = 7.26638 R-Sq = 10.4% R-Sq(adj) = 10.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	6129.3	6129.3	116.08	0.000

```
Residual Error 998 52694.6 52.8
Total          999 58823.9
```

Ecuación g

MTB > regress c11 2 c6 c10

The regression equation is
 postest 2 = 25.2 + 7.20 grupos + 0.499 pretest 2

Predictor	Coef	SE Coef	T	P
Constant	25.201	1.373	18.36	0.000
grupos	7.1992	0.4148	17.36	0.000
pretest 2	0.49920	0.02696	18.52	0.000

S = 6.27125 R-Sq = 33.3% R-Sq(adj) = 33.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	19613.3	9806.7	249.35	0.000
Residual Error	997	39210.6	39.3		
Total	999	58823.9			

Source	DF	Seq SS
grupos	1	6129.3
pretest 2	1	13484.0

Con estos ejemplos de simulación se demuestra, entonces, la importancia del control en el diseño cuasi-experimental sobre todo cuando los grupos no son equivalentes. Si no se controlan las variables extrañas asociadas a la selección de los grupos, el sesgo puede sobrerrepresentar o subrepresentar drásticamente el impacto del programa, llevando a conclusiones erróneas. En estos ejemplos se han controlado los sesgos de selección mediante el uso de una variable instrumental, una covariante asociada a las puntuaciones iniciales (pretest) de los grupos. Pero esto se ha podido hacer, porque ya estaba planificado desde el diseño, por eso se recomienda planificar cuidadosamente el diseño, incluso cuando se tiene que ajustar estadísticamente⁷¹. Por eso se recomienda la observación directa del proceso de selección para incorporar al modelo analítico todas aquellas variables relevantes –pero desconocidas- que están influyendo en el proceso de selección de los grupos. Además, es conveniente practicar simulaciones previas múltiples bajo diferentes supuestos sobre los factores que pueden afectar directamente a los grupos.

4.4.3. Discontinuidad de la regresión

En este subtítulo se crea y analiza datos para un diseño de discontinuidad de la regresión. Tal como se mencionó previamente, este diseño, a pesar de ser

⁷¹ Las técnicas de ajuste que aquí se han utilizado se han podido realizar porque han estado planificadas, pero existen otros procedimientos que también podrían utilizarse, como el emparejamiento (matching) y la estratificación.

escasamente utilizado, es muy recomendado por sus excelentes perspectivas interpretativas. Incluso, muchos autores lo consideran a medio camino entre el diseño experimental propiamente dicho y el diseño cuasi-experimental de control no-equivalente (Anguera, et al, 1995).

No debe olvidarse que hay al menos tres variables a considerar en un diseño de discontinuidad de la regresión: a) la variable de asignación conocida a partir del cual se define el punto de corte para asignar a los sujetos al GC y GE, b) la variable tratamiento (el programa) y c) la variable dependiente (postest) el cual registra los efectos.

En este caso, igual que en los ejemplos anteriores, se generan 1000 casos de variables pretest y postest, con sus respectivos errores de medición de cinco puntos de desviación; sin embargo, aquí la asignación al grupo control se hace sobre un criterio “conocido”, generalmente utilizando las puntuaciones del pretest, estableciendo un “punto de corte” entre grupo control y experimental, asignando las puntuaciones superiores a un grupo y las puntuaciones inferiores al otro grupo.

```
MTB > random 1000 c1;
SUBC> normal 50 5.0.
MTB > random 1000 c2;
SUBC> normal 0 5.0.
MTB > random 1000 c3;
SUBC> normal 0 5.0.
MTB > add c1 c2 c4
MTB > code (0:50) 1 c4 c5
MTB > code (50:100) 0 c5 c5

MTB > table c5
Rows: C5
      Count
0      484
1      516
All    1000

MTB > let c6=c1+c3+(10*c5)

MTB > table c5;
SUBC> mean c4 c6;
SUBC> stdev c4 c5.
Rows: grupos
      pretest  postest  pretest  grupos
      Mean    Mean     StDev   StDev
0      55.95   53.15   4.361   0.0000
1      44.03   57.18   4.340   0.0000
All    49.80   55.23   7.374   0.5000

MTB > Plot 'postest'*'pretest';
SUBC> Symbol 'grupos';
SUBC> Regress 'grupos';
SUBC> Overlay.
```

Tal como se observa en los comandos, se ha asignado al grupo experimental aquellos individuos ($n=516$) con puntuaciones inferiores al promedio de corte del pretest (50 puntos), mientras que el grupo control estaría formado por individuos con puntuaciones superiores al punto de corte ($n=484$). En este experimento simulado, se ha aplicado el programa al grupo con menores puntuaciones en el pretest (variables que mida, p.e. calidad de vida). Esta asignación es válida en la medida que una de las políticas gubernamentales es dirigirse a los más necesitados.

Posteriormente se ha supuesto que el programa aplicado ha sido efectivo en 10 puntos teóricos. De lo dicho, el siguiente gráfico demuestra el punto de corte existente entre ambos grupos y el impacto del programa. El impacto del programa se mide a través de la diferencia (en unidades del postest) de las intercepciones de ambas líneas de regresión en el punto de corte.

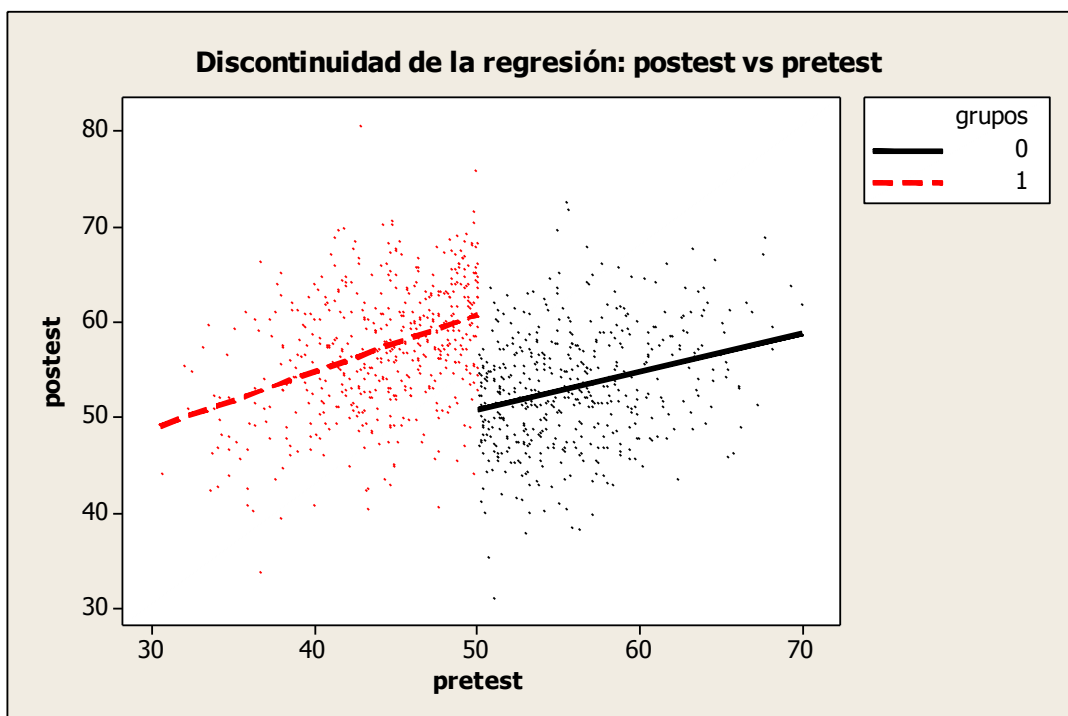


Figura 4.12. Relación entre postest y pretest cuando existe impacto del programa en 10 puntos y cuando los grupos no tienen asignación conocida (Fuente: Elaboración propia).

Si el programa no hubiese sido efectivo (impacto=0), entonces la gráfica de dispersión hubiese sido como la siguiente:

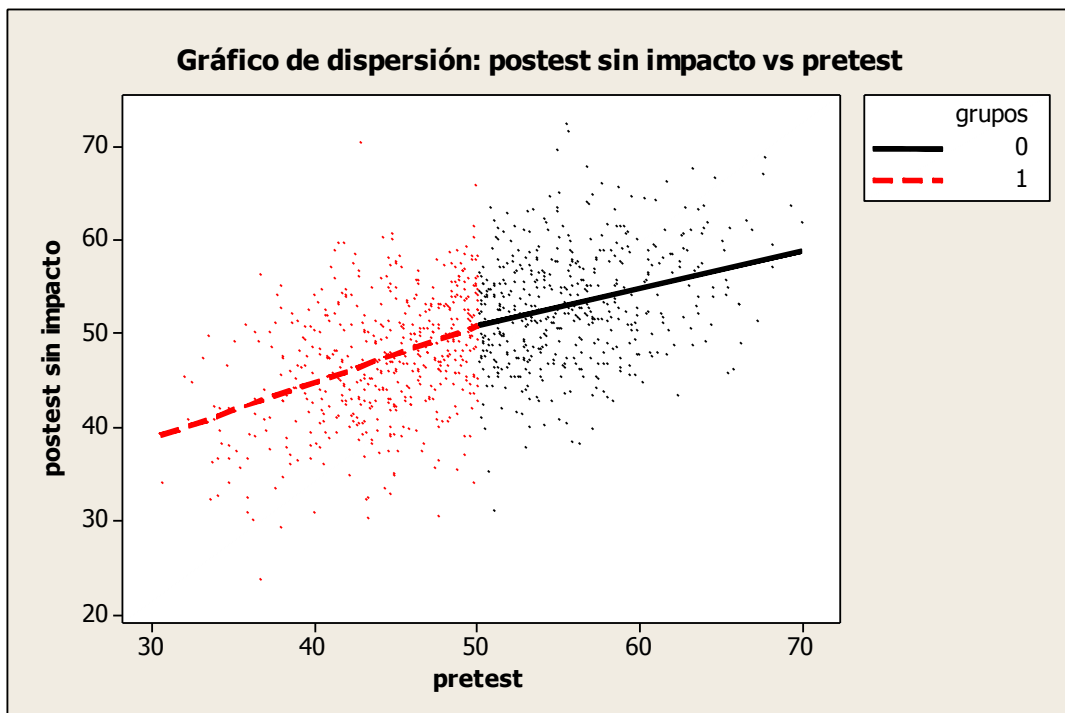


Figura 4.13. Relación entre posttest y pretest cuando no existe impacto del programa y cuando los grupos son de asignación conocida (Fuente: Elaboración propia).

En este caso (Figura 4.13) no se observa ninguna “discontinuidad” de la regresión, por lo que se deduce la inexistencia de impacto a favor de alguno de los grupos.

En el modelo de la discontinuidad de la regresión, se utiliza una variable de asignación conocida para formar los grupos control y experimental. En la ecuación (h), se presenta la regresión sin incluir la variable de asignación. En este caso, aparentemente el programa ha tenido un impacto de 4 puntos, pero el porcentaje de ajuste es bajo (9.1%). Como se conoce la variable de asignación, entonces se ingresa a la ecuación como covariante (ecuación i), aumentándose el ajuste en 19.9% y estimando el impacto del programa con un margen de error mínimo (impacto= 10.0437). Por eso, muchos autores consideran que el diseño de discontinuidad de la regresión es tan robusto como un diseño experimental puro, porque las diferencias de estimación son casi próximas a cero.

Ecuación h

MTB > regress c6 1 c5

The regression equation is
 posttest = 53.2 + 4.03 grupos

Predictor	Coef	SE Coef	T	P
Constant	53.1515	0.2899	183.33	0.000
grupos	4.0318	0.4036	9.99	0.000

S = 6.37824 R-Sq = 9.1% R-Sq(adj) = 9.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	4059.6	4059.6	99.79	0.000

```
Residual Error 998 40600.6 40.7
Total          999 44660.2
```

Ecuación i

MTB > regress c6 2 c7 c5

The regression equation is
 postest = 50.2 + 0.505 pre-corte + 10.0 grupos

Predictor	Coef	SE Coef	T	P
Constant	50.1506	0.3760	133.39	0.000
pre-corte	0.50473	0.04360	11.58	0.000
grupos	10.0437	0.6430	15.62	0.000

S = 5.99142 R-Sq = 19.9% R-Sq(adj) = 19.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	8870.9	4435.4	123.56	0.000
Residual Error	997	35789.4	35.9		
Total	999	44660.2			

Ecuación j

MTB > regress c12 1 c5

The regression equation is
 postest sin impacto = 53.2 - 5.97 grupos

Predictor	Coef	SE Coef	T	P
Constant	53.1515	0.2899	183.33	0.000
grupos	-5.9682	0.4036	-14.79	0.000

S = 6.37824 R-Sq = 18.0% R-Sq(adj) = 17.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	8895.8	8895.8	218.67	0.000
Residual Error	998	40600.6	40.7		
Total	999	49496.4			

Ecuación k

MTB > regress c12 2 c5 c7

The regression equation is
 postest sin impacto = 50.2 + 0.044 grupos + 0.505 pre-corte

Predictor	Coef	SE Coef	T	P
Constant	50.1506	0.3760	133.39	0.000
grupos	0.0437	0.6430	0.07	0.946
pre-corte	0.50473	0.04360	11.58	0.000

S = 5.99142 R-Sq = 27.7% R-Sq(adj) = 27.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	13707.0	6853.5	190.92	0.000
Residual Error	997	35789.4	35.9		
Total	999	49496.4			

En el ejemplo de simulación donde no existe impacto del programa, la estimación es similar. Cuando se predice el impacto sin controlar la variable de asignación (ecuación j), la estimación indica un impacto negativo de -5.9682 con una predicción del 18%; sin embargo, incluyendo la covariante de asignación

conocida (ecuación k) la predicción aumenta a 27.7% y la estimación se aproxima a la real (0.0437), demostrando ausencia de impacto del programa.

En consecuencia, se demuestra el enorme poder de los diseños de discontinuidad de regresión para controlar las diferencias de asignación entre grupos, convirtiendo variables de asignación desconocidas y descontroladas en variables de asignación conocidas y ahora controladas.

4.4.4. Series temporales interrumpidas

Tal como se mencionó en el subtítulo 4.3.2.9.1, en el diseño simple de series temporales, existen varios posibles resultados, algunos relacionados al cambio de nivel (impacto del programa), otros relacionados al cambio de la tendencia o de dirección.

Para que en un diseño simple sólo se observe cambio de nivel, es suficiente utilizar los siguientes comandos:

```
MTB > set c1
DATA> 1:20
DATA> end
MTB > random 20 c2;
SUBC> normal 50 5.
MTB > set c3
DATA> 1:20
DATA> end
MTB > code (1:10) 0 c3 c3
MTB > code (11:20) 1 c3 c3
MTB > random 20 c4;
SUBC> normal 0 1.

MTB > add c2 c4 c5.
MTB > let c6 = c5 + (5*c3)

MTB > TSPlot 'P. sin cambio de nivel';
SUBC> Symbol 'grupos (antes-después)';
SUBC> Connect 'grupos (antes-después)'.

MTB > TSPlot 'P. con cambio nivel';
SUBC> Symbol 'grupos (antes-después)';
SUBC> Connect 'grupos (antes-después)'.

```

En este caso, se han creado veinte observaciones continuas, con un promedio de 50 y una desviación estándar de cinco, distribuidos aleatoriamente en 10 observaciones anteriores a la aplicación del programa y 10 observaciones posteriores a la aplicación del programa. Se le ha sumado un error de medición (media=0, D.E.=1) y al grupo de observaciones posteriores a la aplicación del programa se le ha sumado un impacto de nivel de 5 puntos.

Tal como se muestra en la Figura 4.14, el grupo de observaciones posterior al tratamiento tiene un cambio de nivel superior al grupo anterior de observaciones.

Sin embargo, las observaciones sin cambio de nivel (triángulos) es “aparentemente” también significativo, pero recuérdese, que de antemano se elaboró los datos indicando ausencia de impacto, así que esa apariencia es engañosa.

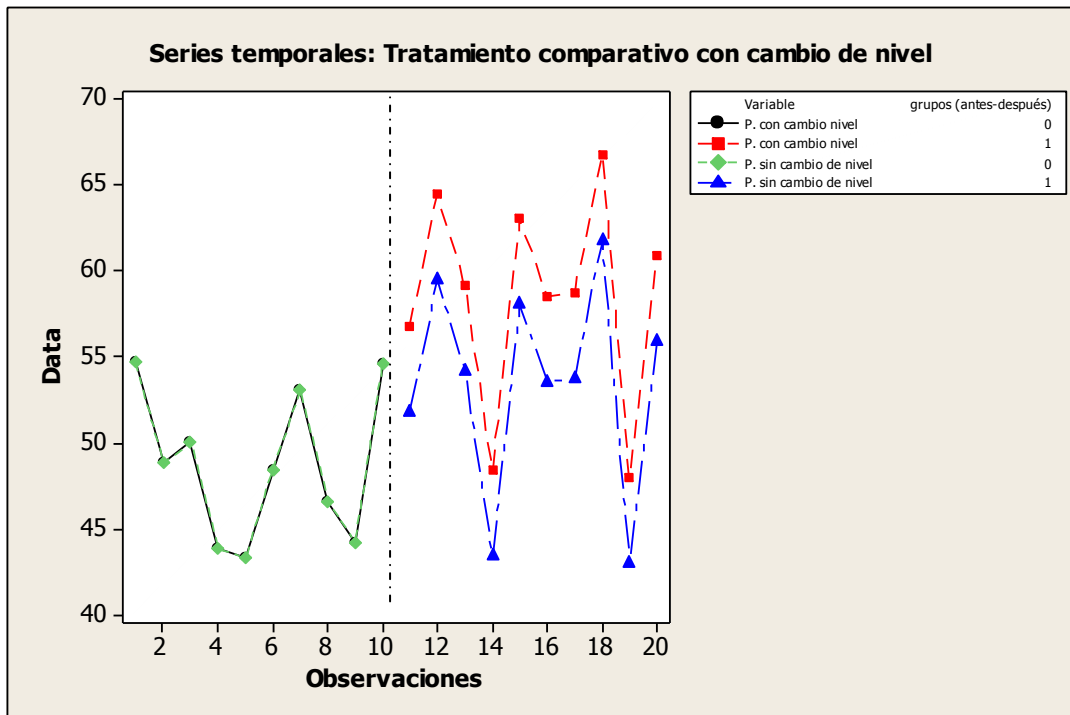


Figura N° 4.14. Serie simple con cambio de nivel posprograma, en supuesto significativo y no significativo (Fuente: Elaboración propia).

Por ello, para controlar esos efectos azarosos, no es suficiente el análisis gráfico, se requiere del análisis estadístico de la regresión.

```

MTB > tabla c3;
SUBC> mean c5 c6.
Rows: grupos (antes-después)
      P. sin
      cambio de nivel
      P. con
      cambio de nivel
      Mean      Mean
0      48.78    48.78
1      53.50    58.50
All    51.14    53.64

```

Ecuación 1

```

MTB > regress c5 1 c3
The regression equation is
P. sin cambio de nivel = 48.8 + 4.72 grupos (antes-después)
Predictor      Coef    SE Coef      T      P
Constant      48.783  1.693    28.81  0.000
grupos (antes-después)  4.720  2.395    1.97  0.064

```

S = 5.35434 R-Sq = 17.8% R-Sq(adj) = 13.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	111.39	111.39	3.89	0.064
Residual Error	18	516.04	28.67		
Total	19	627.43			

Ecuación m

MTB > regress c6 1 c3

The regression equation is

P. con cambio nivel = 48.8 + 9.72 grupos (antes-después)

Predictor	Coef	SE Coef	T	P
Constant	48.783	1.693	28.81	0.000
grupos (antes-después)	9.720	2.395	4.06	0.001

S = 5.35434 R-Sq = 47.8% R-Sq(adj) = 44.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	472.39	472.39	16.48	0.001
Residual Error	18	516.04	28.67		
Total	19	988.43			

Tal como se observa en los resultados, aunque aparentemente existe una diferencia promedio de 5 y 10 puntos para ambos grupos, sólo existe cambio de nivel para el segundo grupo de observaciones simuladas, pues sólo en la ecuación (m) el coeficiente de grupos es significativo ($p=0.001$), explicando 30% más que la ecuación (l).

Puede ocurrir que durante la aplicación de un programa social exista un cambio de tendencia en los resultados, es decir, estos pueden incrementarse o decrementarse a medida que pase el tiempo de aplicación. En ese sentido, el impacto del programa se mezcla con estas variaciones temporales de error. Si se sospecha de esta posibilidad, es necesario ajustar el modelo simple de series temporales.

Para crear una serie temporal sin cambio de nivel, pero con cambio de tendencia, se requieren los siguiente comandos:

```
MTB > set c1
DATA> 1:20
DATA> end
MTB > random 20 c2;
SUBC> normal 50 5.
MTB > random 20 c3;
SUBC> normal 0 1.
MTB > add c2 c3 c4
MTB > set c5
DATA> 1:20
DATA> end
MTB > code (1:10) 0 c5 c5
```

```

MTB > code (11:20) 1 c5 c5
MTB > add c1 c3 c8
MTB > add c4 c8 c9
MTB > TSPlot 'P.sin cambio, con tendencia';
SUBC> Symbol 'Grupo (antes-después)';
SUBC> Connect 'Grupo (antes-después)';
SUBC> Project 'Grupo (antes-después)'.

```

Cuando una serie temporal no presenta cambio de nivel pero sí de tendencia, la representación gráfica es como sigue:

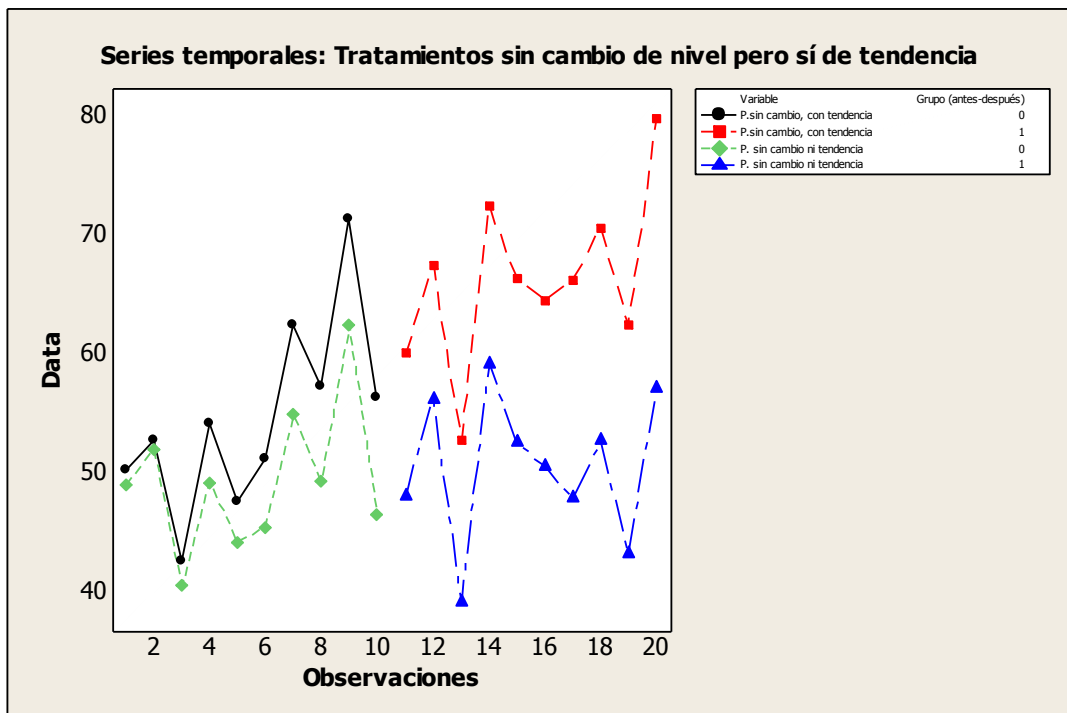


Figura N° 4.15. Serie simple sin cambio de nivel y con cambio de tendencia del posprograma, en supuesto significativo y no significativo (Fuente: Elaboración propia).

A simple vista se puede suponer que el programa ha sido efectivo, pero eso no es así, porque el nivel no ha cambiado, sólo ha cambiado la tendencia, y esta está asociada a la maduración o la estacionalidad de los datos. Las siguientes ecuaciones de regresión demostrarán lo afirmado.

```

MTB > regress c4 1 c5
The regression equation is
P. sin cambio ni tendencia = 49.4 + 1.27 Grupo (antes-después)

```

Predictor	Coef	SE Coef	T	P
Constant	49.448	1.956	25.28	0.000
Grupo (antes-después)	1.267	2.767	0.46	0.652

S = 6.18666 R-Sq = 1.2% R-Sq(adj) = 0.0%

Ecuación n


```

Analysis of Variance
Source          DF      SS      MS      F      P
Regression      1      8.03     8.03    0.21   0.652
Residual Error  18     688.95   38.27
Total           19     696.98

```

Ecuación o

MTB > regress c9 1 c5

The regression equation is
P.sin cambio, con tendencia = 54.7 + 11.7 Grupo (antes-después)

```

Predictor          Coef  SE Coef      T      P
Constant           54.673  2.426    22.54  0.000
Grupo (antes-después) 11.657  3.430     3.40  0.003

```

S = 7.67013 R-Sq = 39.1% R-Sq(adj) = 35.7%

```

Analysis of Variance
Source          DF      SS      MS      F      P
Regression      1     679.48   679.48  11.55  0.003
Residual Error  18    1058.96   58.83
Total           19    1738.43

```

Ecuación p

MTB > regress c9 2 c5 c1

The regression equation is
P.sin cambio, con tendencia = 46.4 - 3.47 Grupo (antes-después) + 1.51 observación

```

Predictor          Coef  SE Coef      T      P
Constant           46.355  3.370    13.75  0.000
Grupo (antes-después) -3.466  5.685    -0.61  0.550
observación         1.5123  0.4929     3.07  0.007

```

S = 6.33189 R-Sq = 60.8% R-Sq(adj) = 56.2%

```

Analysis of Variance
Source          DF      SS      MS      F      P
Regression      2    1056.85   528.43  13.18  0.000
Residual Error  17     681.58   40.09
Total           19    1738.43

```

```

Source          DF  Seq SS
Grupo (antes-después) 1  679.48
observación         1  377.38

```

En la ecuación (m) se demuestra que no existe impacto significativo (sin cambio de nivel) del programa. Sin embargo, en la ecuación (n) existe – aparentemente- un impacto de 11.657 puntos a favor del programa, lo cual es falso porque previamente se supuso que no existía cambio de nivel alguno. Es el cambio de tendencia – historia de los datos, maduración, eventos asociados no controlados- la que está produciendo los resultados artificiales. Entonces, para controlar esta tendencia se introduce una covariante (el número de observación), demostrando en la ecuación (p) que no existe ningún impacto significativo del programa, es decir que no existe ningún cambio de nivel.

Cuando existe cambio de nivel y cambio de tendencia, el impacto del programa tiende a mezclarse con estos errores, sobre-estimándolo. Por ejemplo, si

se simula un programa efectivo en 10 puntos pero con la presencia de tendencia ascendente en los datos, entonces, el impacto será sobre representado.

```
MTB > let c11=c9+(10*c5)
```

Ecuación q

```
MTB > regress c11 1 c5
```

The regression equation is

P.con cambio, con tendencia = 54.7 + 21.7 Grupo (antes-después)

Predictor	Coef	SE Coef	T	P
Constant	54.673	2.426	22.54	0.000
Grupo (antes-después)	21.657	3.430	6.31	0.000

S = 7.67013 R-Sq = 68.9% R-Sq(adj) = 67.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2345.2	2345.2	39.86	0.000
Residual Error	18	1059.0	58.8		
Total	19	3404.2			

En la ecuación (q) se observa que el impacto del programa de 10 puntos ha sido sobre-representado en 11.657 puntos adicionales. En la figura siguiente se observa, sin embargo, la tendencia ascendente de los datos, por lo que se requiere su control ingresándola como covariante.

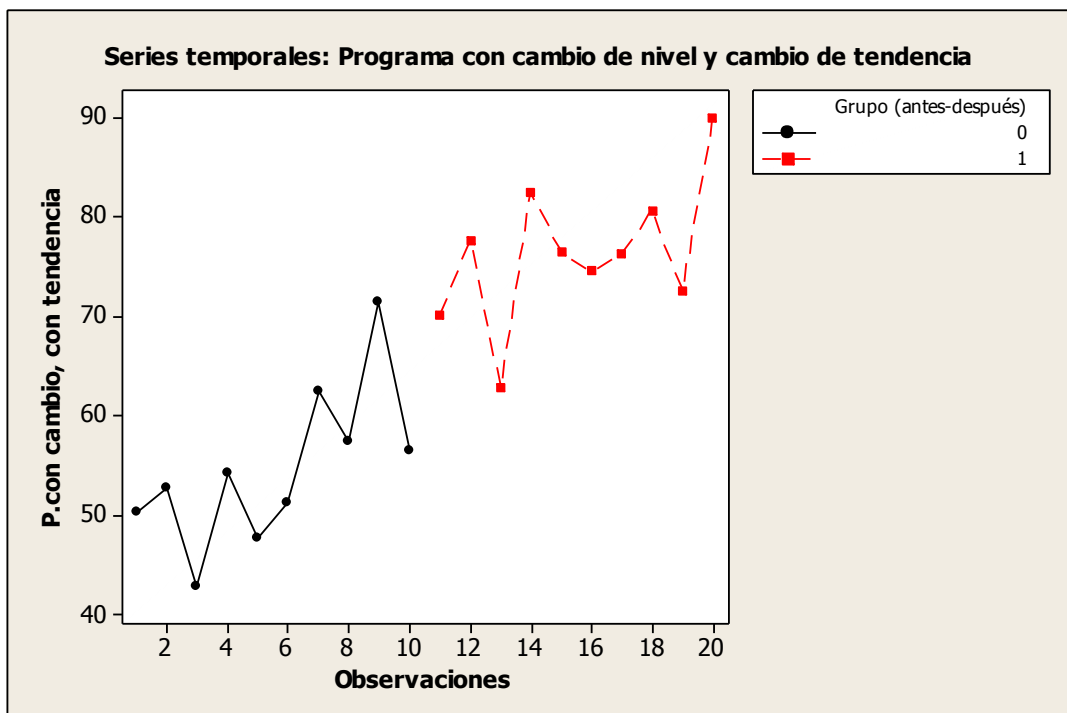


Figura N° 4.16. Serie simple con cambio de nivel y con cambio de tendencia del posprograma, en supuesto significativo (Fuente: Elaboración propia).

Si se incluye la tendencia como covariante (ecuación r) se consigue un ajuste al 80%, y el impacto del programa se reduce a 6.434 puntos. Pero en este caso, no sólo debe controlarse la tendencia global de los datos, sino también la interacción de la tendencia con el cambio de nivel. Por eso, en la ecuación (s) se incluye esta covariante adicional y se encuentra que el impacto del programa se ha modificado a 11.40 puntos, un valor más cercano al supuesto inicialmente.

Ecuación r

MTB > regress c11 2 c5 c1

The regression equation is

P.con cambio, con tendencia = 46.4 + 6.53 Grupo (antes-después)
+ 1.51 observación

Predictor	Coef	SE Coef	T	P
Constant	46.355	3.370	13.75	0.000
Grupo (antes-después)	6.534	5.685	1.15	0.266
observación	1.5123	0.4929	3.07	0.007

S = 6.33189 R-Sq = 80.0% R-Sq(adj) = 77.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2722.6	1361.3	33.95	0.000
Residual Error	17	681.6	40.1		
Total	19	3404.2			

Ecuación s

MTB > regress c11 3 c5 c1 c10

The regression equation is

P.con cambio, con tendencia = 45.1 + 11.4 Grupo (antes-después)
+ 1.74 observación
- 0.46 tendencia post tratamiento

Predictor	Coef	SE Coef	T	P
Constant	45.082	4.430	10.18	0.000
Grupo (antes-después)	11.40	12.09	0.94	0.360
observación	1.7438	0.7139	2.44	0.027
tendencia post tratamiento	-0.463	1.010	-0.46	0.653

S = 6.48430 R-Sq = 80.2% R-Sq(adj) = 76.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2731.44	910.48	21.65	0.000
Residual Error	16	672.74	42.05		
Total	19	3404.17			

A estos diseños simples antes-después, se puede agregar un grupo control, mejorando las perspectivas del análisis y el control de las amenazas contra la validez, principalmente aquellas relacionadas a la maduración, instrumentación, medición, regresión a la media e historia.

En este ejemplo de simulación se tiene dos grupos de datos, 20 de control (GC) y 20 experimental (GE) con medidas antes (10 primeras observaciones) y

después de la intervención del programa (10 últimas observaciones). Se ha supuesto un impacto de 10 puntos para el GE y 0 puntos para el GC.

Los resultados de la simulación se representan en la siguiente figura, donde claramente se observa un cambio de nivel a favor del GE, pero como existe una tendencia positiva para ambos grupos, el valor del impacto está confundido con ellas.

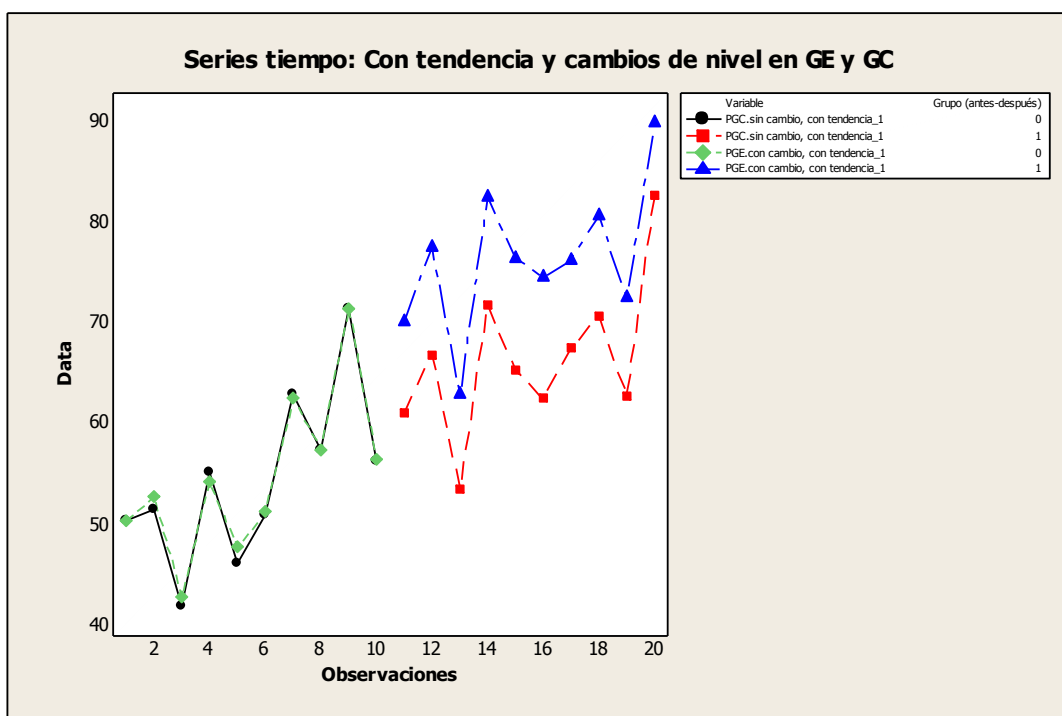


Figura N° 4.17. Serie temporal con con cambio de nivel y con cambio de tendencia del posprograma en GE, y sin cambio de nivel y con tendencia en GC (Fuente: Elaboración propia).

En efecto, en la ecuación (t) se observa el valor del impacto del programa sobre-representado, pues está confundido con el error de tendencia. En la ecuación anterior (s), se logró reducir ese sesgo utilizando como covariantes la tendencia y la interacción de la tendencia con el tratamiento. En la ecuación (s) el ajuste logrado era del 80.2%. Sin embargo, utilizando en las ecuaciones de regresión a los valores del grupo control como covariante, se puede corregir con mayor precisión los sesgos en la estimación del impacto. En efecto, en la ecuación (u) se ha incluido a los valores del grupo control como covariante y se ha logrado un ajuste del 99.4%, reduciendo el valor del impacto de 21.657 a 10.3236, casi idéntico al supuesto teóricamente.

En definitiva, el diseño de series temporales interrumpidas con grupo control permite un mayor nivel de certeza del impacto del programa, pues se controlan muchas amenazas contra la validez.

MTB > regress c13 1 c5

The regression equation is

PGE.con cambio, con tendencia_1 = 54.7 + 21.7 Grupo (antes-después)

Predictor	Coef	SE Coef	T	P
Constant	54.673	2.426	22.54	0.000
Grupo (antes-después)	21.657	3.430	6.31	0.000

S = 7.67013 R-Sq = 68.9% R-Sq(adj) = 67.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2345.2	2345.2	39.86	0.000
Residual Error	18	1059.0	58.8		
Total	19	3404.2			

MTB > regress c13 2 c5 c12

The regression equation is

PGE.con cambio, con tendencia_1 = 3.50 + 10.3 Grupo (antes-después)
+ 0.941 PGC.sin cambio, con tendencia_1

Predictor	Coef	SE Coef	T	P
Constant	3.498	1.703	2.05	0.056
Grupo (antes-después)	10.3236	0.5988	17.24	0.000
PGC.sin cambio, con tendencia_1	0.94075	0.03071	30.63	0.000

S = 1.05276 R-Sq = 99.4% R-Sq(adj) = 99.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	3385.3	1692.7	1527.26	0.000
Residual Error	17	18.8	1.1		
Total	19	3404.2			

Source	DF	Seq SS
Grupo (antes-después)	1	2345.2
PGC.sin cambio, con tendencia_1	1	1040.1

Estos han sido algunos experimentos simulados mediante computadora sobre los posibles resultados que se pueden obtener si no se controlan las amenazas contra la validez. Cada diseño cuasi-experimental se dirige al control de algunas amenazas, y aunque existen procedimientos estadísticos para ajustar estos sesgos, la propuesta del Modelo Arysis se fundamenta en la planificación y previsión analítica de la evaluación de impacto.

4.5. CONTRASTACIÓN DE HIPÓTESIS

En la primera hipótesis se planteó que *“Aunque existe un desarrollo teórico amplio sobre los programas sociales y la evaluación de programas; no existe un nivel teórico comprehensivo de la evaluación de impacto de los*

programas sociales”. Al respecto, los resultados de la investigación permiten confirmarla.

La teoría de la evaluación de impacto es dispersa y no tiene un referente integrado. Generalmente, las publicaciones sobre el tema se centran en aspectos propedéuticos, introductorios, demasiado esquemáticos y no abordan problemas teóricos-metodológicos con la profundidad necesaria. La evaluación de impacto es analizada dentro del contexto general de la evaluación de programas, pero no se le dedica un apartado especial. Los pocos libros en habla hispana (Ej. Baker, 2000), son limitados, se centran en casos, y están un tanto desfasados con relación al desarrollo metodológico actual. En efecto, la revisión biblio-integrativa nos informa que más del 77% de las publicaciones científicas sobre el tema datan *posteriori* al año 2000, por lo que las publicaciones de habla hispana –todas anteriores a esa época- están desactualizadas.

La teoría de la causación –en estos textos- se limita a unas cuantas líneas y los diseños de investigación se circunscriben a dos o tres de naturaleza cualitativa o cuantitativa. Ni siquiera se consideran los principios del evaluador o de la evaluación. No se hace un análisis de las bondades o desventajas de cada diseño ni se demuestran sus limitaciones o alcances. El lector no adquiere una comprensión suficiente para valorar las amenazas contra la validez de los resultados, la calidad y criterios de este tipo de evaluación; ni siquiera desarrolla una actitud positiva hacia ella.

Por otro lado, estas limitaciones bibliográficas contrastan con los procedimientos metodológicos presentes en los artículos e informes de evaluación de impacto publicados recientemente, donde técnicas sofisticadas de análisis, de ajuste de sesgos, de supuestos causales, escapan a la comprensión del gerente social por cuanto no ha obtenido la formación suficiente para entenderlas.

En la segunda hipótesis se formuló que **“Existe un desfase significativo entre el estado actual del desarrollo académico-metodológico de la evaluación de impacto en el ámbito internacional y el desarrollo metodológico nacional”**. Al respecto, los resultados de la investigación permiten confirmar la hipótesis.

En el Perú la evaluación de impacto de los programas no es un proceso tan común como en el extranjero, ni tampoco lo es el desarrollo académico-metodológico sobre el tema; sin embargo, para el caso de los programas financiados por cooperación internacional (Ej. BID, Banco Mundial, USAID, etc.) existen informes bien estructurados. Ello es posible porque estos organismos internacionales son exigentes con la evaluación y, muchas veces, contratan a expertos extranjeros o convocan a empresas nacionales especializadas para realizarlas. En el caso de los programas financiados por el gobierno, la evaluación de impacto es casi inexistente. Salvo algunos programas de formación laboral juvenil, o de infraestructura vial, o de empleo temporal (A trabajar urbano, a trabajar rural, gerenciados por FONCODES), el resto de programas sociales no ha realizado evaluaciones de impacto. Los programas de salud, alimentarios, seguridad, importantísimos para el desarrollo nacional, no tienen retroalimentación oportuna sobre el impacto de sus acciones; por el contrario, las pocas evaluaciones

en estos sectores son guardadas celosamente, catalogados como *top secret* y no se difunden a la ciudadanía.

En el extranjero la situación es diferente. La revisión meta-analítica nos ha demostrado que existe una política fuerte de evaluación, principalmente en Europa y América del Norte. Pero Latinoamérica no se ha quedado atrás y durante la última década ha impulsado significativamente la evaluación de impacto, siendo México, Argentina y Brasil los países donde se realizan más evaluaciones de este tipo. En el Perú, con la implantación del Sistema Nacional de Inversión Pública (SNIP) existe la esperanza de que una política de evaluación fuerte se instale. Sin embargo, tal proceso recién empieza y los manuales y reglamentación que emite aún está a nivel de estudios de pre-inversión. Aun se guarda silencio sobre los estudio de post-inversión y, dentro de ella, sobre la evaluación de impacto.

Similar al ámbito internacional, las EI en el Perú han aumentado geométricamente durante la última década, con una disminución en la década del 90 quizá debido a la coyuntura política.

En el aspecto metodológico, el Perú también está en desventaja. A pesar que diversos especialistas nacionales –que son pocos- dominan los procedimientos de análisis de evaluación de impacto y lo han plasmado en informes detallados, es innegable que la gran mayoría de informes de evaluación nacional adolecen de una serie de falencias y deficiencias metodológicas. En efecto, las EI nacionales subinforman u omiten importantes aspectos metodológicos en sus informes, dando indicios de deficiencias. Igual que en el extranjero, el diseño más frecuente es el cuasi-experimental, aunque en nuestro país no se ha realizado diseños experimentales ni meta-analíticos; si se ha empleado más técnicas de control de sesgos como el matching o las variables instrumentales. En cuanto al impacto de los programas evaluados, en el Perú se reportan menos programas con efectos negativos, pero 24.2% no especifica adecuadamente impacto alguno, hecho debido a las deficiencias metodológicas mencionadas.

Finalmente, en la tercera hipótesis se formuló que ***“Los fundamentos del Sistema de Evaluación de Impacto ARYSIS de los programas sociales en el Perú son: los principios internacionales de evaluación, la teoría de la causación, el diseño cuasi-experimental avanzado y la estadística de simulación vía montecarlo”***. Al respecto, los resultados de la investigación permiten confirmarla.

Debido a que la teoría de la evaluación de impacto está bibliográficamente dispersa, se requería un modelo integrativo fundamentado en tales teorías. El Modelo ARYSIS es resultado de ello. En la revisión bibliográfica se encontró que las Asociaciones Internacionales de Evaluación, los Comités Científicos de evaluación se guían siempre por principios rectores, los cuales son normas de procedimiento que orientan la conducta del evaluador así como su quehacer. Por eso el Modelo ARYSIS necesitaba fundamentarse en tales principios, reconocidos y aceptados por la comunidad científica internacional.

Por otro lado, la teoría general de la causación era el único parámetro posible para fundamentar el Modelo. La teoría de la inferencia causal, el análisis de las amenazas contra la validez de la inferencia, el análisis de los diversos modelos de causalidad y validez son aportes científicos de naturaleza teórica reconocida

mundialmente y que ya tiene un amplio desarrollo dentro de las ciencias experimentales.

Finalmente, la metodología cuasi-experimental, la cual fue creada para medir causalidad en contextos sociales, fue el criterio metodológico ineludible para fundamentar el Modelo. Su versatilidad, flexibilidad, multiplicidad de diseños y posibilidad analíticas, fueron aspectos válidos para su elección. A ello, se le suma la ventaja de la simulación experimental computarizada vía Montecarlo, la cual permite definir y probar el modelo *a priori*, modificarlo y optimizarlo, obteniendo un modelo realista y con el mayor control de las amenazas contra la validez.

CONCLUSIONES

1. A pesar que cada año el Perú destina mayores recursos para el desarrollo social, aún no existe un sistema de evaluación objetivo que permita conocer el impacto preciso que tiene este gasto sobre la población. Ejecutar acciones sin conocer su real efecto trae consigo, con toda certeza, el desperdicio de los recursos, la subcobertura de los programas y un bajo impacto de los mismos sobre la población objetivo. De hecho, aunque existe evidencia de algunos indicadores positivos de la política social, no puede negarse que se desconoce tanto el efecto preciso, como el impacto de las acciones gubernamentales. Además, hasta hoy la evaluación de los programas se ha concentrado en aspectos administrativos básicos como la cobertura y operación, descuidando los objetivos y fines propios de la intervención. De nada sirve saber a cuántas personas atiende el programa o qué se les entrega si no se sabe si realmente los beneficia en algo o –por el contrario- los perjudica. Solo la evaluación de impacto proporciona esa información.
2. La evaluación de impacto identifica de manera sistemática los efectos (positivos o negativos, esperados o no) sobre las personas, hogares e instituciones, generados por un programa de desarrollo social. Es totalmente distinta a las evaluaciones de procesos, de gestión y de resultados, pues es la única que determina la causalidad entre la intervención y sus efectos observados. Así, permite entender la magnitud de los cambios generados por dicha acción sobre el bienestar y, en especial, sobre la reducción de la pobreza. La evaluación de impacto apunta, también, a retroalimentar y mejorar la efectividad de los programas sociales, comparándolos con intervenciones alternativas; sus resultados proveen información para la toma de decisiones relacionadas con mantener, ampliar, reorientar o eliminar un programa existente o diseñar una nueva política pública; así como hacer posible la rendición de cuentas al público.

3. La principal fuente de difusión y producción de EI en el mundo son los Organismos Internacionales, quienes han aportado más del 75% de la producción de informes, centrándose en los países en desarrollo y evaluando aspectos educativos, de salud, pobreza y empleo. Las EI han aumentado geométricamente durante los últimos 30 años, concentrándose más del 77% en el periodo 2000-2006.
4. Existen diversas metodologías para evaluar el impacto de programas sociales, sin embargo, el método más robusto es la evaluación de diseño cuasi-experimental. Este diseño usa tanto encuestas directas aplicadas a los distintos actores involucrados en la ejecución de los programas como información secundaria. Complementariamente, usa las entrevistas a profundidad, los grupos focales, la observación no participante y la revisión documental, para así entender los procesos y condiciones detrás de los impactos observados, así como la percepción y prioridades de los individuos acerca de su bienestar.
5. Un importante porcentaje de EI internacionales (37.5%) utiliza datos secundarios, y en el caso de los primarios, el instrumento más usado es el cuestionario. El tamaño de la muestra es muy variable (33.4% supera las mil unidades) y se centra principalmente en personas y familias. En cuanto al uso de técnicas de control especiales, su uso es muy limitado para el caso de las variables instrumentales (7.6%) y limitado para el *matching* (19.7%). La técnica estadística más empleada es la regresión (38%).
6. Dada su riqueza en términos de información y resultados, las evaluaciones de impacto proveen elementos conceptuales y analíticos que apoyan la toma de decisiones relacionada con los programas evaluados. Considerando que por lo menos 2 de cada 10 programas evaluados no muestran impacto positivo alguno, la EI es una herramienta de gran utilidad para el gobierno en materia de reingeniería de la política social, en la optimización de la inversión y reducción de la pobreza; destinando más recursos a los programas exitosos y redefiniendo aquellos sin impacto.
7. En el Perú la escasa evaluación de impacto de los programas sociales es una constante. Sumado a ello, los pocos informes de evaluación de impacto que se registran en las instituciones públicas tienen carácter de secreto y son guardados recelosamente, atentando contra la naturaleza teleológica de este método. Sólo pocos estudios nacionales no adolecen de significativas deficiencias metodológicas, teóricas y procedimentales, las cuales cuestionan seriamente la validez de los resultados y conclusiones obtenidas. Según los resultados obtenidos, la única fuente de difusión y producción de EI en el Perú son las Instituciones Nacionales y Organismos Internacionales, quienes han aportado el 100% de la producción de informes. La principal área de interés es el empleo y luego la lucha contra la pobreza. Similar al ámbito internacional, las EI en el Perú han

aumentado geométricamente durante la última década, con una disminución en la década del 90 quizá debido a la coyuntura política. En términos metodológicos, las EI nacionales subinforman u omiten importantes aspectos metodológicos en sus informes, dando indicios de deficiencias. Igual que en el extranjero, el diseño más frecuente es el cuasi-experimental, aunque en nuestro país no se ha realizado diseños experimentales ni meta-analíticos; si se han empleado más técnicas de control de sesgos como el *matching* o las variables instrumentales. En cuanto al impacto de los programas evaluados, en el Perú se reportan menos programas con efectos negativos, pero 24.2% no especifica adecuadamente impacto alguno, hecho debido a las deficiencias metodológicas mencionadas.

8. En contraste, la literatura científica internacional se ha desarrollado significativamente sobre la base de la metodología cuasiexperimental (41%, principalmente con el diseño de “diferencias en diferencias”). El avance metodológico en el sentido de mejorar los métodos cuasi-experimentales es muy importante a nivel académico y de la práctica de programas en países desarrollados. Es indudable su importancia con el fin de poder corregir a tiempo o eliminar programas que no alcancen, o lo hagan parcialmente, los objetivos fijados.
9. La aceptación de la necesidad de evaluar los programas con técnicas científicamente robustas es creciente a nivel mundial. De los métodos revisados se puede inferir que la necesidad de contrastar los resultados obtenidos de un conjunto de participantes con un grupo de control, es vital para poder llegar a una conclusión sobre la eficiencia de las intervenciones, pero también para monitorear el cumplimiento de las metas a lo largo de la ejecución de los programas y no únicamente al final. La transparencia o “*accountability*” de los mismos se logra con evaluaciones científicas y robustas desde el punto de vista técnico.
10. La aparición de novedosos desarrollos metodológicos, dentro de los diseños cuasi-experimentales, posibilitan una evaluación de impacto válida. En ese contexto, el uso de la simulación experimental permite la comprensión y análisis *ex ante* de las posibilidades analíticas del diseño de evaluación, optimizándolo, sin mayores gastos. Así, la simulación experimental se convierte en un aliado metodológico en la planificación del diseño de evaluación.
11. El Modelo ARYSIS hace uso de principios y supuestos tanto teóricos como metodológicos. Utilizando la metodología cuasi-experimental más robusta, presenta un sistema teórico-metodológico flexible para modelar los diseños de evaluación de impacto de los programas sociales. La ventaja del modelo radica en su fundamentación teórica, de naturaleza integrativa, que evita una aproximación meramente intuitiva y restringe la arbitrariedad en los resultados. Además, permite una previsión deductiva de los posibles

escenarios en los cuales los resultados pueden acontecer, identificando variables de sesgo, de selección, interacciones de confusión y proponiendo variables instrumentales, control por emparejamiento y otros procedimientos necesarios para evitar la sobreestimación o subestimación de los resultados.

RECOMENDACIONES

1. Se recomienda la institucionalización de la evaluación de impacto como instrumento de apoyo a la toma de decisiones de política social. El Gobierno puede incorporarla como política institucional dentro de la actual Reforma del Estado y simplificación de los Programas Sociales. Esta recomendación es necesaria en la medida que permite el logro de dos fines asociados: el aprendizaje institucional y el uso racional de recursos escasos. De esta forma, servirá para generar aportes estratégicos de política, orientar y ajustar los programas sociales. Los argumentos presentados indican la pertinencia de utilizar la evaluación de impacto como instrumento clave en el proceso de toma de decisiones de la asignación del gasto social. Su utilización contribuirá en la armonización de la oferta programática, en su eficiencia y efectividad, y en la orientación de recursos del presupuesto nacional hacia aquellos programas que generen los mejores resultados en términos de reducción de la pobreza.
2. Se recomienda la difusión del método de evaluación de impacto a través de la actualización curricular de la Gestión Pública y la Gerencia Social. Los gerentes sociales y administradores públicos de programas de inversión necesitan conocer –aunque sea de modo propedéutico– las características y naturaleza de la evaluación de impacto. El currículo de estas especializaciones se debe orientar al desarrollo de una actitud positiva hacia la evaluación de impacto, así como a la formación de criterios de calidad de la misma.
3. El Sistema Nacional de Inversión Pública contempla dentro de su proceso de post-inversión a la evaluación de impacto; sin embargo, aún no ha elaborado guías ni manuales sobre su formulación, ejecución o informe. Por ello, se recomienda la consideración del Modelo ARYSIS para el diseño fundamentado de tales manuales y documentos. No es suficiente que el SNIP provea de material referencial centralizado en estudio de casos. Aparte de ello, debe proveer referencia teórica-metodológica que fundamenten la elección de diseños especializados de evaluación de impacto apropiados para cada sector o programa, pues lo que generalmente ocurre es que los documentos basados en

estudio de casos se usan como “plantilla” de “copia y pega”, con consecuencias negativas obvias para la validez de la inferencia.

4. Finalmente, se recomienda la difusión institucionalizada de las evaluaciones de impacto realizadas en el país. En el extranjero existe una preocupación por la política de “rendición de cuentas”, por ello se publican los informes de evaluación de impacto de los programas sociales, siendo de acceso libre para los interesados. Esta situación, lamentablemente, no ocurre en el país, atentado con el principio de transparencia gubernamental y con el acceso informativo de la ciudadanía. Los informes deben estar abiertos al público y a la prensa, y servir también para la discusión en la Comisión de Presupuesto. Se requiere promover su difusión activamente para fortalecer el proceso de toma de decisiones y mejorar el gasto social.
5. La difusión de estos informes se puede realizar mediante bases de datos en los portales web de cada Ministerio, o mediante la integración de todos los informes en el Sistema Nacional de Inversión Pública, del Ministerio de Economía y Finanzas. El pueblo tiene derecho a saber de la eficacia gubernamental y los investigadores necesitamos esa información para proponer nuevas estrategias y soluciones. Adicionalmente, la mejor forma de asegurar la difusión de las lecciones y el conocimiento adquirido con la evaluación de impacto es mejorar tanto el contenido de los informes como la presentación de los mismos. Un paso en esta dirección sería establecer requerimientos estandarizados a la hora de encomendar las evaluaciones y formatos para la realización de informes, haciendo hincapié en el tipo de valoraciones principales y globales discutidas en la investigación.

REFERENCIAS

1. ABADIE A, ANGRIST J. & IMBENS G. (2002). Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. *Econometrica*, Econometric Society, Vol. 70(1), Pp. 91-117, January.
2. ABDALA, E. (2001). Modelos de evaluación para programas de capacitación de jóvenes. Montevideo: Cinterfor/OIT.
3. ABDALA, E. (2004). Manual para la evaluación de impacto en programas de formación para jóvenes. Montevideo: Cinterfor/OIT.
4. ADATO M, COADY D. & RUEL M. (1999) Evaluación de operaciones de Progresá desde la perspectiva de los beneficiarios, las promotoras, directores de escuela y personal de salud. Instituto Internacional de Investigación sobre Políticas Alimentarias, Washington, DC.
5. ADATO M, DE LA BRIÈRE B, MINDEK D. & QUISUMBING A. (2000). The impact Progresá on women's status and intrahousehold relations. International Food Policy Research Institute. Washington. USA.
6. ADATO M. (2000). El impacto de Progresá sobre las relaciones sociales en la comunidad. International Food Policy Research Institute. Washington. USA Julio.
7. ADATO, M. (2000). Final report: The impact of PROGRESA on community social relationships. September. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, D.C.
8. ADU-GYAMFI JJ, TWUM-AMPOFO K, AKUAMOAH J. & KWAKU A. (2006) ADRA/Ghana's Food Security Program (PL 480 Title II). En. Base de Datos USAID
9. AEDO, C. & NÚÑEZ, S. (2004). The Impact of Training Policies in Latin America and the Caribbean: The Case of Programa Joven. Research Network Working Paper N° 483. Bases de datos electrónica BIF.
10. AGUILAR, M. & ANDER-EGG, E. (1992). Evaluación de servicios y programas sociales. Madrid: Siglo XXI.
11. AHMED, A. & DEL NINNO, C. (2002). The Food for Education Program in Bangladesh: An Evaluation of its impact on Educational Attainment and Food Security. Food Consumption and Nutrition Division Discussion Paper N°138.

- International Food Policy Research Institute, Food Consumption and Nutrition Division, Washington, D.C.
12. AHMED, A., DEL NINNO, C. & CHOWDHURY, H. (2004). Investing in Children through the Food for Education Program. En DOROSH, P., DEL NINNO, C. & SHAHABUDDIN, Q. (Eds.). *The 1998 Floods and Beyond: Towards Comprehensive Food Security in Bangladesh*. Dhaka: The University Press Limited.
 13. AKHTER U. AHMED & MARY ARENDS-KUENNING (2003). Do Crowded Classrooms Crowd Out Learning? Evidence From the Food for Education Programme in Bangladesh, próximo a publicarse. Instituto Internacional de Investigación sobre Políticas Alimentarias, Washington, D.C.
 14. AKUOKO-ASIBEY, A. (1997) Views of selected government officials on the impact of a rural water supply program in Ghana. *Evaluation and Program Planning*, Vol. 20, No. 2, pp. 225-230.
 15. ALATAS, V. & CAMERON, L. (2003). The Impact of Minimum Wages on Employment in a Low Income Country: An Evaluation using the Difference-in-Differences Approach. World Bank Policy Research Working Paper N° 2985. Banco Mundial. Washington.
 16. ALCAZAR, L., XU LIXIN, C. & ZULUAGA, A. (2000). Institutions, Politics and Contracts: The attempt to privatize the Water and Sanitation utility of Lima, Peru. World Bank Policy Research Working Paper N° 2478. The World Bank, Development Research Group, Regulation and Competition Policy, Washington D.C.
 17. ALCAZAR, L.; ABDALA, M. & SHIRLEY, M. (2000). The Buenos Aires Water Concession. World Bank Policy Research Working Paper N°. 2311. Development Research Group, Regulation and Competition Policy, World Bank, Washington D.C.
 18. ALDERMAN, H., BRITTO, P., ENGLE, P. & SIDDIQI, A. (2004). Longitudinal Evaluation of Uganda Nutrition and Early Child Development Program. Working Paper. World Bank. Washington D.C.
 19. ALEXIEVA A, ALEXIEVA K. & VASEV I.(2003) Impact evaluation of the peace corps small project assistance (SPA) project. En: Base de Datos USAID
 20. ALFIE, E. (2003). Diseño metodológico de investigación para la evaluación de resultados en programas sociales. II Congreso Argentino de Administración Pública, Sociedad, Estado y Administración.
 21. ALIAGA, F. (2000). Validez de la Investigación causal. Tipologías y evolución. *Bordón*, 52 (3): 301-321. Disponible en Internet: [<http://www.uv.es/~aliaga/curriculum/Validez.htm>] Acceso el 13 de octubre de 2005.
 22. ALKIN, M. (1990). Debates on evaluation. London. Sage.
 23. ALVERT, B. (2001). Using time-series analysis to evaluate the impact of policy initiatives in child welfare *Evaluation and Program Planning*, 24: 109-117
 24. ALVIRA, F. (1985). La evaluación evaluativa: una perspectiva experimentalista. *Revista Española de Investigaciones Sociológicas*, 29: 129-141.

25. ALWANG J. (2002). The impact of The International Food Policy Research Institute's research program on rural finance policies for food security for the poor. International Food Policy Research Institute. Impact Assessment Discussion Paper N° 16
26. AMERICAN EVALUATION ASSOCIATION (2004). Guiding Principles for Evaluators. USA.
27. ANDERSON GW & VADERVOORT CHG. (1982) Rural Roads Evaluation Summary Report. A.I.D. Program Evaluation Report No. 5 Base de datos electrónica USAID.
28. ANDERSON, P.; THOULESS, D.; ABRAHAMS, E. & FISHER. D. (1980). New method for a scaling theory of localization. *American Physical Society*. 22 (8): 3519–3526.
29. ANDERSON, S. & BALL, S. (1983). The profession and practice of program evaluation. San Francisco, Ca: Jossey-Bass.
30. ANDERSSON, L. (1985) Intervention against loneliness in a group of elderly women: an impact evaluation. *Soc. Sci. Med.* Vol. 20. No. 4. 355-364.
31. ANGELES G, GUILKEY DK. & MROZ TA. (2003) The Effects of Education and Family Planning Programs on Fertility in Indonesia. Measure Evaluation, Working Paper-03-73.
32. ANGELUCCI M (2004) Aid and Migration: An Analysis of the Impact of Progresá on the Timing and Size of Labour Migration. Forschungsinstitut zur Zukunft der Arbeit (IZA). Discussion Paper No. 1187
33. ANGRIST J, BETTINGER E. & KREMER M. (2004). Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. Serie Documentos de Trabajo (US) No. 10713, 1–33. National Bureau of Economic Research, agosto.
34. ANGRIST J. & LAVY V. (2001) New Evidence on Classroom Computers and Pupil Learning. IZA Discussion Paper No. 362
35. ANGRIST, J., BETTINGER, E., BLOOM, E., KING, E. & KREMER. M. (2002). Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *The American Economic Review*, 92 (5): 1535-1558.
36. ANGUERA, M. (1989). Innovaciones en la metodología de evaluación de programas. *Anales de Psicología*. N° 5: 13-42. Murcia: Secretario de publicaciones e intercambio científico.
37. ANGUERA, M., ARNAU, J., ATO, M. MARTINEZ ARIAS, R., PASCUAL, J. & VALLEJO, G. (1995). Métodos de investigación en psicología. Madrid: Síntesis.
38. ANOLIN ALC. (2000). Women and Micro-Finance Programs. Civil Society and Governance Programme, IDS.
39. ANUATTI-NETO F, BAROSSÍ-FILHO M, GLEDSON DE CARVALHO A. & MACEDO R. (2003) Costs and Benefits of Privatization: Evidence from Brazil. Research Network Working Paper N°-455. Bases de datos electrónica BIF.

40. AOS S, LIEB R, MAYFIELD J, MILLER M & PENNUCCI A. (2004) Benefits and costs of prevention and early intervention programs for youth. Olympia: Washington State Institute for Public Policy.
41. APPELA, P., SMITH, R, SCHMEIDLER, J. & RANDELL, J. (2000). Impact of a vocational counselor on employment-related outcomes among methadone patients. *Evaluation and Program Planning*, 23: 437-448.
42. ARAUJO F, VIVEROS AM. & MURPHREY J (1985) Agricultural credit in the dominican republic .AID. Project impact evaluation report N°. 58. Base de datos electrónica USAID.
43. ARCE, R. (2005). Formulación y gestión de políticas públicas. Presentación del Diplomado de Gestión Pública. Escuela de Gerencia Continental. Lima.
44. ARELLANO, A. (2006). Una revisión sobre los métodos de estudio y evaluación en las políticas activas de empleo. Instituto Valeriano de Investigaciones científicas.
45. ARROYO, J. (2001). La función del gobierno, la intersectorialidad y la sociedad civil en salud. En: Políticas de Salud 2001-2006. Consorcio de Investigación Económica y Social. Lima.
46. ASHWORTH K, HARDMAN J, LIU WCH, MAQUIRE S & MIDDLETON S. (2001). Education Maintenance Allowance: The First Year. A Quantitative Evaluation. Department for Education and Employment. Research Report RR257.
47. ASKIN P, CHILDRESS M, DELMARE R, ESTES V, SMITH M. & SULLIVAN G. (1996) Impact evaluation of the NIS Farmer-to-Farmer Program. Base de datos electrónica USAID
48. ATANASIO O, FITZSIMONS, GÓMEZ A, LÓPEZ D, MEGHIR C, MESNARD A. (2006) Child Education and Work Choices in the Presence of a Conditional Cash Transfer Programme in Rural Colombia. The Institute For Fiscal Studies. WP06/13.
49. ATANASIO O. & MESNARD A. (2005) The impact of a conditional cash transfer programme on consumption in Colombia. The Institute For Fiscal Studies. Report Summary Familias 02.
50. ATO, M. (1991). Metodología de la Investigación en Ciencias del Comportamiento. I: Fundamentos. Barcelona: PPU-DM.
51. ATO, M., QUIÑONES, E. ROMERO, A, & RABADÁN, R. (1989). Evaluación de programas: Aspectos básicos. *Anales de Psicología*, 5 (1-2): 1-12.
52. ATTANASIO O, FITZSIMONS E. & GÓMEZ A. (2005). The Impact of a Conditional Education Subsidy on School Enrolment In Colombia. The Institute For Fiscal Studies Report Summary Familias 01.
53. ATTANASIO O, GÓMEZ LC, HEREDIA P. & VERA-HERNÁNDEZ M. (2005). The short-term impact of a conditional cash subsidy on child health and nutrition in Colombia. The Institute For Fiscal Studies Report Summary Familias 03.
54. ATTANASIO O, MEGHIR C. & SANTIAGO A. (2005). Education Choices in México: Using a Structural Model And a Randomized Experiment to Evaluate Progres. The Institute for Fiscal Studies EWP05/01.

55. ATTANASIO O, SYED M. & VERA-HERNÁNDEZ M. (2004). Early Evaluation of a New Nutrition and Education Programme in Colombia. The Institute For Fiscal Studies. Briefing Note No. 44.
56. ATTANASIO OP. & VERA-HERNÁNDEZ M. (2004) Medium and Long Run Effects of Nutrition and Child Care: Evaluation of a Community Nursery Programme in Rural Colombia.
57. ATTANASIO, O. (2004). Baseline Report on the Familias En Accion. Institute of Fiscal Studies. London, UK.
58. ATTANASIO, O., & VERA-HERNANDEZ, M. (2004). Medium and Long Run Effects of Nutrition and Child Care: Evaluation of a Community Nursery Programme in Rural Colombia. Institute of Fiscal Studies, London, UK.
59. ATTANASIO, O., MEGHIR, C. & SANTIAGO, A. (2005). Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. Institute of Fiscal Studies. Working Paper EWP05/01. London, UK.
60. AUCOIN, P. (2005). Decision-Making in Government: The Role of Program Evaluation. Discussion Paper. Marzo 29 de 2005. Artículo disponible en Internet: [\[http://www.tbs-sct.gc.ca/eval/tools_outils/Aucoin/Aucoin_e.asp#4.3.5\]](http://www.tbs-sct.gc.ca/eval/tools_outils/Aucoin/Aucoin_e.asp#4.3.5) Acceso el 27 de enero de 2006.
61. AVERY RJ, BRYANT WK, MATHIOS A, KANG H, BELL B. (2006) Electronic Course Evaluation; Does an on-line delivery system bias student evaluation? Journal of Economic Education. Winter.
62. AWASUM D, SIENCHÉ C. & OBWAKA E. (1999). Break the Silence Talk about AIDS. Johns Hopkins University.
63. AZARANG, M. & GARCÍA, E. (1996). Simulación y análisis de modelos estocásticos. McGrawHill. México D.F.
64. BAKER, J. (2000). Evaluación del impacto de los proyectos de desarrollo en la pobreza. Manual para profesionales. Banco Mundial, Washington D.C.
65. BALL, S. (1997). Unintended Effects in Educational Research. En KEEVES, J. (Ed.). Educational Research, Methodology and Measurement. An International Handbook. 2ª Edición. Londres: Pergamon.
66. BALLART X. & RIBA C. (1995). Impact of Legislation Requiring Moped and Motorbike Riders to Wear Helmets. *Evaluation and Program Planning*, 18 (4): 311-320.
67. BAMBERGER, M. (2000). Integrating Quantitative and Qualitative Methods in Development Research. Washington, D.C. Banco Mundial.
68. BAMBERGER, M. (2006). Realización de evaluaciones de impacto de calidad con limitaciones de presupuesto, tiempo e información. Grupo temático para el análisis, seguimiento y evaluación del impacto en la pobreza Red PREM. Banco Mundial. Washington D.C.
69. BAMBERGER, M.; RUGH, J. & MABRY, L. (2006). Real World Evaluation: Working under Budget, Time, Data and Political Constraints. Thousand Oaks, CA: Sage.
70. BANCO INTERAMERICANO DE DESARROLLO (2000). Economic and Social Report 1998-1999, Washington, D.C., BID.

71. BANCO INTERAMERICANO DE DESARROLLO, Oficina de Evaluación (BID-EVO), (1997). Evaluación: Una herramienta de gestión para mejorar el desempeño de los proyectos.
72. BANCO MUNDIAL (s/f). PovertyNet. Pobreza. Evaluación del Impacto. Documento institucional disponible en Internet: [<http://www.worldbank.org/poverty/spanish/impact/overview/howtoevl.htm>] SIID. SIA. División de Política Social.
73. BANCO MUNDIAL. (2001). *Attacking Poverty*. Washington, D.C. World Development Report 2000/2001.
74. BANERJEE A, COLE S, DUFLO E. & LINDEN L. (2005) *Remedying Education: Evidence from Two Randomized Experiments in India*. by National Bureau of Economic Research, Inc in its series NBER Working Papers N°11904.
75. BARKAT A, KHAN SH, RAHMAN M, ZAMAN S, PODDAR A, HALIM S, RATNA NN, MAJID M, MAKSUD AKM, KARIM A. & ISLAM K. (2002) *Economic and social impact evaluation of the Rural Electrification Program in Bangladesh*. Base de datos electrónica USAID
76. BARNES C, GAILE G. & KIBOMBO R. (2001). *The Impact of Three Microfinance Programs in Uganda*. Washington, D.C. AIMS. Disponible en: <http://www.usaidmicro.org/pubs/aims>
77. BARRY, B. (1965). *Political Argument*, London: Routledge & Kegan Paul.
78. BARTOLOMÉ, M. (1990). Evaluación y optimización de los programas de intervención. *Revista de Investigación Educativa*, 16. 39-61.
79. BARZELAY, M. (2001). La nueva Gerencia Pública. Un ensayo bibliográfico para estudiosos latinoamericanos. *Reforma y Democracia*, 19, CLAD. Disponible en Internet: [<http://www.clad.org.ve/reforma.html>] Acceso el 29 de julio de 2004.
80. BAYA B, SANGLI G. & MAIGA A. (2004). *Measuring the Effects of Behavior Change Interventions in Burkina Faso with Population –Based Survey Result*. JHPIEGO, an affiliate of Johns Hopkins University,
81. BECKER, S. y ICHINO, A. (2002). Estimation of Average Treatment Effects Based On Propensity Scores. *Stata Journal*, 2 (4): 358-377.
82. BEDI, T., BHATTI, S., GINE, X., GALASSO, E., GOLDSTEIN, M. & LEGOVINI, A. (2006). *Impact Evaluation and the Project Cycle*. Doing Impact Evaluation Series N°1. Thematic group on Poverty Analysis, Monitoring and Impact Evaluation. Banco Mundial. DC. Washington.
83. BEEBY, C. E. (1977). *The Meaning of Evaluation*. *Current Issues in Education*, N° 4. Willington.
84. BEHRMAN JR, MARTORELL R. & STEIN AD. (2003). *The Impact of Experimental Nutritional Interventions on Education into Adulthood in Rural Guatemala: Preliminary Longitudinal Analysis*. Second Meeting of the Social Policy Monitoring Network Health and Nutrition November 6-7.
85. BEHRMAN JR, BIRDSALL N. & SZÉKELY M. (2000). *Economic Reform and Wage Differentials in Latin America*. Research Network Working Paper N°-435. Bases de datos electrónica BIF.

86. BEHRMAN JR, BIRDSALL N. & SZÉKELY M. (2001) Pobreza, desigualdad, y liberalización comercial y financiera en América Latina. Research Network Working Paper N°-449. Bases de datos electrónica BIF.
87. BEHRMAN JR, MARTORELL R. & STEIN AD. (2003). The Impact of Experimental Nutritional Interventions on Education into Adulthood in Rural Guatemala: Preliminary Longitudinal Analysis. Second Meeting of the Social Policy Monitoring Network Health and Nutrition November 6-7.
88. BEHRMAN JR, PARKER SW. & TODD PE. (2004). Medium-Term Effects of the Oportunidades Program Package, including Nutrition, on Education of Rural Children Age 0-8 in 1997. Technical Document Number 9 on the Evaluation of Oportunidades 2004. Philadelphia: University of Pennsylvania.
89. BEHRMAN JR, SEGUETA P. & TODD PE. (2000). El impacto de Progresá sobre el rendimiento escolar durante el primer año de operación. Instituto Internacional de Investigación sobre Políticas Alimentarias, Washington, D.C.
90. BEHRMAN JR. & HODDINOTT J. (2000). Evaluación del impacto de Progresá en la talla del niño en edad preescolar. Instituto Internacional de Investigación sobre Políticas Alimentarias, Washington, DC.
91. BEHRMAN, J. & HODDINOTT, J. (2001). Program valuation with unobserved heterogeneity and selective implementation: the Mexican PROGRESA impact on child nutrition. Penn Institute for Economic Research. *Working Paper* 02-006.
92. BEHRMAN, J., PIYALI, S. & TODD, P. (2001). Progressing through Progresá: an impact assessment of a school subsidy experiment of México. Penn Institute for Economic Research. *Working Paper* 01-033.
93. BELL. LA (1995). The Impact of Minimum Wages in México and Colombia. Policy Research Working Paper 1514. The Policy Research Department.
94. BENOLIEL S, ILON L, SUTTON M, KARMACHARYA DM, LAMICHHANE S, RAJBHANDRY P, DER KAFLE B. & GIRI S. (1998) Promoting education girls in Nepal. Impact Evaluation, N° 5. Base de datos electrónica USAID
95. BENOLIEL S, ILON L, SUTTON M, KARMACHARYA DM, LAMICHHANE S, RAJBHANDRY P, DER KAFLE B. & GIRI S. (1998) Promoting education girls in Nepal. Impact Evaluation, N° 5. Base de datos electrónica USAID
96. BERK, R. & RAUMA, D. (1983). Capitalizing on nonrandom assignment to treatments: A regression-discontinuity evaluation of a crime-control program. *Journal of the American Statistical Association*, 78:21-27.
97. BERNBAUM M, FAIR K, MISKE S, MOREAU T, NYIRENDA D, SIKES J, WOLF J, HARBER RB, HARTWELL A. & SCHWARTZ B. (1999) Promoting primary education for girls in Malawi. Impact Evaluation, N° 5. Base de datos electrónica USAID
98. BERNER, H; COOPER, R; GUZMÁN, M. & GUZMÁN, N. (2005). Metodología evaluación de impacto. Chile. División de Control de Gestión.
99. BESSINGER, R., KATENDE, C., & GUTPA, N. (2004). Multimedia campaign exposure effects on knowledge and use of condoms for STI and

- HIV/AIDS prevention in Uganda. *Evaluation and Program Planning*, 27: 397-407.
100. BIBI, S. (1998). Comparing Effects of General Subsidies and Targeted Transfers on Poverty: Robustness Analysis Using Data Set from Tunisia. Working Paper 0125. Economic Research Forum.
 101. BIFULCO, R. (2002) Addressing Self-selection Bias in Quasi-experimental Evaluations of Whole-school Reform A Comparison of Methods. *Evaluation Review*, 26 (5): 545-572.
 102. BISHOP, J. & TREMBLEY, E. (1987). Counseling centers and accountability: Immovable objects, irresistible forces. *Journal of Counseling and Development*, 65, 491-494.
 103. BISQUERRA, R. (1989). Métodos de investigación educativa. Guía práctica. Barcelona: Ceac.
 104. BITLER MP, GELBACH JB. & HOYNES HW. (2004). What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments. Bases de datos electrónica BIF.
 105. BLANET, J et al (eds.) (2000). Learning for Change: Issues and Experiences in Participatory Monitoring and Evaluation, Londres, Intermediate Technology Publications.
 106. BLUE, RN, GALATY M. & GREEN A. (2006) The CEE/SEE Program. Final Draft submitted to The Internacional Center for Not-for-Profit Law. Washington, DC. En: Base electrónica de Datos USAID.
 107. BLUNDELL R, BREWER M. & SHEPHARD A. (2005) Evaluating the Labor Market impact of Working Families` Tax Credit using difference in differences. Institute for Fiscal Studies, junio.
 108. BLUNDELL R, COSTA DIAS M, MEGHIR C, VAN REENEN J. (2003) Evaluating the Employment Impact of a Mandatory Job Search Programme". Centre for Economic Policy Research (CEPR). Discussion Paper No. 3786. Disponible en SSRN: <http://ssrn.com/abstract=394602>
 109. BLUNDELL R, DEARDEN L. & MEGHIR C. (1996). The Determinants and Effects of Work-Related Training in Britain. Londres. Institute for Fiscal Studies, abril.
 110. BLUNDELL R, DUNCAN A, & MEGHIR C. (2002) Evaluating the Working Families Tax Credit. Background Paper for ‘Structural versus Non-Structural approaches to Evaluation Social Policy Monitoring Network IFS, November 20-21. Bases de datos electrónica BIF.
 111. BLUNDELL R, DUNCAN A, MCCRAE J. & MEGHIR C. (2000) The Labour Market Impact of the Working Families’ Tax Credit. *Fiscal Studies*. vol. 21, no. 1, pp. 75–104
 112. BOBONIS GJ, MIGUEL E. & PURI C. (2004). Iron Deficiency Anemia and School Participation. Poverty Action Lab Paper N° 7. March.
 113. BOLTVINIK, J. (2000). Métodos de medición de la pobreza. Una evaluación crítica en Socialis. *Revista latinoamericana de política social*, N° 2, mayo, Buenos Aires.

114. BOOTHROYD, P. (1998). Social Assessment Research: The establishment, the underground a state of the art. International Development Research Center.
115. BORUCH, R. & GOMEZ, H. (1977). Sensitivity, bias and theory in impact evaluations. *Professional Psychology*, 8: 411-434.
116. BOUILLON CP. & TEJERINA L. (2006) Do we know what works? A Systematic Review of Impact Evaluations of Social Programs in Latin America and the Caribbean. Inter-American Development Bank. Working Paper
117. BOURGUIGNON F, FERREIRA HG. & LEITE PG. (2003). Condicional Cash Transfers, Schooling and Child Labor: Micro-Simulating Bolsa Escola. Texto para Discussão N° 477. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economía.
118. BOURGUIGNON, F., De MELO, J. & SUWA, A. (1991). Distributional Effects of Adjustment Policies: Simulations for Archetype Economies in Africa and Latin America. *World Bank Economic Review*, 5 (2):339-66.
119. BOVENS, M.; HART, P. & KUIPERS, S. (2005). The politics of policy evaluation. En: MORAN, M., REIN, M. & GOODIN, R. (Ed.). *The Oxford handbook of public policy*. Cap. 15. Pp. 317-333.
120. BOX, G. & JENKINS, G. (1976). *Time Series Analysis. Forecasting and Control*. Revised Edition. Oakland. California: Holden Day.
121. BOX, G. & TIAO, G. (1965). A change in level of nonstationary time series. *Biometrika*, 52: 181-192.
122. BOX, G. & TIAO, G. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of American Statistical Association*, 70: 70-79.
123. BRADSTOCK, A. (2005) Land reform and its Impact on Livelihoods: Evidence from eight land reform groups in the Northern Cape Province of South Africa. Policy & Research Series N° 4. Londres. Farm-África.
124. BRAUN M.; GIUGLIANI, E., MATOS, M., GIUGLIANI, C., PROENCO, A. & MACHADO, A. (2003). Evaluation of the impact of the Baby-Friendly Hospital Initiative on Rates of Breastfeeding. *American Journal of Public Health*, 93 (8): 1277-1279.
125. BRINKERHOFF DW, FOTZO PT, ORMOD BJ. (1983) Haiti: Hacho Rural Community Development. AID Project Impact Evaluation Report N° 49. Base de datos electrónica USAID
126. BRIONES, G. (1985) Evaluación de programas sociales. Santiago: PIIE.
127. BROWN L, COX, GB, JONES WE, SEMKE J, ALLEN DG, GILCHRIST L. & SUTPHEN-MROZ, J. (1994) Effects of mental health reform on client characteristics, continuity of care and community tenure. *Evaluation and Program Planning*, Vol. 17, No. I, pp. 63-72.
128. BRUGIAVINI, A. & PERACCHI, F. (2005) Fiscal implications of Pension Reforms in Italy. Research. Centre for International Studies on Economic Growth.(CEIS) Paper Series, Vol. 23, N° 67.
129. BRYK, A. (1978). Stakeholder-based evaluation. San Francisco: Jossey - Bass.

130. BRYSON, A.; DORSETT, R. & PURDON, S. (2002). The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies. *Working Paper* N°. 4. Policy Studies Institute, U.K. Department of Work and Pensions.
131. BUENDÍA, L.; COLÁS, P & HERNÁNDEZ, F. (1997). Métodos de investigación en Psicopedagogía. Madrid: McGraw-Hill.
132. BURCHFIELD S, HUA H, BARAL D. & ROCHA V. (2002). A Longitudinal Study of the Impact of Integrated Literacy and Basic Education Programs on Women's Participation in Social and Economic Development in Nepal. Girls' and Women's Education Policy Research Activity. USAID.
133. BUSTELO, M. (2004). El potencial impacto del establecimiento de normas, estándares y códigos en la creación de una cultura de evaluación. IX Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Madrid, España, 2-5 de Noviembre.
134. CAIDEN, G. (1998). Enfoques y lineamientos para el seguimiento, la medición y la evaluación del desempeño en programas del sector público. *Reforma y Democracia*, 12, CLAD. Disponible en Internet: [<http://www.clad.org.ve/reforma.html>] Acceso el 26 de agosto de 2004.
135. CALDERÓN C, CHONG A. & VALDÉS R. (2004) Labor Market Regulations and Income Inequality. Research Department Working Paper N°-514. Bases de datos electrónica BIF.
136. CALDERÓN-MADRID, A. (2006) . Revisiting the Employability Effects of Training Programs for the Unemployed in Developing Countries. Working Paper N° R-522
137. CALDÉS N, COADY D. & MALUCCIO JA. (2004) The cost of poverty alleviation transfer programs: a comparative analysis of three programs in Latin America. Food Consumption and Nutrition Division discussion Paper N°. 174 International Food Policy Research Institute. Washington DC.
138. CAMERON, L. (2002). Did Social Safety Net Scholarships Reduce Drop-Out Rates during the Indonesian Economic Crisis? World Bank Policy Research Working Paper N° 2800. Poverty Team, Development Research Group, World Bank, Washington D.C.
139. CAMPBELL, D. (1957). Factor Relevant to the Validity of Experiments in Social Settings. *Psychological Bulletin*, 54 (4): 297-312.
140. CAMPBELL, D. & FISKE. D. (1959). Convergent and Discriminant Validation by the Multitrat-Multimethod Matrix. *Psychological Bulletin*, 56: 81-105.
141. CAMPBELL, D. & OVERMAN, E. (1988). Methodology and epistemology for social science: Selected papers. Chicago: University of Chicago Press.
142. CAMPBELL, D. & STANLEY, J. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally (Traducción española: Diseños experimentales y cuasiexperimentales en la investigación social. Buenos Aires: Amorrortu. 1973, 1982).

143. CAMPBELL, D. (1974). Qualitative knowing in action research. Comunicación presentada en la Reunión anual de la American Psychological Association, Los Angeles California.
144. CAMPBELL, D. (1986). Relabeling internal and external validity for applied social scientists. In W. Trochim (Eds.) *Advances in quasi-experimental design and analysis* (pp. 67-78). San Francisco: Jossey-Bass.
145. CANTON, E. & BLOM, A. (2004). Can Student Loans Improve Accessibility to Higher Education and Student Performance? An Impact Study of the Case of SOFES, Mexico. World Bank Policy Research Paper N° 3425. World Bank, Washington D.C.
146. CARDOSO, E. & PORTELLA, A. (2004). The Impact of Cash Transfers on Child Labor and School Attendance in Brazil. Working Paper N° 04-W07. Vanderbilt University, Department of Economics, Nashville, TN.
147. CARDOZO, M. (2003). Evaluación de Políticas de Desarrollo Social. *Política y Cultura*, 20: 139-154.
148. CARE (2004). Jaaa Wazima Project: A Household Livelihood Security Project in Siaya and Busia Districts. Report Prepared for CARE Kenya by Strategic Public Relations and Research Ltd. P.O Box 7201 00100, Nairobi.
149. CARIDE, J. (1989). De la evaluación de necesidades a la evaluación de programas sociales en el desarrollo comunitario. En MARÍN IBÁÑEZ, R & PÉREZ SERRANO, G. (Eds). *Investigación en animación sociocultural*. UNED. Madrid, pp. 133-152.
150. CARRERA J, CHECCHI D, & FLORIO M. (2005) Privatization discontent and its determinants::evidence from Latin América. JEL Numbers: H 32, G 14, L 33. Bases de datos electrónica BIF
151. CASTRO MF. (2004). Impact evaluation of Empleo en Acción: Lessons and relevante for Colombia. The third Meeting of the Social Policy Network. Departamento Nacional de Planeación. República de Colombia.
152. CASTRO, G.; CHAVES, P. (1994). Metodología Evaluación de impacto de proyectos sociales. UNESCO. Unidad Regional de Ciencias Humanas y Sociales para América Latina y el Caribe. Caracas.
153. CAZARES, A. & BEATTY, L. (Eds.) (1994). *Scientific methods for prevention intervention research*. Rockville, MD: National Institute on Drug Abuse.
154. CENTRE FOR CONFLICT RESOLUTION (2004) Centre for Conflict Resolution: Impact Evaluation Report. Kenya. Nyamakoroto House. Base de datos electrónica USAID.
155. CENTRE FOR REVIEWS AND DISSEMINATION. (2001). *Undertaking Systematic Reviews of Research on Effectiveness*. CRD Report Number 4. 2° Ed. Marzo.
156. CEPAL (1998). *Gestión de programas sociales para América Latina*. Volumen I. Serie Políticas Sociales 25. Comisión Económica para América Latina y el Caribe. Santiago de Chile.
157. CHACALTANA, J. (2003). El impacto del Programa "A Trabajar Urbano". Ganancias de ingresos y utilidades de las obras. Consorcio de investigación

- Económica y Social y Centro de Estudios para el Desarrollo y la Participación. Lima.
158. CHACÓN, S. & LÓPEZ, J. (1993). Metodología de la evaluación de programas de intervención: una aplicación en centros asistenciales infantiles. *Apuntes de Psicología*, 37: 41-60.
 159. CHAN TY. & HAMILTON BH. (2003) Learning, private information and the economic evaluation of randomized experiments. *Journal of Political Economy*.
 160. CHASE, R. & SHERBURNE-BENZ, L. (2001). Household Effects of African Community Initiatives: Evaluating the Impact of the Zambia Social Fund. World Bank, Washington D.C.
 161. CHASE, ROBERT S. (2002). Supporting Communities in Transition: The Impact of the Armenian Social Investment Fund. *The World Bank Economic Review*, 16 (2): 219-240.
 162. CHAUDHURY, N., HAMMER, J. & MURRUGARRA, E. (2003). The effects of a Fee-Waiver Program on Health Care Utilization among the Poor: Evidence from Armenia. World Bank Policy Research Working Paper N° 2952. World Bank, Development Research Group and Human Development Department, Europe and Central Asia, Washington D.C.
 163. CHEE G, FIELDS R, HSI N. & SCHOTT W, (2004). Evaluation of GAVI Immunization Services Support Funding. Abt Associate Inc.
 164. CHELIMSKY, E.; SHADISH, W. (1997). Evaluation for the 21st Century. A Handbook. USA: Sage Publications.
 165. CHEN S. & RAVALLION M. (2003). Hidden Impact? Ex-Post Evaluation of an Anti-Poverty Program. World Bank Policy Research Working Paper 3049.
 166. CHEWYND E. & DWORKIN DM. (1981) Korean Potable Water System Project: Lessons from Experience. Project Impact Evaluation No.20. Base de datos electrónica USAID
 167. CHIFUNYISEA, T., BENOYB, H. & MUKIIBI, B. (2002). An impact evaluation of student teacher training in IHV/AIDS education in Zimbabwe. *Evaluation and Program Planning*, 25: 377-385.
 168. CHOKSI, A. (1995). Evaluation in the Bank: taking stock of 50 years of development work. En: Evaluation and Development, Proceedings of the 1994 World Bank Conference.
 169. CHONG A, GALDO V. & TORERO M. (2005). Does Privatization Deliver? Access to Telephone Services and Household Income in Poor Rural Areas Using a Quasi-Natural Experiment in Peru. Latin American Research Network. Red de Centros de Investigación N° 535.
 170. CHONG A. & GALDO J. (2006) Does the Quality of Training Programs Matter? Evidence from Bidding Processes Data. Latin American Research Network. Red de Centros de Investigación N° 555.
 171. CLAD (1999). Una Nueva Gestión Pública para América Latina. Informe Institucional. Disponible en Internet: [<http://www.clad.org.ve/reforma.html>] Acceso el 26 de agosto de 2004.

172. COADY D, DAI X. & WANG L. (2001). Community Programs and Women`s Participation: The Chinese Experience. Policy Research Working Paper 2622. The World Bank
173. COADY D. (2000). La aplicación del análisis social costo-beneficio a la evaluación de Progres. Internacional Food Policy Research Institute.Washington, DC.
174. COADY DP. & LEE R. (2000). Análisis del equilibrio general del impacto de las transferencias de Progres. sobre el bienestar. Internacional Food Policy Research Institute.Washington, DC.
175. COADY DP. & PARKER SW. (2002). A cost-effectiveness analysis of demand- and supply-side education interventions: the case of Progres. in México. Food Consumption and Nutrition Division. Discussion Paper N° 127. Internacional Food Policy Research Institute.Washington, DC.
176. COADY DP. (2001). An evaluaron of the Distributional Power of PROGRESA`S Cash. Bases de datos electrónica BIF.
177. COADY, D., OLINTO, P. & CALDES, N. (2003). Coping with the Coffee Crisis in Central America: The Role of Social Safety Nets in Honduras. International Food Policy Research Institute, Washington D.C.
178. COCHRAN, W. (1983). Planning and Analisis of Observational Studies (ed.), L. E. Moses y F. Mosteller. New York, NY: Wiley.
179. COHEN, E. (2001). Reforma del Estado, modernización administrativa y evaluación del desempeño de la gestión. En Los desafíos de la reforma del Estado en los programas sociales: tres estudios de caso, Santiago de Chile, CEPAL (Serie Políticas Sociales, N° 45).
180. COHEN, M. & SNODGRASS, DR. (2002) Clients in context: The impacts of microfinance in three countries. Assessment the Impact of Microenterprise Services (AIMS). Washington.
181. COLÁS, M. & REBOLLO, M. (1993). Evaluación de programas: una guía práctica. Sevilla, Kronos.
182. COLÁS, M^aP. & BUENDÍA, L. (1994). Investigación educativa. 2^a edición. Sevilla: Alfar.
183. COLEMAN, F. (1972). The evaluation of Equality of Educational Opportunity. En F. MOSTELLER & D. MOYNIHAN (Eds.). On Equality of Educational Opportunity. Nueva York: Random House
184. CONE, J. (1977). The relevance of reliability and validity for behavioral assessment. *Behavior Therapy*, 8: 411-426.
185. Consejo Nacional de Política Económica y Social - CONPES (2002). Evaluación de impacto de programas sociales. Bogotá-Colombia. Documento N° 3188.
186. CONTRERAS D, FLORES L. & LOBATO F. (2003). Monetary Incentives for Teachers and School Performance. The evidence for chile. Departament of Economics University of Chile.
187. COOK D, MULROW C, HAYNES R. (1997). Systematic reviews: synthesis of best evidene for clinical decisions. *Annals Internal Medical*, 126: 376-380.

188. COOK, C., DUNCAN, T., JITSUCHON, S., SHARMA, A & GUOBAO, W. (2004). Assessing the Impact of Transport and Energy Infrastructure on Poverty Reduction. Regional Assistance Technical Report N° 5947. Asian Development Bank, Manilla, Philippines.
189. COOK, T. & CAMPBELL, D. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 28: 141-180.
190. COOK, T. & CAMPBELL, D. (Eds.). (1979). Quasi-experimentation: Design and analysis for field settings. Chicago: Rand McNally.
191. COOK, T. & SHADISH JR., W. (1986). Program Evaluation: The World Science. *Annual Review of Psychology*, 37: 139-232
192. COOK, T. & SHADISH, W. (1994). Social experiments: some developments over the past fifteen years. *Annual Review of Psychology*, 45: 545-580.
193. COOK, T.; CAMPBELL, D. & PERACCHIO, L. (1990). Quasi Experimentation. En M. DUNNETTE & L. HOUGH (Eds.). Handbook of industrial and Organizational Psychology. Vol.1, pags. 491-576. 2ª Edición. Palo Alto (CA). Consulting Psychologist Press.
194. COOK, T.; LEVITON, L. & SHADISH, W. (1985). Program evaluation. En: LINDZEY, G & ARONSON, E. (Eds.). Handbook of social psychology. Nueva York, Random House.
195. COOKE & KOTHARY (2001). Participation: The New Tyranny? Londres, Zed Books.
196. CORDRAY, D. (1986). Quasi-Experimental Analysis: A Mixture of Methods and Judgment. En: W. TROCHIM (Ed.). Advances in Quasi-Experimental Design and Analysis. San Francisco: Jossey-Bass.
197. COWLES, M. (1989). Statistic in Psychology. An Historical Perspective. Hillsdale, New Jersey: LEA.
198. COYLE, S.; BORUCH, R. & TURNER (1991). Evaluating AIDS prevention programs: expanded edition. Washington DC: National Academy Press.
199. CREATIVE ASSOCIATES & BENCHMARKS, INC. (1985). Small Project Assistance (SPA): A Joint Project of AID and the Peace Corps. Washington. Base de datos electrónica USAID - A.I.D. Evaluation Special Study No. 24.
200. CREATIVE ASSOCIATES INTERNATIONAL (1996). Reinsertion of ex-combatants in El Salvador. USAID. National Reconstruction Secretariat.
201. CRONBACH, L. (1963). Course improvement through evaluation. *Teachers College Record*, 64: 672-683.
202. CRONBACH, L. (1974). Beyond the two disciplines of the scientific psychology. Comunicación a la asamblea de la APA, 2 de Septiembre. (Traducción española en F. ALVIRA, M. AVIA, R. CALVO & F. MORALES (Eds.). Los dos métodos de las Ciencias Sociales. Madrid: C.I.S.).
203. CRONBACH, L. (1982). Designing evaluations of educational and social programs. San Francisco, Jossey-Bass.
204. CRONBACH, L.; HAMBRON, S., DORNBUSCH, S., HESS, R., HORNICK, R., PHILLIPS, D., WALKER, D. & WEINER, S. (1980). Towards

- reform in program evaluation: Aims, methods and institutional arrangements. San Francisco: Jossey-Bass.
205. CUANTO (2000). Perú: Informe final de evaluación del proyecto de Caminos Rurales. Lima: Ministerio de Transporte, Comunicaciones, Vivienda y Construcción. Lima.
 206. CUETO, B & MATO, J. (2004). El matching como técnica de evaluación de políticas: una aplicación a los programas de fomento del empleo. VII Encuentro de Economía Aplicada. 3,4 y 5 de junio.
 207. CUETO, S. & CHINEN, M. (2000). Impacto educativo de un programa de desayunos escolares en escuelas rurales del Perú. Lima: GRADE, Documento de trabajo, 34.
 208. CUNILL, N. & OSPINA, S. (2003). Evaluación de Resultados para una Gestión Pública moderna y democrática. CLAD-MAP-AECI.
 209. CURTIS V, KANKI B, COUSENS S, DIALLO I, KPOSEHOUE A, SANGARÉ M. & NIKIEMA M. (2001). Evidence of behaviour change following a hygiene promotion programme in Burkina Faso. *Bulletin of the World Health Organization*, 79: 518–527.
 210. DAR, A., TZANNATOS, Z. (1999). Active Labour Market Programs: A Review of the Evidence from Evaluations. Social Protection Discussion Paper, 9901. Washington, D.C.: World Bank.
 211. DAVISON, A. & HINCKLEY, D. (1997). Bootstrap methods and their application. New York: Cambridge University Press.
 212. DE GIORGI G. (2005) Long Term Effects of a Mandatory Multistage Program: The New Deal for Young People in the UK. The Institute For Fiscal Studies WP05/08.
 213. DE LA ORDEN, A. (1993) La evaluación educativa. Proyecto CINAIE. Buenos Aires.
 214. DE MIGUEL, M. (1998). Modelos y diseños en la evaluación de Programas. Documento fotocopiado.
 215. DE MIGUEL, M. (1999). La evaluación de programas. *Revista de Investigación Educativa*, 17 (2): 345-348.
 216. DE MIGUEL, M. (2000). Evaluación externa de un programa de educación social. En PÉREZ SERRANO, G. (Cord.). Modelos de investigación cualitativa en educación social y animación sociocultural. Aplicaciones prácticas. Madrid: Narcea. Pp. 287-319.
 217. DE RUS, G. & INGLADA, V. (1997) Cost-Benefit Analysis of the High-Speed Train in Spain. *Annals of Regional Science*, Springer Berlin Heidelberg. Vol. 31, Nº 2: 175-188.
 218. DE VREESE, CH. (2004). Primed by the Euro: The Impact of a Referendum Campaign on Public Opinion and Evaluations of Government and Political Leaders. *Scandinavian Political Studies*, 27: 45-64.
 219. DEARDEN L, EMERSON C, FRAYNE C. & MEGHIR C. (2005). Education subsidies and School Drop-Out Rates. The Institute For Fiscal Studies WP05/11.

220. DEARDEN L, EMERSON C, FRAYNE C. & MEGHIR C. (2005). Education subsidies and School Drop-Out Rates. The Institute For Fiscal Studies WP05/11.
221. DEARDEN L. & HEATH A. (1996) Income Support and Staying in School: What Can We Learn from Australia's AUSTUDY Experiment?. Fiscal Studies. vol. 17, no. 4, pp. 1-30
222. DEBONO, E. (1999). Six Thinking Hats, Londres, Backbay.
223. DEERY HA, DAY LM. & FILDES BN. (2000) An impact evaluation of a falls prevention program among older people Accident Analysis and Prevention 32 (2000) 427-433.
224. DEHEJIA, R., WAHBA, S. (1999). Causal effects in nonexperimental studies: reevaluating the evaluation training programs. Journal of the American Statistical Association 94, 1053-1062.
225. DELGADO-RODRÍGUEZ, M. (2002). Glosario de Metanálisis. Panacea. 3(8): 19-22.
226. DENNIS M, SCOTT CK, FUNK R. (2003). An experimental evaluation of recovery management checkups (RMC) for people with chronic substance use disorders. An experimental evaluation of recovery management checkups (RMC) for people with chronic substance use disorders. *Evaluation and Program Planning*, 26: 339-352.
227. DERRIENNIC Y, WOLF K. & KIWANUKA-MUKIIBI P. (2005). An Assessment of Community- Based Health Financing Activities in Uganda. Bethesda, MD: The Partners for Health Reformplus Project, Abt Associates Inc.
228. DEUTSCH, R. (1998). Does Child Care Pay?: Labor Force Participation and Earnings Effects of Access to Child Care in the Favelas of Rio de Janeiro. Banco Interamericano de Desarrollo Oficina del Economista Jefe Documentode Trabajo N° 384
229. DEVELOPMENT ASSISTANCE COMITÉ (2002). Glosario de los principales términos sobre evaluación y gestión basada en resultados. París. Disponible en internet [www.oecd.org/dac/evaluation] Acceso el 26 de agosto de 2004.
230. DIAGNE A. & ZELLER M. (2001). Access to Credit and Its Impact on Welfare in Malawi Research Report 116 International Food Policy Research Institute Washington, D.C.
231. DICCIONARIO DE LA LENGUA ESPAÑOLA (22ª ED.) de la Real Academia Española. Editorial: ESPASA-CALPE SA, 2001. Dos volúmenes.
232. DIVISIÓN DE CONTROL DE GESTIÓN (2003). Evaluación de Programas Gubernamentales (EPG). Ministerio de Hacienda. Santiago de Chile.
233. DOLLAR D. & LEVIN V. (2005) Sowing and Reaping: Institutional Quality and Project Outcomes in Developing Countries. World Bank Policy Research Working Paper 3524.
234. DOROSH P, HAGGBLADE S, LUNGREN C, RAZAFIMANANTENA T. & RANDRIANMIARANA Z. (2003) Economic Motors for Poverty Reduction in Madagascar. Working Paper INSTAT – USAID.

235. DOWNES A, MAMINGI M. & BELLE ANTOINE RB (2000). Labor Market Regulation and Employment In the Caribbean. Latin American Research Network Red de Centros de Investigación Research network Working paper N° -388. Bases de datos electrónica BIF.
236. DRYDEN A, QUERAL M. & TAUCHEN H. (2001). Evaluation of Rhode Island's Family Independence Program. Wellesley College Department of Economics. Wellesley Child Care Research Partnership.
237. DU BOIS, F. (2004). Programas sociales, salud y educación en el Perú. Un balance de las políticas sociales. Democratizando el gasto social. Instituto Peruano de Economía Social de Mercado & Fundación Konrad Adenauer Stiftung. Lima.
238. DUFLO E, DUPAS P, KREMER M. & SNEI S. (2006) Education and HIV/AIDS Prevention: Evidence from a randomized evaluation in Western Kenya. Poverty Action Lab. Working Paper, February.
239. DUFLO E. & HANNA R. (2006). Monitoring Works: Getting Teachers to Come to School. BREAD Working Paper No. 103.
240. DUFLO, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *The American Economic Review*, 91 (4): 795-813.
241. DUGGAN A, FUDDY L, BURRELL L, HIGMAN SM, MCFARLANE E, WINDHAM A. & SIA C. (2002). Randomized trial of a statewide home visiting program to prevent child abuse: impact in reducing parental risk factors. *Child Abuse and Neglect*. Vol. 28, N° 6, 625-645.
242. DUNN, W. (1982). Reforms as arguments. *Knowledge. Creation, Diffusion, Utilization*, 3: 293-326.
243. DUPAS P. (2006). Relative Risks and the Market for Sex: Teenagers, Sugar Daddies and HIV in Kenya. En Poverty Action Lab. Mayo.
244. EBRAHIM R. (1996) .United States Agency for Intenarional Development, ATechnical Report 12. Base de datos electrónica USAID
245. EDWARD, M. & KREMER, M. (2001). Worms: Education and Health Externalities in Kenya. National Bureau of Economic Research Working Paper N° 8481. National Bureau of Economic Research, Cambridge, MA.
246. EDWARD, M. & KREMER, M. (2003). Health Behavior and the Design of Public Health Programs: Evidence from Randomized Evaluations. World Bank, Washington D.C.
247. EFRON, B. (1979). Bootstrap methods. *Annals of statistics*, 7: 1-26.
248. EFRON, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap, and other resampling methods. *Biometrika*, 68: 589-599.
249. EFRON, B. (1987). Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, 82: 171-20.
250. EFRON, B., & TSIBIRANI, R. (1993). An introduction to the Bootstrap. N.Y.: Chapman & Hall
251. ELÍAS VJ, RUIZ F, COSSA R., & BRAVO B. (2004). An Econometric Cost-Benefit Analysis of Argentina's Youth Training Program. Latin American

- Research Network Red de Centros de Investigación Research Network Working Paper #R-482. Bases de datos electrónica BIF.
252. EMMERSON C, FRAYNE C., MCNALLY S. & SILVA O. (2005). Evaluation of Aimhigher: Excellence Challenge The Early Impact of Aimhigher: Excellence Challenge on Pre-16 Outcomes: An Economic Evaluation The Institute For Fiscal Studies. Research Report RR652.
 253. EZEMENARI, K. & SUBBARAO, K. (1998). Jamaica's Food Stamp Program: Impacts on Poverty and Welfare. Policy Research Working Paper N° 2207. World Bank, Poverty Reduction and Economic Management Network, Poverty Division, Washington D.C.
 254. EZEMENARI, K.; RUQVIST, A. & SUBBARAO, K. (1999). Impact Evaluation: A Note on Concepts and Methods. World Bank Poverty Reduction and Economic Management Network, procesado. Washington, D.C.: Banco Mundial.
 255. FARM COMMUNITY TRUST OF ZIMBABWE (2001). The Impacto of Land Reform on Comercial Farm Workers' Livelihoods. Report Date: Report 9th.
 256. FAUTH G. & DANIELS B. (2001) Youth Reintegration Training and Education For peace (YRTEP) Program. United States Agency for International Development Office of Transition Initiatives. En base electrónica USAID
 257. FEDER G, MURGAI R. & QUINZON J. (2004) Sending Farmers Back to School: The Impact of Farmer Field Schools in Indonesia. Review of Agricultural Economics, vol. 26. N° 1: 45-62
 258. FEIN, D. (2001). Will welfare reform influence marriage and fertility? Early evidence from the ABC demonstration. *Evaluation and Program Planning*, 24: 427-444.
 259. FEIN, D. (2001). Will welfare reform influence marriage and fertility? Early evidence from the ABC demonstration. *Evaluation and Program Planning*, 24: 427-444.
 260. FÉRNÁNDEZ-BALLESTEROS, R. (1996). Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de la salud. Madrid: Síntesis Psicológica.
 261. FERRANDO D, SERRANO N. & PURE C. (2002) Midterm Impact Evaluation of the Reproductive health in the Community (reprosalud) Project. The United States Agency for International Development in Perú.. Base de datos electrónica USAID
 262. FETTERMAN, D. (2000). Foundations of Empowerment Evaluation. Thousand Oaks, CA: SAGE.
 263. FINK, A. (1993). Evaluation Fundamental. Newbury Park, Sage.
 264. FISHER, F. (1997). Evaluating Public Policy. Chicago: Ed. Nelson Hall Publishers.
 265. FISHMAN, G. (1996). Monte Carlo, concepts, algorithms and applications. Springer.
 266. FLORES R, MORRIS S, OLINTO P, MEDINA J. & NEIDECKER O. (2003) Evaluation of the Family Allowance Program (PRAF) in Honduras:

- Health and Nutrition Impacts. International Food Policy Research Institute, November.
267. FLORES R. (2003). Impact Evaluation of the Pilot Phase of the Nicaraguan. Red de Protección Social, FCN Discussion Paper, IFPRI, Washington D.C. (forthcoming).
 268. FLORIO, M (2002) A State without ownership: the welfare impact of British privatisations 1979-1997. Department of Economics University of Milan Italy in its series Departmental Working Papers with number 2002-24
 269. FORREST S, MASTERS H. & MILNE V. (2004). Evaluating the impact of training in psychosocial interventions: a stakeholder approach to evaluation – part II. *Journal of Psychiatric and Mental Health Nursing*, 11: 202–212.
 270. FRAKER, T. & MAYNARD, R. (1987). Evaluating Comparison Group Designs with Employment-Related Programs. *Journal of Human Resources*, 22: 194–227.
 271. FRANCESCONI, M. & VAN DER KLAUW, W. (2004) The Consequences of ‘In-Work’ Benefit Reform in Britain: New Evidence from Panel Data. IZA Discussion Paper No. 1248.
 272. FRANKENBERG E. & DUNCAN T. (2001) Women’s health and pregnancy outcomes: do services make a difference? *Demography*. Vol.38. Nº 2, 253-265.
 273. FRECHTLING, J., STEVENS, F., LAWRENZ, F. & SHARP, L. (1993). The User-Friendly Handbook for Project Evaluation: Science, Mathematics and Technology Education. NSF 93-152. Arlington, VA: NSF.
 274. FREEMAN, H. & ROSSI, P. (1993). Evaluation: A Systematic Approach. Sage Publications, Newberry Park, C.A
 275. FRIEDKIN T, LESTER R, BLANK H. & AHMED N. (1983) Bangladesh Small-Scale Irrigation. ALD Project Impact Evaluation Report Na 42. Base de datos electrónica USAID
 276. FRIEDLANDER, D. & ROBINS, P. (1995). Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods. *American Economic Review* 85 (4): 923–937.
 277. FULOP N, ELSTON J, HENSHER M, MCKEE M. & WALTERS R. (2000) Health Policy. Lessons for Health Strategies in Europe: The Evaluation of a National Health Strategy in England . *European Journal of Public Health*, vol. 10. Nº 1.
 278. FURUKAWA, SH. & HOSHINO, Y. (2001). Knowledge-Based Governance by Performance Measurement. *The Japanese Journal of Evaluation Studies*, 1 (2): 13-27.
 279. GALABAWA, J., OBELEAGU, A. & MIYANAZAWA, I. (2002). The impact of school mapping in the development of education in Tanzania: an assessment of the experiences of six districts. *Evaluation and Program Planning*, 25: 23-33.
 280. GALASSO M. & RAVALLION M. (2002). Decentralized Targeting of an Anti-Poverty Program. *Journal of Public Economics*, 85: 705-727

281. GALASSO, E. & RAVALLION, M. (2003). Social Protection in a Crisis: Argentina's Plan Jefes y Jefas. World Bank Policy Research Working Paper N° 3165. Development Research Group. World Bank, Washington D.C.
282. GALASSO, E., RAVALLION, M. & SALVIA, A. (2001). Assisting the Transition from Workfare to Work: A Randomized Experiment. World Bank Policy Research Working Paper N° 2738. World Bank, Washington D.C.
283. GALDO, J. (2000). Una Metodología de Evaluación de Impacto de los Programas de Educación y/o Capacitación en el Mercado Laboral: el Caso de ProJoven. En Impacto de la Inversión Social en el Perú, Ed. Enrique Vásquez, Lima CIUP-IDRC.
284. GALDO, V. & BRICENO, B. (2004). Evaluating the Impact on child Mortality of a Water Supply Project and Sewerage Expansion in Quito, Ecuador. Inter-American Development Bank, Office of Evaluation and Oversight, Washington D.C.
285. GALIANI, S. & SCHARGRODSKY, E. (2002). Evaluating the Impact of School Decentralization on Educational Quality. *Economía*, 2 (2):275-314.
286. GALIANI, S. & SCHARGRODSKY, E. (2005). Effects of Land Titling. Business School Working Papers proprightspoor, Universidad Torcuato Di Tella.
287. GALIANI, S. (2006). Sesión I: Inferencia causal. Ponencia presentada en el Taller organizado por el Banco Mundial en Cuernavaca, Mayo de 2006.
288. GALIANI, S. (2006b). Sesión II: Estudios aleatorizados. Ponencia presentada en el Taller organizado por el Banco Mundial en Cuernavaca, Mayo de 2006.
289. GALIANI, S. (2006c). Sesión IV: Diferencias en diferencias (datos panel). Ponencia presentada en el Taller organizado por el Banco Mundial en Cuernavaca, Mayo de 2006.
290. GALIANI, S. (2006d). Sesión VI: Técnicas de pareamiento. Ponencia presentada en el Taller organizado por el Banco Mundial en Cuernavaca, Mayo de 2006.
291. GALIANI, S., GERTLER, P. & SCHARGRODSKY, E. (2002). Water for Life: The Impact of the Privatization of Water Services on Child Mortality. Stanford Institute for Economic Policy Research (SIEPR), Stanford University.
292. GALIANO S, GERTLER P, SCHARGRODSKY E. & STURZENEGGER F. (2003) The Costs and Benefits of Privatization in Argentina: A Microeconomic Analysis. Latin American Research Network Red de Centros de Investigación Research Network Working Paper #R-454. Bases de datos electrónica BIF.
293. GARCÍA, J. (1991). Recursos metodológicos en la Evaluación de Programas. *Bordón*, 43 (4): 461-476.
294. GARRETT J. (2001). Lessons from the Urban Food-for-Work Program: CAREEthiopia. Notes and Observations. IFPRI, Washington, D.C
295. GAVIRIA A. & NÚÑEZ J. (2003). Evaluation the impact of SENA on earnings and employment. Archivos de Economía. Documento 220, abril. Departamento Nacional de Planeación. Colombia.

296. GAVIRIA, A. & NÚÑEZ, J. (2003). Evaluation the impact of SENA on earnings and employment. Archivos de Economía. Documento 220, abril.
297. GEORGE, R., KOSEC, K. & WALLSTEN, S. (2004). Has Private Participation in Water and Sewerage Improved Coverage? Empirical Evidence from Latin America. *World Bank Policy Research Working Paper* N° 3445. World Bank, Development Research Group, Washington D.C.
298. GERTLER P, LEVINE DI. & MORETTI E. (2002). Do Microfinance Programs Help Families Insure Consumption Against Illness?. University of California, Berkeley Center for International and Development Economics Research (CIDER) working paper C03-129.
299. GERTLER P, PATRINOS H. & RUBIO-CODINA M. (2006). Empowering parents to improve education: evidence from rural Mexico. *World Bank Policy Research Working Paper* 3935.
300. GERTLER P. (2000). El Impacto del Programa de Educación Salud y Alimentación (Progesa) sobre la Salud. Washington, D.C.: International Food Policy Research Institute (IFPRI).
301. GERTLER PJ. & BOYCE S. (2001) An Experiment in Incentive-Based Welfare: The Impact of PROGESA on Health in México. Mimeo, UC Berkeley..
302. GERTLER, P. (2000). Final Report: The Impact of PROGRESA on Health. International Food Policy and Research Institute, Washington D.C.
303. GIBSON PG, SHAH S. & MAMOON HA. (1998) Peer-Led Asthma Education for Adolescents: Impact Evaluation. *Journal of Adolescent Health*; 22: 66 -72.
304. GIEDION U, GONZÁLES L. & ACOSTA OL. (2001). Efectos de la reforma en salud sobre las conductas irregulares en los hospitales públicos: El caso de Bogotá, Distrito Capital, Colombia. Latin American Research Network Red de Centros de Investigación Research Network Working paper #R-426. Bases de datos electrónica BIF.
305. GILBERT, N. & TROITZSCH, K. (1999). *Simulation for the Social Scientist*. Open University Press. Buckingham, Reino Unido.
306. GILL, J., BITTNER, H., TOWNSEND, B. & MAINOUS, A. (2005). Impact of Providing a Medical Home to the Uninsured Evaluation of a Statewide Program. *Journal of Health Care for the Poor and Underserved*, 16 (3): 515-535.
307. GLASS, D. & ELLETT, F. (1980) Evaluation research. *Annual Review of Psychology*, 31: 211-228.
308. GLASS, G., WILLSON, V. & GOTTMAN, J. (1975). *Design and analysis of time-series experiments*. Boulder, CO: Colorado Associated University Press
309. GLEWWE, P. & OLINTO, P. (2004). Evaluating the impact of Conditional Cash Transfers on Schooling: An experimental analysis of Honduras' PRAF program. Final Report for USAID. International Food Policy Research Institute. Washington D.C.

310. GLEWWE, P., KREMER, M. & MOULIN, S. (2000). Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya. Science Foundation, World Bank Research Committee.
311. GLEWWE, P., KREMER, M., MOULIN, S. & ZITZEWITZ, E. (2000). Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya. Working Paper N° 8018. National Bureau of Economic Research, Cambridge MA.
312. GLEWWE, P., NAUMAN, I. & KREMER, M. (2003). Teacher Incentives. Working Paper N° 9671. National Bureau of Economic Research, Cambridge, MA.
313. GLINER JA. & SAMPLE P. (1996) A multimethod approach to evaluate transition into community life. *Evaluation and Program Planning*, Vol. 19, No. 3, pp. 22-233
314. GODDARD PO, GÓMEZ G, HARRISON P. & HOOVER G (1981). The Product is Progress: Rural Electrification in Costa Rica . Proyect Impact Evaluation Report Na 22. Base de datos electrónica USAID
315. GODOY, L. & RANGEL, M. (1997). Nuevas experiencias en Política Social: Los Fondos de Inversión Social en América Latina y el Caribe en los Programas Sociales. Documento LC/R 1744, CEPAL, Santiago de Chile.
316. GONZÁLEZ, A. (1987). Evaluación de centros y programas educativos. *Revista Comunidad Educativa*, 15; 6-9.
317. GOODMAN, C. (1990). Evaluation of a Model Self-Help Telephone Program: Impact on Natural Networks. *Social Work*, 36 (6): 556-562.
318. GREENBERG M, MANTELLA N, LAHRA M, FRISCHB M, WHITEC K. & DAVID KEHLERD D. (2005) Evaluating the economic effects of a new state-funded school building program: the prevailing wage issue. *Evaluation and Program Planning*, 28:33-45.
319. GREENBERG M, MANTELLA N, LAHRA M, FRISCHB M, WHITEC K. & DAVID KEHLERD D. (2005) Evaluating the economic effects of a new state-funded school building program: the prevailing wage issue. *Evaluation and Program Planning*, 28:33-45.
320. GROSH, M.E. & MUÑOZ, J. (1996). A Manual for Planning and Implementing the Living Standards Measurement Study Survey. LSMS, Documento de trabajo N° 126. Banco Mundial, Washington, D.C.
321. GUBA, E. & LINCOLN, Y. (1985). *Naturalistic Inquiry*. Sage. Beverly Hills.
322. GUBA, E. (1978). *Toward a methodology of naturalistic inquiry in educational evaluation*. Los Ángeles. Center for the study of evaluation.
323. GUETZKOW, H. (1962). *Simulation in Social Science*. Englewood Cliffs, New Jersey. Prentice Hall.
324. GUGERTY M.K. & KREMER M. (2004). "The Rockefeller Effect". University of Wahington. Working Paper.
325. GUMUSELI AI. & ERGIN B. (2002) The Manager's Role in Enhancing the Transfer of Training: A Turkish Case Study . *International Journal of Training Development*, vol. 6: 80-97.

326. GUPTA N, KATENDE C. & BESSINGER R. (2004). An Evaluation of Post-campaign Knowledge and Practices of Exclusive Breastfeeding in Uganda. *Journal of Health Populaton and Nutrition*. Vol. 22, 4, 429-439,
327. HACKMAN, J., PEARCE, J., & WOLFE, J. (1978). Effects of changes in job characteristics on work attitudes and behaviors: A naturally occurring quasi-experiment. *Organizational Behavior and Human Performance*, 21: 289-304.
328. HALLER, E. (1974). Cost analysis for educational program evaluation. In W. James Popham, (Ed.), *Evaluation in Education: Current Applications*. Berkeley, CA: McCutchan Publishing Corporation.
329. HANDA S, HUERTA MC, PÉREZ R. & STRAFFON B. (2000) Pobreza, desigualdad y efectos indirectos del Programa de Educación, Salud y Alimentación. Washington, D.C.: International Food Policy Research Institute (IFPRI).
330. HANDA S. & HUERTA MC. (2003). On the Bias in Estimating the Impact of a Nutrition Intervention Using Clinic Based Data Disponible en: www.unc.edu/~shanda/research/Handa_Huerta_Program_Bias_V1.pdf.
331. HANSEN, C., GULLICKSON, A. & LAWRENZ, F. (2003). Assessing the Impact and Effectiveness of the Advanced Technological Education (ATE) Program. The Evaluation Center Western Michigan University.
332. HASEEN, F. (2006). Change in Food and Nutrient Consumption Among the Ultra Poor: Is the CFPR/TUP Programme Making a Difference?.. Working Paper Series No. 11. Challenging the Frontiers of Poverty Reduction/ Targeting the Ultra Poor.
333. HAWTHORNE, G. (1996). Education: estimating drug use prevalence among Victorian primary school students and the statewide effect of the Life Education programme. *Addiction*, 91 (8): 1151-1159.
334. HAYES L, QUINE S. & BUSH J. (1994) Peer-Led Asthma Education for Adolescents: Impact Evaluation Attitude change amongst nursing students towards Australian Aborigines. *International Journal of Nurse Studies*., Vol. 31, No. I, pp. 61 16, 1994.
335. HAYES RJ, CHANGALUCHA J, ROSS DA, GAVYOLE A, TODD J, OBASI AIN, PLUMER M, WIGHT D. MABEL D. & GROSSKURTH H. (2005) The MEMA kwa Vijana Project: Design of a community randomised trial of an innovative adolescent sexual health intervention in rural Tanzania. *Contemporary Clinical Trials*, 26: 430-442.
336. HAYS C., HAYS, S., DEVILLE, J., & MULHALL, P. (2000). Capacity for effectiveness: the relationship between coalition structure and community impact. *Evaluation and Program Planning*, 23: 373-379.
337. HEADY C, MITRAKOS T. & TSAKLOGLOU P. (2001). The Distributional Impact of Social Transfers in the European Union: Evidence from the ECHP Fiscal Studies, vol. 22, no. 4, pp. 547-565.
338. HECKMAN J. & PAGÉS C. (2000). The Cost of Job Security Regulation: Evidence from Latin American Labor Markets. Research department Departamento de investigación Working Paper #430. Bases de datos electrónica BIF.

339. HECKMAN, J., ICHIMURA, H., TODD, P. (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies* 64, 605–654.
340. HECKMAN, J., LALONDE, R. & SMITH, J. (1999). The Economics and Econometrics of Active Labour Market Programs. En: ASHENFELTER, O. & CARD, D. (Eds). *The Handbook of Labor Economics*, Volume III. Amsterdam: Elsevier Science.
341. HEINRICH CJ. & CABROL M. (2005) Programa Nacional de Becas Estudiantiles Impact Evaluation Findings. Office of Evaluation and Oversight. Working Paper. 06. julio.
342. HENGEL. P. & CARLSON, C. (2002). Enhancing Learning through Evaluation: Approaches, Dilemmas and Some Possible Ways Forward. Ponencia presentada en la V Conferencia Europea de Evaluación, Sevilla, octubre. Disponible en internet [www.europenevaluation.org].
343. HEVIA, H. (2001) Evaluación del impacto en los beneficiarios del programa de titulación masiva de tierras. Agencia de los Estados Unidos para el Desarrollo Internacional. Base de datos electrónica USAID
344. HODDINOTT J, SKOUFIAS E. & WASHBURN R. (2000). El impacto de Progresas sobre el consumo: informe final. Instituto Internacional de Investigación sobre Políticas Alimentarias, Washington, DC.
345. HOLLORAN S, COREY GL & MAHONEY T. (1982) Sederhan; Indonesi Small-Scale Irrigation. AID, Project Impact Evaluation Report N° 29. Base de datos electrónica USAID
346. HOLLORAN S, COREY GL & MAHONEY T. (1982) Sederhan; Indonesi Small-Scale Irrigation. AID, Project Impact Evaluation Report N° 29. Base de datos electrónica USAID
347. HOPENHAYN HA. (2001). Labor Market Policies and Employment Duration: The Effects of Labor Market Reform in Argentina Latin American Research Network Red de Centros de Investigación Research Network Working paper #R-407. Bases de datos electrónica BIF.
348. HORN, R. (1993). Statistical indicators for the economic and social sciences. Cambridge, University Press, Hong Kong,
349. HOTCHKISS DR, KRASOVEC K, EL-IDRISSI MDZ, ECKERT E. & KARIM AM.(2003) The Role of User Charges and Structural Attributes of Quality on the Use of Maternal Health Services in Morocco. Measure Evaluation, Working Paper-03-68.
350. HOTCHKISS DR, MAGNANI RJ, BROWN LF. & FLORENCE CS. (1998) Family Planning Program Effects on Contraceptive Use in Morocco, 1992-1995 .Measure Evaluation, Working Paper -98-08.
351. HOTZ, J., IMBENS, G. & MORTIMER, J. (1999). Predicting the Efficacy of Future Training Programs Using Past Experiences. National Bureau of Economic Research Technical. *Working Paper*, N°. 238.
352. HOUSE, E. (1980). *Evaluating with validity*, Beverly Hills: Sage Publications.
353. HOUSE, E. (1993). *Professional Evaluation Social Impact and Political Consequences*. Newbury Park: Sage.

354. HUGHES-D'AETH, A. (2002) Evaluation of HIV/AIDS peer education projects in Zambia. *Evaluation and Program Planning* 25 (2002) 397–407
355. HUNTER, W., SCHMIDT E., & ZAKOCS, R. (2005) Assessment Approach for Enhancing the Organizational Capacity of State Injury Prevention Programs. *Journal of Public Health Management Practice*, 11 (1): 29-36.
356. HUTTON, R. & McNEIL, D. (1981). The value of the incentives instimulating energy conservation. *Journal of Consumer Research*, 8: 291-298.
357. IKHWELONET CONSORTIUM (2002) Ikhwelo Pilot Projects: Final Report. Impact Evaluation Research Report, november 2001 and march 2002. Base de datos electrónica USAID
358. ILPES & CEPAL (2003). Bases conceptuales para el ciclo de cursos sobre gerencia de proyectos y programas. Serie Manuales N° 24. Santiago de Chile. Instituto Latinoamericano del Caribe y de Planificación Económica y Social. Edición: Naciones Unidas.
359. INSTITUTO APOYO. (2000). Sexta Evaluacion Ex-Post del FONCODES: Evaluacion de Impacto y Sotenibilidad. Lima. Peru.
360. IRENE NORLUND, THU NHUNG MLO DUON DU, NGO HUU TOAN (2004). NAV Evaluation Report - The Integrated Rural Development Program 1994 - 2004
361. ISAAC, W. (1999). Dialogue and the Art of Thinking Together: A Pioneering Approach to Communicating in Bussines and in Life, Nueva York.
362. JACINTO, C. & GALLART, A. (1998). La evaluación de programas de capacitación de jóvenes desfavorecidos: Una ilustración con programas para jóvenes desempleados en los países del Cono Sur. Paris. Instituto Internacional de Planeamiento de la educación/UNESCO.
363. JACKSON, B. & MOHR, L. (1986). Rent subsidies: an impact evaluation and an application of the random-comparison-group design. *Evaluation Review*, 10: 483-517.
364. JACKSON, N., TAYLOR, L., QUIGLEY, R. (En prensa). Findings from a retrospective process evaluation of five HIA studies. London: Health Development Agency.
365. JAGANNATHAN, R., CAMASSO, M. & KILLINGSWORTHUSA, M. (2004). Do Family Caps on Welfare Affect Births Among Welfare Recipients? Reconciling Efficacy and Effectiveness Estimates of Impact through a Blended Design Strategy. *American Journal of Evaluation*, 25 (3): 295–319.
366. JALAN, J. & RAVALLION, M. (1998). Geographic Poverty Traps? Institute for Economic Development. IED Discussion Paper Series, N° 86, Mayo, Boston University.
367. JALAN, J. & RAVALLION, M. (2003). Does Piped Water Reduce Diarrhea for Children in Rural India? *Journal of Econometrics*, 112 (1): 153-173.
368. JALAN, J. & RAVALLION, M. (2003). Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching. *Journal of Business and Economic Statistics*, 21 (1): 19-30.

369. JALAN, J., & GLINSKAYA, E. (2004). Improving Primary School Education in India: An Impact Assessment of DPEP-Phase I. World Bank, Washington, D.C.
370. JARIOT, M. (2001). La evaluación de la intervención mediante programas de orientación: El proceso de evaluación del desarrollo de una intervención desde un modelo de consulta. Tesis Doctoral. Universidad Autónoma de Barcelona.
371. JEE-PANG, T., LANE, J. & LASSIBILLE, G. (1999). Student Outcomes in Philippine Elementary Schools: An Evaluation of Four Experiments. *The World Bank Economic Review*, 12 (3): 493-508.
372. JENKINS GP. & KUO CY. (2006) Evaluation of the Benefits of Transnational Transportation Projects. *Journal of Applied Economics*. Vol 9, No. 1, pp. 1-17.
373. JERRELL JM, & RIDGELY MS. (1999) The relative impact of treatment program 'robustness' and 'dosage' on client outcomes. *Evaluation and Program Planning* 22 323-330
374. JIMENEZ, E. & SAWADA, Y. (1998). Do Community-Managed Schools Work? An Evaluation of El Salvador's EDUCO Program. Paper N° 8 of Working Paper Series on Impact Evaluation Reforms. Development Research Group, World Bank, Washington D.C.
375. JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1981), Standards for evaluations of educational programs, projects, and materials. New York. McGraw-Hill.
376. JONES, P.(1997). The Impact of Minimum Wage Legislation in Developing Countries where Coverage is Incomplete. Centre for the Study of African Economies. Institute of Economics and Statistics. University of Oxford WPS/98-2
377. JONSON WF, FERGUSON CE. & FIKRY M. (1983) The Wheat Development Program. A.I.D. Project Impact Evaluation No. 48. Base de datos electrónica USAID
378. JONSON-WELCH C. (1999). Focusing on Women Works: Research on Improving Micronutrient Status through Food-Based Intervention International Center for Research on Women. Synthesis Paper.
379. JOSHI PK, JHA AK, WANI SP, JOSHI L. & SHIYANI RL. (2005). Meta-Analysis to Assess Impact of Watershed Program and People's Participation. Comprehensive Assessment of Water Management in agriculture. Research Report 8.
380. JUDD, C. & KENNY, D. (1981). Process Analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5(5), 602-619.
381. JULNES, G. (2004). Review of Experimental and Quasi-experimental Designs for Generalized Causal Inference: By W.R. Shadish, T.D. Cook, D.T. Campbell, 2002; Houghton-Mifflin, Boston. *Evaluation & Program Planning*, Mayo de 2004, Vol. 27 N° 2, p.173.
382. JUMBE CH.BL.. & ANGELSEN A. (2006). Do the poor benefit from devolution policies? Evidence from Malawi's forest co-management program. Forthcoming in *Land Economics*, noviembre.

383. KARUGIA JT, MWAI OA, KAITHO R, DRUCKER AG, WOLLNY CBA. & REGE JEO, (2001) Economic Analysis of Crossbreeding Programmes in Sub-Saharan Africa: A Conceptual Framework and Kenyan Case Study. Working paper N° 106. Fondazione Eni Enrico Mattei.
384. KAUFFMAN, R., & ENGLISH, F. W. (1979) Weeds assessment concepts and application. Englewood Cliffs, NJ: Educational Technology Publications.
385. KAZDIN, A. (1984). Statistical analysis for single-case experimental designs. En D. Barlow y H. Hersen (Eds.): Single-Case Experimental Designs: Strategies for Studying Behavior Change. 2ª ed. pp. 265-316. Nueva York: Pergamon Press.
386. KEEVES, J. (Ed.) (1997). Educational Research, Methodology and Measurement. An International Handbook. 2ª Edición. Londres: Pergamon.
387. KESSLER JL. BALLANTYNE J, MAUSHAMMER R. & ROMERO N.(1981) Ecuador: Rural electrification. AID. Project Impact Evaluation No. 48. Base de datos electrónica USAID
388. KIGER, D. (2000). The tribes process TLC: A preliminary evaluation of classroom implementation & Impact on student achievement. *Education*, 120 (3): 586-592.
389. KILLICK, T. (1995). IMF Programmes in Developing Countries, Design and Impact. London: England.
390. KIM, J., ALDERMAN, H. & ORAZEM, P. (1998). Can Cultural Barriers Be Overcome in Girls' Schooling?: The Community Support Program in Rural Balochistan. Working Paper Series on Impact Evaluation of Education Reforms, Paper N° 10. World Bank, Development Research Group, Washington D.C.
391. KIM, J., ALDERMAN, H. & ORAZEM, P. (1998). Can Private School Subsidies Increase Schooling for the Poor?: The Quetta Urban Fellowship Program. Working Paper Series on Impact Evaluation of Education Reforms, Paper N° 11. World Bank, Development Research Group, Washington D.C.
392. KING, E., OLZER, B. & RAWLINGS, L. (1999). Nicaragua's School Autonomy Reform: Fact or Fiction? Working Paper Series on Impact Evaluation of Education Reforms, Paper N° 19. World Bank, Development Research Group, Washington D.C.
393. KIRBY D, LARIS BA. & ROLLERI L. (2005). Impact of Sex and HIV Education Programs on Sexual Behaviors of Youth in Developing and Developed Countries. Family Health International. FHI Working Paper Series N°. 2.
394. KISH, L. (1965). Survey Sampling. Nueva York: John Wiley and Sons.
395. KISH, L. (1987). Statistical Design for Research. John Wiley & Sons.
396. KLAWITTER, MM. & FLATT, V. (1998) The Effects of State and Local Antidiscrimination Policies on Earnings for Gays and Lesbians. *Journal of Policy Analysis and Management*, Vol. 17, No. 4, 658–686.
397. KLIKSBURG, B. (1997). Pobreza, un tema impostergable: nuevas propuestas a nivel mundial, México, CLAD/FCE/ Programa de Naciones Unidas.

398. KLOEKA GC, VAN LENTHEA FJ, VAN NIEROPB PWM, KOELENC MA. & MACKENBACH JP. (2006) Impact evaluation of a Dutch community intervention to improve health-related behaviour in deprived neighbourhoods. *Health & Place*, 12: 665-677.
399. KLUGER A. (2001) The Incidence of Job Security Regulations on Labor Market Flexibility and Compliance in Colombia: Evidence from the 1990. Reform. Research Network Working papers ; R-393. Bases de datos electrónica BIF.
400. KRATOCHWILL, T. (1978). Single Subject Research. Strategies for evaluating change. Nueva York: Academic Press.
401. KREMER M, MOULIN S. & NAMUNYU R. (2003). Decentralization: A Cautionary Tale. Harvard University, Cambridge, Mass.
402. KREMER, M., EDWARD, M., THORNTON, R. & OZIER, O. (2004). Incentives to Learn. Policy Research Working Paper N° 3546. World Bank, Washington D.C.
403. KRÖGER, C., WINTER, H. & SHAW, R. (1998). Guía para la evaluación de las intervenciones preventivas en el ámbito de las drogodependencias. Manual para responsables de planificación y evaluación de programas. Lisboa: Observatorio Europeo de Drogas y Toxicomanías.
404. KRUGLANSKI, A. & KROY, M. (1976). Outcome validity in experimental research: A reconceptualization. *Representative Research in Social Psychology*, 7: 166-178.
405. KRUSE, D. & SCHUR, L. (2003) Employment of People with Disabilities Following the ADA. *Industrial Relations*, Vol. 42, No. 1.
406. LALONDE, R. & MAYNARD, R. (1987). How Precise Are Evaluations of Employment and Training Programs? Evidence from a Field Experiment. *Evaluation Review*, 11(4): 428-451.
407. LALONDE, R. (1986). Evaluating the Econometric Evaluations of Training Programs. *American Economic Review*, 76: 604-620.
408. LANE, J. (2000). *New Public Management*. Sage, London.
409. LATORRE, A.; DEL RINCÓN, D. & ARNAL, J. (1996). Bases metodológicas de la investigación educativa. Barcelona: Hurtado Ediciones.
410. LAVY, V (2003) Pay for Performance: the Effect of Teachers' Financial Incentives on Students' Scholastic Outcomes' CEPR Discussion Paper No. 3862
411. LEACH, F, ABDULLA S, APPLETON H, EL-BUSHRA, CARDENAS N, KEBEDE K, LEWIS V. & SITARAM S. (2000). The Impact of Training on Women's Micro-Enterprise Development. Education Research Paper No. 40, 139 p.
412. LECHNER, M. & VAZQUEZ-ALVAREZ, R. (2003) The Effect of Disability on Labour Market Outcomes in Germany: Evidence from Matching IZA. Discussion Paper. N° 967.
413. LECHNER, M.(1999) Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification, *Journal of Business & Economic Statistics*, 17, 74-90.

414. LEE, D. (2005) Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. National Bureau of Economic Research (NBER) Working Paper. N° 11721.
415. LEHMAN, R. (1980). What simulation can do to the statistics and design course? *Behavior Research Methods and Instrumentation* 12: 157-159.
416. LEÓN, M. & YOUNGER, SD. (2004) Transfer Payments, Mother's Income, and Child Health in Ecuador. Cornell Food and Nutrition Policy Program Working Paper No. 172.
417. LEVIN, H. (1975). Cost-Effective Analysis in Evaluation Research. In Guttentag, M. & Struening, E. (Eds.). *Handbook of Evaluation Research*, Vol. 2, Sage, Beverly Hill.
418. LEVY I, ZUVEKAS C. & STEVENS C. (1981). Philippines: Rural Roads I and II. AID Project Impact Evaluation Report N° 18. Base de datos electrónica USAID
419. LEVY, A. et al. (1985). The impact of a nutrition information program on food purchases. *Journal of Public Policy and Marketing*, 4: 1-13.
420. LINDEN L, BANERJEE A. & DUFLO E. (2003). Computer-Assisted Learning: Evidence from a Randomized Experiment . Cambridge, MA: Poverty Action Lab.
421. LIPSEY, M (1996). Key issues in intervention. A program evaluation perspective. *American Journal of Industrial Medicine*, 29, 298-302.
422. LOGAN TK, HOYT W., MCCOLLISTER K, FRENCH M, LEUKEFELD C, & MINTON L. (2004) Economic evaluation of drug court: methodology, results, and policy implications. *Evaluation and Program Planning* 27 (2004) 381–396.
423. LOHR, B. (1972). An historical of the research on the factors related to the utilization of health services. Duplicated research report, Bureau for Health Services Research and Evaluation, Social and Economic Analysis Division, Rockville, MD, January.
424. LOKSHIN, M. & YEMTSOV, R. (2003). Evaluating the Impact of Infrastructure Rehabilitation Projects on Household Welfare in Rural Georgia. World Bank Policy Research Working Paper N° 3155. World Bank, Development Economics Research Group and Eastern Europe and Central Asia, Poverty Reduction and Economic Management, Washington, D.C.
425. LÓPEZ, JH. (2004) Pro-growth, pro-poor: Is there a tradeoff? World Bank Policy Research Working Paper 3378.
426. LÓPEZ-BARAJAS, E.; LÓPEZ. E. & PÉREZ, R. (1987). *Pedagogía experimental I*. Madrid: UNED.
427. LORA, E. & OLIVERA M. (1998) “Macro Policies and Employment Problems in Latin America”. Inter-American Development Bank, Working Paper No. 372.
428. LUCAS K, DAVIS T. & RINKARD K. (1996) Agricultural Transportation Assistance Program (ATP). Disponible en base de datos electrónica USAID.
429. MADUEÑO, M., LINARES, J. & ZURIA, A. (2004). Instrumento estandarizado de identificación de beneficiarios para programas sociales en el

- Perú. Bethesda, MD: Socios para la reforma del sector salud, Abt Associates Inc.
430. MAGNANI R, TOURKIN S. & HARTZ S. (1984) Evaluation of the Provincial Water Project in the Philippines. S. Agency for International Development. Project 492-U-033. Base de datos electrónica USAID
431. MAGNANI RJ, HOTCHKISS DR, FLORENCE CS.& LEIGH LA.(1999) The impact of Family Planning Suplí Environment on Contraceptive Intentions and Use in Morocco. *Studies in Family Planning*, vol. 30, 2: 120-132.
432. MAGNANI, R. ET AL. (2003) The Impact of Life Skills Education on Adolescent Sexual Risk Behaviors. Horizons Research Summary. Washington, D.C.: Population Council.
433. MAKANZA M. & MUNYARADZI R. (2004) Impact Evaluation on Implementatio of the WTO valuation agreement for Mozambique. Regional Center for Southern África, U.S. Agency for International Development Gaborone, Botswana. Base de datos electrónica USAID
434. MALUCCIO, J. & FLORES, R. (2004). Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red De Proteccion Social. Food Consumption and Nutrition Division Discussion Paper N° 184. International Food Policy and Research Institute, Food Consumption and Nutrition Division, Washington D.C.
435. MALUCCIO, J. (2003). Education and Child Labor: Experimental Evidence from a Nicaraguan Conditional Cash Transfer Program. International Food Policy and Research Institute, Washington D.C.
436. MANDEL DH, ALLGEIR PF, WASSERMAN G, HICKEY G, SALAZAR R. & ALVIAR J. (1980) The Philippines: Rural Electrification. Project Impact Evaluation N°. 15 Base de datos electrónica USAID
437. MANOHAR, V., DES ROCHES, E. & FERNEAU, E. (1976). An education Program in Alcoholism for Social Workers: Its Impact on Attitudes and Treatment-Orientede Behavior. *British Journal Addiction*, 71: 225-234.
438. MARK, M. (1986). Validity typologies and the logic and practice of quasi-experimentation. En M.K. Trochim (Ed.). *Advances in quasi-experimental design and analysis* (pp. 47-67). San Francisco: Jossey-Bass.
439. MARK, M. (2001). Evaluation Capacity and the Development of Contextually-Sensitive Evaluation Practice: The Role of the Evaluation Journal. *The Japanese Journal of Evaluation Studies*, 1 (1).
440. MARKIEWICZ, A (2005). A balancing act: resolving multiple stakeholder interests in program evaluation. *Evaluation Journal of Australasia*, 4 (1, 2): 13-21.
441. MÁRQUEZ G. & PAGÉS C (1997) Trade and Employment: Evidence from Latin America and the Caribbean. Inter-American Development Bank, Working Paper No. 373.
442. MARTÍNEZ, C. (1996). Evaluación de programas educativos. Investigación evaluativa. Modelos de evaluación de programas. Madrid, UNED.

443. MASUD N. & YONTCHEVA B. (2005). Does Foreign Aid Reduce Poverty? Empirical Evidence from Nongovernmental and Bilateral Aid. International Monetary Fund. Working Paper 05/100.
444. MATEU, P. & VILCA, J. (2004). Modelo de medición de impacto sobre el bienestar objetivo y subjetivo: un análisis de caso del Proyecto de Reducción y Alivio a la Pobreza (PRA). Lima. Universidad del Pacífico CIUP.
445. MATYAS, T. & GREENWOOD, K. (1990). Visual analysis of single-case time-series: Effects of variability, serial dependence, and magnitude of intervention effect. *Journal of Applied Behavior Analysis*, 23: 341-351.
446. MATYAS, T. & GREENWOOD, K. (1991). Problems in the estimation of autocorrelation in brief time series and some implications for behavioral data. *Behavioral Assessment*, 13: 137-157.
447. MAXWELL J, CHAPIN N, DEMETRE MC. & FLETCHER L. (1981). Honduras Rural Roads: Old Directions and New. AID Agect Impact Eduation Report No. 17. Base de datos electrónica USAID
448. MAXWELL, S. & DELANEY, H. (1990). Desgning experiments and analyzing data: A model comparison perspective. Belmont, CA: Wadsteorth.
449. MCCLELLAND DG, JOHNSTON AG, PEDERSEN BR, SCHWARTZ JB. & SCHMEDING R.(1993) Evaluaron of AID Family Planning Programs: Ghana Case Study. Center for Development Information and Evaluation. AID Technical Report N° 13. Disponibel en Base de datos electrónica. USAID
450. MCCORD, A. (2004). Policy Expectations and Programme Reality: The Poverty Reduction and Labour Market Impact of Two Public Works Programmes in South Africa. Economics and Statistics Analysis Unit. University of Cape Town. Working Paper. 8.
451. MCDOWELL J. & MOULTON J. (1990). Liberian Primary Education Program (PEP):Final Impact Evaluation of the PEP Pilot Radio Broadcast. The Academy for Educational Development and The Bureau of Primary Education, Ministry of Education, Liberia
452. MCGAHA A, BOOTHROYD RA, POYTHRESS NG, PETRILA J. & ORT RG. (2002) Lessons from the Broward County Mental Health Court Evaluation. *Evaluation and Program Planning*, 25: 125-135.
453. MCKAY, H.; SINISTERRA, L., MCKAY, A., GÓMEZ, H. y LLOREDA, P. (1978). Improving cognitive ability in chronically deprived children. *Science*, 200: 270-278.
454. MCKERNAN, S. (2002). The Impact of Microcredit Programs on Self-employment Profits: Do Noncredit Program Aspects Matter? *The Review of Economics and Statistics*, 84 (1): 93-115.
455. MCLAUGHLIN, M. (1976). Implementation as Mutual Adaptation. *Teachers College Record*, 77: 339-351.
456. MEDIANERO, D. (1996). Principios básicos de la evaluación de impacto. *Revista de la Facultad de Ciencias Económicas*, UNMSM, AÑO VI, N° 19, marzo. Lima.
457. MEDIANERO, D. (2001). El Enfoque del Marco Lógico en la Gestión de Proyectos. CINDEH.

458. MEDINA GIOPP, A. & MEJIA, J. (1993). El Control en la Implantación de la Política Pública. México, Plaza y Valdés.
459. MEDINA GIOPP, A. (2002). La eficiencia en los programas sociales: Análisis de las situaciones y propuestas para su mejoramiento. VII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Lisboa, Portugal, 8-11 Oct.
460. MEDINA, C. & NÚÑEZ, J. (2005) The Impact of Public and Private Job Training in Colombia. Inter-American Development Bank. Research Network Working papers ; R-484.
461. MEEHAN R. & VIVEROS-LONG AM. (1982) Liberian Primary Education Program (PEP):Final Impact Evaluation of the PEP Pilot Radio Broadcast.
462. MEEHAN, R.; VIVEROS-LONG, AM (1982). Panamá: Rural Water. Project Impact Evaluation N° 32. Base de datos electrónica USAID.
463. MEEKERS D, AGHA S. & KLEIN M. (2004). The impact on condom use of the "100% Jeune" social marketing program in Cameroon. Journal of Adolescent Health. Vol. 36, N° 6. p. 530.
464. MEEKERS, D. & VAN ROSSEM, R. (2004). The Reach and Impact of Social Marketing and Reproductive Health Communication Campaigns in Zambia. Measure Evaluation. Working Paper -04-77.
465. MEGHIR C. & PALME M. (2003). Ability, Parental Background and Education Policy: Empirical Evidence from a Social Experiment. The Institute For Fiscal Studies WP03/05.
466. MELCHIORA LA, HUBAA GJ, BROWNB VB. & SLAUGHTERB R. (1999) Evaluation of the effects of outreach to women with multiple vulnerabilities on entry into substance abuse treatment. Evaluation and Program Planning, 22: 9-277.
467. MENARD, C. & GEORGE, C. (2000). A transitory Regime: Water Supply in Conakry, Guinea." World Bank Policy Research Working Paper N° 2362. World Bank, Development Research Group, Public Economics and Regulation and Competition Policy, Washington D.C.
468. MENARD, C., GEORGE, C. & ZULUAGA, A. (2000). The Welfare Effects of Private Sector Participation in Guineas Urban Water Supply. World Bank Policy Research Working Paper N° 2361. World Bank, Development Research Group, Public Economics and Regulation and Competition Policy, Washington D.C.
469. MENY, Y. & THOENIG, J. (1992). Las políticas públicas. Barcelona: Editorial Ariel.
470. MERTENS, D. & RUSSON, C. (2000). A proposal for the International Organization for Cooperation in Evaluation. American Journal of Evaluation, 21, 275-283.
471. MEYER, M. & CONRAD, S. (1992). The Case of Bilingual Education Strategies. Washington, D.C.: National Academy Press.
472. MICHAEL J. CAMASSO MJ, JAGANNATHANB R. & WALKER CC. (2004) New Jersey's Transitional Housing Demonstration Program: the

- relationship of service delivery structure and process to the attainment of more permanent forms of housing. *Evaluation and Program Planning* . vol. 27: 45–58
473. MICHAEL, J.; SCHWERINA, M., MICHAELB, P., & GLASERC, D. & FARRAR, K. (2002). A cluster evaluation of Navy quality of life programs. *Evaluation and Program Planning*, 25: 303–312.
474. MIDDLETON S, PERREN K, MAGUARI S, RENNISON J, BATTISTIN E, EMERSON C. & FITZSIMONS E. (2005). Evaluation of Education Maintenance Allowance Pilots: Young People Aged 16 to 19 Years. Final Report of the Quantitative Evaluation .Department for Education and Skills, Nottingham, RR 499.
475. MILES, M. & HUBERMAN, A. (1994). *Qualitative Data Analysis*. Londres: Sage Publications.
476. MILLER, J. (2002). Impact Evaluation: Resettlement Grant Activity. Emergency Recovery: Agricultura and Comercial Trade (ER:ACT). Base de datos electrónica USAID
477. MINES, R., GRESSARD, C., & DANIELS, H. (1982). Evaluation instudent services: A metamodel. *Journal of College Student Personnel*, 23 (3): 195–201.
478. MINISTERIO DE ECONOMÍA Y FINANZAS (2002). Manual para la evaluación ex-post de proyectos de riego grandes y medianos. Dirección General de Programación Multianual del Sector Público. Lima.
479. MINTON, J. (1975). The impact of “Sesame Street” on reading readiness of kindergarten children. *Sociology of Education*, 48: 141-151.
480. MIRANDA M, PORRAS IT. & MORENO IL. (2003). The social impacts of payments for environmental services in Costa Rica: A quantitative field survey and analysis of the Virilla watershed . International Institute for Environment and Development, London.
481. MOCK, NB. ET AL. (1993). The utility of case-control methods for health policy and planning analysis .*Evaluation and Program Planning*, Vol. 16, pp. 199-205, 1993
482. MOCTEZUMA, E & ROEMER, A. (1999). Por un Gobierno con Resultados. FEC.
483. MOFFITT, R. (1991). The use of selection modelling to evaluate AIDS interventions with observational data. *Evaluation Review*, 15: 291-314.
484. MOHR, L. (1995). *Impact Analysis for Program Evaluation*. Sage Publications.
485. MOKATE, K. (2000). Convirtiendo el monstruo en aliado: la evaluación como herramienta de la gerencia social”. Documentos de trabajo del INDES, BID. Versión modificada, Washington D.C.
486. MOKATE, K. (2001) Eficacia, eficiencia, equidad y sostenibilidad: ¿qué queremos decir? Instituto Interamericano para el Desarrollo Social – INDES. Documento de Trabajo I-24, Washington D.C.
487. MONDITO G. & MONTOYA S (2000). The Effects of Labor Market Regulations on Employment Decisions by Firms: Empirical Evidence for Argentina Latin American Research Network Red de Centros de Investigación Research Network Working Paper #R-391

488. MONDRAGÓN, A. (2002). ¿Qué son los indicadores? Instituto Nacional de Estadística, Geografía e Informática (INEGI), N° 19, Julio-septiembre.
489. MONNIER, E. (1991). Objectifs et destinataires des évaluations. *Documentación Administrativa*, N° 224/225: 131-157. España.
490. MORA JO, PIELMEIER N, ÁVILA P. & RAMÍREZ M. (1988). Growth Monitoring and Nutrition Education: Impact evaluation of an effective applied nutrition program in the Dominican Republic. U.S. Agency for International Development Bureau for Science and Technology Office of Nutrition. Washington. Base de datos electrónica USAID
491. MORAH E, MEBRATHU S. & SEBHATU K. (1998) Evaluation of the orphans reunification project in Eritrea. *Evaluation and Programming Planning*, 21: 437-448.
492. MORALES, P. (1993). Líneas Actuales de Investigación en Métodos Cuantitativos: El Meta - Análisis o la Síntesis Integradora. *Revista de Educación*. 300: 191-221.
493. MOREL, J. (2002). Evaluating the impact of an electronic business system in a complex organizational setting: the case of Central Contractor. *Evaluation and Program Planning*, 26: 429-440.
494. MOSTAJO, R. (2002). El sistema presupuestario en el Perú. Serie Gestión Pública N° 17. ILPES & CEPAL. Santiago de Chile.
495. MOSTELLER, F. (1990) Improving research methodology: an overview. En SEHREST, I.; PERRIN, E. & BUNKER, L. (Eds.) *Research Methodology: Strengthening Causal Interpretation of Nonexperimental Data*, pp. 221-230. Rockville, MD: Agency for Health Care Policy and Research.
496. MUDYARABIKWA, O. (2000). An examination of Public Sector subsidies to the Private Health Sector: A Zimbabwe Case Study. Regional Network for Equity in Health in South Africa. Equinet Policy Series. N° 8.
497. NACIONES UNIDAS (1998) Gestión de Programas Sociales en América Latina. Volumen 1. Serie Políticas Sociales N° 25. Comisión Económica para América Latina y el Caribe. P. 14.
498. NAVAIE-WALISER, M; MARTIN S.; TESSARO I.; CAMPBELL M. & CROSS, A. (2000). Social Support and Psychological Functioning among High-Risk Mother: The impact of the Baby Love Maternal Outreach Worker Program. *Public Health Nursing*, 17 (4): 280-291.
499. NEUMARK, D. & ADAMS, S. (2003) Detecting Effects of Living Wage Laws. *Industrial Relation*. Vol. 42. N° 4.
500. NEWMAN, J., PRADHAN, M., RAWLINGS, L. RIDDDER, G. COA, R. & EVIA, J. (2002). An Impact Evaluation of Education, Health and Water Supply Investments by the Bolivian Social Investment Fund. *The World Bank Economic Review*, 16 (2): 241-274.
501. NICHOLSON E, MILES HL, JOHNSON DN. (1983) U. S. Aid to Education in Paraguay: The Rural Education Development Project AID Project Impact Evaluation Report N° 46. Base de datos electrónica USAID
502. NICHOLSON E, MILES HL, JOHNSON DN. (1983) U. S. Aid to Education in Paraguay: The Rural Education Development Project AID Project Impact Evaluation Report N° 46. Base de datos electrónica USAID

503. NICKSON, A. (2002). Transferencia de políticas y reforma en la gestión del sector público: el caso del New Public Management. *Reforma y Democracia*, 24. Caracas, CLAD.
504. NIOCHE, J. (1982). De la evaluación al análisis de políticas públicas. *Revue Francaise de Science Politique* N°1, Francia, febrero.
505. NORES M, BELFIELD, C.; STEVEN, W; SCHWEINHART (2005). Updating the Economic Impacts of the High/Scope Perry Preschool Program. *Educational Evaluation and Policy Analysis*. 27 (3): 245-261.
506. NORES, M., BELFIELD, C. & SCHWEINHART, S. (2005). Updating the Economic Impacts of the High/Scope Perry Preschool Program. *Educational Evaluation and Policy Analysis*. 27, 3: 245-261.
507. NORTH SOUTH CONSULTANTS EXCHANGE (2003). Impact Study of the New Horizons Program in Egypt. Final Report. 03-07.
508. NUSSBAUM, M. & SEN, A. (1999) La calidad de vida. Fondo de Cultura Económica, Serie Economía contemporánea.
509. NYARKO P, PENCE B. & DEBPUUR C. (2001). Immunization Status and Child Survival in Rural Ghana. Population Research Division Working Paper No. 147, Population Council, New York.
510. ÑOPO, H. & ROBLES, M. (2002). Evaluación de Programas Sociales: Importancia y Metodologías. Estimación Econométrica para el Caso de PROJoven. Informe Final. Consorcio de Investigación Económica y Social Proyectos Medianos, Lima.
511. ÑOPO, H. & ROBLES, M. (2004). Evaluación de los programas sociales: estimación para el caso de ProJoven. *Economía y Sociedad* N° 52, CIES, julio de 2004. Pp. 58-62.
512. ÑOPO, H; ROBLES, M. & SAAVEDRA, J. (2002). Una medición del impacto del programa de capacitación laboral juvenil Projoven. – Lima: GRADE. Documento de trabajo, N°36.
513. OBASI A ET AL. (2003). Mema Kwa Vijana. A randomised controlled trial of an adolescent sexual and reproductive health intervention programme in rural Mwanza, Tanzania. Intervention and Process Indicators. African Medical & Research Foundation (AMREF),
514. OCDE, (2001) Evaluation Feedback for Effective Learning and Accountability.
515. OECD (1998) Review of the DAC Principles. Disponible en internet:: [<http://www.oecd.org/dac/evaluation>] Acceso el 26 de agosto de 2004.
516. OLÍAS DE LIMA, B. (2001). La Nueva Gestión Pública. Prentice may: Madrid.
517. ORGANIZACIÓN MUNDIAL PARA LA SALUD - OMS (2003) Programación para la Salud y el Desarrollo de los adolescentes. Informe de un grupo de estudio. OMS/FNUAP/UNICEF sobre programación para la salud de los adolescentes. OMS. Serie de informes técnicos N°886. OMS, Ginebra. Pg. 227-234.
518. ORTHNER, D., COOK P., SABAH, Y. & ROSENFELD, J. (2005). Organizational learning: A cross-national pilot-test of effectiveness in children's services. *Evaluation and Program Planning*, XX: 1-9.

519. ORTHNER, DK; COOK, P; SABAH, Y; ROSENFELD, J. (2006) Organizational learning: A cross-national pilot-test of effectiveness in children's services. *Evaluation and Program Planning*, 29:70–78
520. ORTIZ, G. (2003). Desafíos de las políticas sociales. *Economía y Sociedad* N° 48. CIES: Lima.
521. ORTIZ, G.; SANDOVAL, R. & HUSNI, S. (2001). Estado de los sistemas de monitoreo y evaluación de los programas sociales públicos en el Perú. CIES: Lima.
522. ORTIZ, S., BALCAZAR, M., ROBISON, N., ROJAS R. & ALURRALDE, L. (2004). Evaluación de Impacto del Programa Desarrollo Integral Temprano (PIDI). Reporte de Marketing SRL. La Paz, Bolivia.
523. OSBORNE, D. & GAEBLER T. (1994). La reinención del gobierno. Cómo el espíritu emprendedor está transformando el Sector Público. Paidós. (Versión en inglés de Penguin Books, 1993).
524. OSPINA, S. (2001). Evaluación de la Gestión Pública: conceptos y aplicaciones en el caso latinoamericano. *Reforma y Democracia*, 19, CLAD. Disponible en Internet: [<http://www.clad.org.ve/reforma.html>] Acceso del 26 de agosto de 2004.
525. OWNE, B. & PORTILLO, J.(2003) Legal Reform, Externalities and Economic Development: Measuring the Impact of Legal Aid on Poor Women in Ecuador. Public Law & Legal Theory Working Paper Series Research Paper No. 55
526. PAES R. & CORSEUIL CH. (2001) The Impact of Regulations on Brazilian Labor Market Performance. Latin American Research Network Red de Centros de Investigación. Research Network Working paper #R-427
527. PAINTER JE, BALWIN E, AHMED AS, SIDDIQUI MA.& KHAN MH. (1982) The On-Farm Water Management Project In Pakistan.. AID Project Impact Evaluation Report N° 35. Base de datos electrónica USAID
528. PAKER, C., RIVAS, G. & CAUAS, E. (1999) Evaluación de Impacto en Programas de Superación de Pobreza, el Caso del Fondo de Inversión Social (FOSIS) de Chile. Informe Técnico. Washington, mayo.
529. PANNARUNOTHAI, S. & KITTIDILOKKU, S. (1997). Paying Health Personnel in the Government Sector by Fee-For-Service: a Challenge to Productivity and Quality, and a Moral Hazard . *Human Resources for Health Development (electronic journal)* 1 (2).
530. PARKER SW. & SKOUFIAS. E. (2000) El impacto de Progresas sobre el trabajo, el ocio y el uso del tiempo. Instituto Internacional de Investigación sobre Políticas Alimentarias, Washington, DC.
531. PARLETT, M. & HAMILTON, D. (1989) La evaluación como iluminación. En: Gimeno SACRISTÁN, J. y PÉREZ GÓMEZ, A. (Dir.): *La enseñanza: su teoría y su práctica*. Madrid, Akal ; 450-466
532. PARSON, K.; LI, S.; & MCENTIRE, R. (2002). Data Integration in the Evaluation of Juvenile Justice Education. *Evaluation Review*, 26 (3): 322-339.
533. PATTON, M. (1980). *Qualitative Evaluation methods*. Beverly Hills, Ca.: Sage.

534. PATTON, M. (1982). *Practical Evaluation*. Beverly Hills. Sage Publications.
535. PATTON, M. (1989). A context and boundaries for a theory-driven approach to validity. *Evaluation and Program Planning*, 12: 375-377.
536. PATTON, M. (1997). *Utilization-Focused evaluation*. Thousand Oaks: SAGE Publications, Inc.
537. PAULL G, WALKER I. & ZHU Y. (2000). Child Support Reform: Some Analysis of the 1999 White Paper. *Fiscal Studies*, vol. 21, no. 1, pp. 105–140
538. PAXON, C.; WALDFOGEL, J. (2001) Welfare reforms, family resources, and child maltreatment. In: Meyer B, Duncan G. , editor. *In The incentives of government programs and the well-being of families*. Chicago: Joint Center for Poverty Research; pp. 1–47.
539. PAXSON, C. & SCHADY, N. (2002). The Allocation and Impacts of Social Funds: Spending on School Infrastructure in Peru. *The World Bank Economic Review*, 16(2): 297-319.
540. PEDHAZUR, E. (1982). *Multiple regressions in behavioral research*. Segunda Edición. New York, Holt, Rinehart and Winston.
541. PERACCHIO, L. & COOK, T. (1988). Avances en el diseño cuasi-experimental. En I. DENDALUCE (Coord.). *Aspectos Metodológicos de la Investigación Educativa* (pp. 85-101). País Vasco: Narcea.
542. PÉREZ JUSTE, R. (1985). Diseño experimental. En DE LA ORDEN, A. (Dir.), *Investigación educativa*. Diccionario de Ciencias de la Educación. Madrid: Anaya.
543. PÉREZ, R. (1991) La universidad Nacional a Distancia: a proximación a la evaluación de un modelo innovador. UNED. Madrid.
544. PÉREZ, R. (1992). Evaluación de programas de orientación. V Congreso Iberoamericano de Orientación. Tenerife: AEDEP.
545. PERLOFF, R., PERLOFF, E. & SUSSNA, E. (1976). Program evaluation. *Annual Review of Psychology*, 27, 569-594.
546. PERREN K, MIDDLETON S. & EMERSON C. (2003). Education Maintenance Allowance Transport Pilots – Quantitative Findings from Year 1 and 2 (2000-2001/2001-2002). Department for Education and Skills. Research Repor N° 471. Disponible en Base de Datos The Institute for Fiscal Studies.
547. PERRY HB, SHANKLIN DS. & SCHROEDER DG. (2003). Impact of a Community-based Comprehensive Primary Healthcare Programme on Infant and Child Mortality in Bolivia. *Journal of Health Population and Nutrition*;vol. 21, 4:383-395.
548. PETERSON PE. & CAMPBELL DE. (2001) School Choice in Dayton, Ohio After Two Years: An Evaluation of the Parents Advancing Choice in Education Scholarship Program Kennedy School of Government, Faculty Research. RWP02-021
549. PEWU G. & BELLEH W. (1982) The Monrovia consolidated school system: an impact evaluation En: USAID/ 669 8875 / MCSS.
550. PIANTO, D. & SERGEI S. (2004). Use of survey design for the evaluation of social programs: The PNAD and the program for the eradication of child

- labor in Brazil. Anais do XXXII Encontro Nacional de Economia 133, ANPEC - Associação Nacional do Centros de Pos-graduação em Economia.
551. PIEHL AM, COOPER SJ, BRAGA AA. & KENNEDY, DM.(1999) Testing for Structural Breaks in the Evaluation of Programs. NBER Working Paper N°. 7226.
 552. PLOTNICK, RD.; GARFINKEL, I.; MCLANAHAN, S. KU , I. (2006) The Impact of Child Support Enforcement Policy on Nonmarital Childbearing Evaluation of Aid to Community Associations. Evans School Working Papers Series N° 2006 – 09.
 553. POMEROY RS, POLLNAC RB, PREDO CD. & KANTON BM (1996) Impact Evaluation of Community-Based Coastal Resource Management Projects in The Philippines NAGA, The ICLARM Quarterly. 19(4): 9-12. Base de datos electrónica USAID
 554. POPPELE, J. SUMMARTO, S. & PRITCHETT, L. (1999). Social Impacts of the Indonesia Crisis: New Data and Policy Implications. Social Monitoring Early Response Unit, World Bank, Washington, D.C. Processed.
 555. PRADHAN, M. & RAWLINGS, L. (2002). The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund. *The World Bank Economic Review*, 16 (2): 275-295.
 556. PRADHAN, M., RAWLINGS. L. & RIDDER, G. (1998). The Bolivian Social Investment Fund: An Analysis of Baseline Data for Impact Evaluation. *World Bank Economic Review* 12 (3): 457–82.
 557. PRENNUSHI, G; RUBIO, G. & SUBBARAO, K. (2000). Seguimiento y Evaluación. En: Libro de Consulta para Estrategias de Reducción de la Pobreza. Banco Mundial, Washington D.C. Banco Mundial.
 558. PRODUCTIVITY COMMISSIONS (2003). Evaluation of the Pharmaceutical Industry Investment Program. Research Report. AusInfo, Canberra.
 559. PRZEWORSKI, A. & TEUNE, H. (1970). The logic of comparative social inquiry. John Willey & Son. NY.
 560. PUHANI, P. (1998). Advantage through Training? A Microeconomic Evaluation of the Employment Effects of Active Labour Market Programmes in Poland. CEPR Discussion Papers 2000, CEPR Discussion Papers.
 561. PUSHPANGADAN, K.(2002). Social Returns from Drinking Water, Sanitation and Hygiene Education: A Case Study of Two Coastal Villages in Kerala. Centre for Development Studies. WP. 279.
 562. QUISUMBING AR. (2003) Food Aid and Child Nutrition in Rural Ethiopia Food Consumption and Nutrition Division. Discussion Paper No. 158
 563. RABBANI M, PRAKASH VA, SULAIMAN M. (2006). Impact Assessment of CFPR/TUP: A Descriptive Analysis Based on 2002-2005 Panel Data. CFPR/TUP Working Paper Series No. 12
 564. RAMIO, C. (2001). Problemas de implantación de la Nueva Gestión Pública en América Latina. *Reforma y Democracia*, 12, CLAD. Disponible en Internet: [<http://www.clad.org.ve/reforma.html>].
 565. RANK D. & WILLIAMS D. (1999) Partial bene_t:cost in the evaluation of the Canadian Networks of Centres of Excellence .Evaluation and Program Planning, 22: 121-129.

566. RAO, P. & MILLER, R. (1971). *Applied Econometrics*. Belmont, C.A.: Wadsworth.
567. RAO, V. & IBÁÑEZ, AM. (2003) The Social Impact of Social Funds in Jamaica: A Mixed-Methods Analysis of Participation, Targeting and Collective Action in. World Bank Policy Research Working Paper 2970,
568. RAVALLION, M. & SHAOHUA C. (2005). Hidden Impact? Household Saving in Response to a Poor-Areas Development Project. *Journal of Public Economics*, 89: 2183-2204.
569. RAVALLION, M. & WODON, Q. (1998). Evaluating a targeted social program when placement is decentralized. World Bank Policy Research Paper N° 1945. World Bank, Development Research Group. Washington D.C.
570. RAVALLION, M. & WODON, Q. (1999). Does child labor displace schooling? Evidence on Behavioral responses to Enrollment Subsidy. World Bank Policy Research Paper N° 2116. World Bank, Development Research Group, Poverty and Human Resources and Latin America and the Caribbean Region, Poverty Reduction and Economic Management Sector Unit, Washington D.C.
571. RAVALLION, M. (2001). The Mystery of the Vanishing Benefits: An introduction to Impact Evaluation. *World Bank Economic Review*, 15 (1): 115-140.
572. RAVALLION, M. (2005). Evaluating Anti-Poverty Programs. *Policy Research Working Paper N° 3625*. Washington, D.C.: Banco Mundial.
573. RAVALLION, M., GALASSO, E., LAZO, T. & PHILIPP, E. (2001). Do Workfare Participants Recover Quickly from Retrenchment?" World Bank Policy Research Working Paper N° 2672. World Bank, Development Research Group, Poverty, Washington D.C.
574. RAVINA, R., PAULINI, J. & CANCHO, C. (2002). Costo efectividad del programa de desayunos escolares de FONCODES y el programa de alimentación escolar del PRONAA. Informe Final (sujeto a revisiones y comentarios). Documento no publicado.
575. RAWLINGS, L., & RUBIO, G. (2003). Evaluating the impact of Conditional Cash Transfer Programs: Evidence from Latin America. World Bank Policy Research Working Paper N° 3119. World Bank, Latin America and the Caribbean Region, Human Development Sector Unit, Washington D.C.
576. RAWLINGS, L.; SHERBURNE-BENZ, L. & VAN DOMELLEN, J. (2004). Evaluating Social Funds: A Cross-Country Analysis of Community Investments. Washington D.C.: World Bank.
577. REA S, MARTIN R, YOUNG M, & KRESS D. (1993) Evaluation of A.I.D. Family Planning Programs Tunisia Case Study. AID Technical Report N°, 15. Base de datos electrónica USAID
578. REGALIA, F. (1999) Impact evaluation methods for social programs. Poverty and Inequality Advisory Unit. Technical Note 2. Diciembre, 1.
579. REICHARDT, C. (1985). Reinterpreting Seaver's Study of Teacher Expectancies as a Regression Artifact. *Journal of Educational Psychology*, 77: 231-236.

580. REICHARDT, C. (1986). Estimating Effects. Manuscrito no publicado. Departamento de Psicología, Universidad de Denver.
581. REJDA, G. & SCHMIDT, J. (1997). The Impact of the Social Security Program on Private Pension Contributions. *Journal of Risk and Insurance*. Diciembre de 1997.
582. REPETTO, E. (1987). Evaluación de programas de orientación. Pp. 245-275. En: ÁLVAREZ ROJO, V. (Coord.). Metodología de la Orientación Educativa. Sevilla: Ediciones Alfar.
583. RICSE, C: (2000). Análisis sobre la situación de los recursos humanos. Documento Técnico N°2. OPS/Minsa. Lima.
584. RIECKEN, H. & BORUCH, R. (1974). Social Experimentation: A method for planing and Evaluating Social Intervention. New York: Academic.
585. RIECKEN, H. (1972). Memorandum on Program Evalaution. En WEISS, C. (Ed.). Evaluating Action Programs: Readings in social action and Education. Boston, MA: Allyn and Baoon.
586. RIVLIN, A. & TIMPANE, M. (1975). Planned variation in education. Washington DC, Brooking Institution.
587. ROBERTS JE, CLAPP-WINCEK C, BROKENSHEA DW. (1982) Kenya: Rural Roads. AID Project Impact Evaluation Report N° 26. Base de datos electrónica USAID
588. ROBSON, C. (1993). Real World Research. Londres: Ed. Blackwell.
589. RODRÍGUEZ, A. (2003). Evaluación del Impacto de los Programas de Bolsa Escolar en el Trabajo Infantil en Brasil. Tesis de Maestría. Escuela Superior de Agricultura Luiz de Queiroz. Sao Paulo.
590. RONCONI L, SANGUINETTI J, FACHELLI S, CASAZZA V. & FRANCESCHELLI I. (2006). Poverty and Employability Effects of Workare Programs in Argentina. PMMA Working Paper No. 2006-14 Available at SSRN: <http://ssrn.com/abstract=908566>.
591. ROSENBAUM, P. & RUBIN, D. (1983). The Central Role of Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70 (1): 41-55.
592. ROSENBAUM, P. & RUBIN, D. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, 79: 516-524.
593. ROSENBAUM, P. & RUBIN, D. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, 39: 33-38.
594. ROSHOLM, M. & SKIPPER, L. (2003). Is Labour Market Training a Curse for the Unemployed? Evidence from a Social Experiment IZA Discussion Paper No. 716
595. ROSSI, P. & FREEMAN, H. (1985). Evaluation: a systematic approach (3ª Edic.). Beverly Hills: Sage Publications.
596. ROSSI, P.; FREEMAN, H. & LIPSEY, M. (1999). Evaluation: A Systematic Approach. Sage Publications.
597. RUBIN, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized experiments. *Journal of Educational Psychology*, 66: 688-701.

598. RUBIN, D. (1977). Assignment to treatment on the basis of a covariate. *Journal of Education Statistics*, 2: 1-26.
599. RUBINSTEIN, R. (1981) Simulation and the Monte Carlo method. Wiley series in Probability and Mathematical Statistics. John Wiley and Sons Inc. New York.
600. RUEL MT, DE LA BRIERE B, HALLMAN K, QUISUMBING A. & COJ N. (2002). Does subsidized childcare help poor working women in urban areas? Evaluation of a Government-sponsored program in Guatemala city. International Food Policy Research Institute. Food Consumption and Nutrition Division. Discussion Paper N°. 131
601. RUEL MT. (2003) El Programa de Guarderías Comunes de Guatemala: Focalización Efectiva de la Ayuda en Alimentos en Áreas Urbanas. Instituto Internacional de Investigación sobre Políticas Alimentarias, Washington, D.C.
602. RUTHMAN, L. (1977). Planning useful evaluations. Evaluability assessment. Baberly Hills, Ca: Sage.
603. RYAN JG. & MENG X.(2004). The Contribution of IFPRI Research and the Impact of the Food for Education Program in Bangladesh on Schooling Outcomes and Earnings. Impact Assessment Discussion Paper N°. 22. International Food Policy Research Institute. Washington, DC.
604. RYAN JG. (1999) Assessing the impact of Rice Policy Changes in Viet Nam and the contribution of policy research. Impact Assessment Discussion Paper N°. 8 International Food Policy Research Institute. Washington, DC.
605. SAAVEDRA J. & TORERO M. (2000) Labor Market Reforms and Their Impact on Formal Labor Demand and Job Market Turnover: the case of Peru .Latin American Research Network Red de Centros de Investigación Research network Working paper #R-394
606. SAAVEDRA, J. & PASCÓ-FONT, A. (2001). Reformas estructurales y bienestar. Grupo de Análisis para el Desarrollo. GRADE. Lima.
607. SÁEZ, A., SUÁREZ, J., ALIAGA, F. & BO, R. (1994). La utilización de los procedimientos de comparaciones múltiples en la investigación educativa en España. *Revista de Investigación Educativa*, 23: 396-404.
608. SAHN, D., DOROSH, P. & YOUNGER, S. (1996). Exchange Rate, Fiscal and Agricultural Policies in Africa: Does Adjustment Hurt the Poor? *World Development*, 24 (4): 719-47.
609. SÁNCHEZ, J. (2000). Director del ILPES. Presentación en el Seminario de alto nivel sobre las funciones básicas de la planificación. La Habana, 16 al 17 de Noviembre.
610. SANDOVAL, J. & RICHARD, M. (2003). Los indicadores en la evaluación del impacto de programas. México. Sistema Integral de Información y Documentación.
611. SANGRA, E. (2000). Evaluation an civil society, the example of the Canton of Geneva. Paper presentado en la Cuarta Conferencia de la Sociedad Europea de Evaluación. Lausanne, Suiza. Octubre.
612. SARETSKY, G. (1972). The OEO P.C. experiment and John Henry effect. *Phi Delta Kappa*, 153: 589-591.

613. SAXE L, REBER E, HALLFORS D, KADUSHIN C, JONES D., RINDSKOPF D. & BEVERIDGE A. (1997) Think globally, act locally: assessing the impact of community-based substance abuse prevention Evaluation and Program Planning, Vol. 20, No. 3, pp. 357-366,
614. SCHEFFE, H. (1959). The Analysis of Variance. New York: Willey.
615. SCHIEFELBEIN, WOLFF L. & SCHIEFELBEIN, P. (1998) Cost-Effectiveness of Education Policies in Latin America: A Survey of Expert Opinion.
616. SCHOENI, R. & BLANK, R. (2001). What has welfare reform accomplished? Impact on welfare participation, employment, income, poverty, and family structure. *NBER Working Paper 7627*. Disponible en internet: [<http://www-personal.umich.edu/bschoeni/vjpam3.pdf>] Acceso el 21 de setiembre de 2004.
617. SCHOLLAR, E. (2001). A Review of Two Evaluations of the Application of the READ Primary Schools Program in the Eastern Cape Province of South Africa. *International Journal of Educational Research*, 35 (2): 205-216.
618. SCHUFTAN C, VAN DER VEEN A. & LOTHE P. (2003). Evaluation of Echo's 1999 to 2002 Funded Actions in Sudan. European Comisión Humanitarian Office. Final Report.
619. SCHULTZ TP. (2000a) El impacto de Progresía sobre la inscripción escolar. Internacional Food Policy Research Institute. Washington, DC.
620. SCHUTTA RK, ROSENHECK RE, PENK WE, DREBING CE. & SEIBYL CL. (2005) The social environment of transitional work and residence programs: Influences on health and functioning. *Evaluation and Program Planning* 28: 291-300.
621. SCHWERINA MJ, MICHAELB PG, GLASERC DN. & FARRAR KL. (2002) A cluster evaluation of Navy quality of life programs. *Evaluation and Program Planning*, 25: 303-312
622. SCRIVEN, M. (1967). The Methodology of Evaluation. En: TYLER et al. (1967). *Perspective of Curriculum Evaluation*. Chicago, Rand McNally.
623. SCRIVEN, M. (1974). Pros and cons about goal-free evaluation. *The Journal of Evaluation Comment*, 3(4): 1-4.
624. SCRIVEN, M. (1980). The logic of evaluation. Inverness, California. Edgepress.
625. SEAVER, W. & QUARTON, R. (1976). Regression-discontinuity analysis of Dean's List effects. *Journal of Educational Psychology*, 68: 459-465.
626. SEDLACEK G, YAP, Y. & ORAZEM P. (2000). Evaluating the Impact of PETI on Child Labor Supply and Schooling Demand in Rural Northeastern Brazil: The Case of Pernambuco, Bahia and Sergipe. World Bank Background Paper. Washington, D.C., August. Summary of paper in "Eradicating Child Poverty in Brazil." 2001. Report N° 21858-BR. World Bank, Human Development Department Brazil Country Management Unit, Latin America and Caribbean Regional Office, Washington DC.
627. SENATE STANDING COMMITTEE ON SOCIAL WELFARE (1979a). Through a Glass Darkly: Evaluation in Australian Health and Welfare Services, Volume 1. The Report. Canberra, Australian Government Publishing Service.

628. SHACK, N. (2000). La estrategia de lucha contra la pobreza. En: La Reforma incompleta. Instituto Peruano de Economía y Control de Investigación de la Universidad del Pacífico. Lima, Perú.
629. SHADISH, W.; COOK, T. & CAMPBELL, D. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin Company, Boston MA.
630. SHADISH, W.; COOK, T. & HOUTS, A. (1986) Quasiexperimentation in a Critical Multiplist Mode. En W. Trochim (Ed.), *Advances in Quasi-Experimental Design and Analysis*. San Francisco: Jossey-Bass.
631. SHADISH, W.; COOK, T. & LEVINTON, L. (1991). *Foundations of program evaluation*. New York: Sage Publication.
632. SHAOHUA, CH. & RAVALLION, M. (2003). Hidden Impact? Ex-Post Evaluation of an Anti-Poverty Program. *World Bank Policy Research Working Paper N° 3049*. Development Research Group, World Bank, Washington D.C.
633. SHAPIO, J. & MORENO, J. (2004). An Impact Evaluation Using Propensity Score Matching. World Bank.
634. SHARMA AN, SHARMA R. & RAJ N. (2000). The Impact of Social Labelling on Child Labour in India's Carpet Industry. Institute for Human development New Felhi. ILO/IPEC Working Paper.
635. SHEA B., DUBE C. & MOHER D. (2001). Assessing the quality of reports of systematic reviews: The QUORUM statement compared to other tools. En: EGGER M., SMITH G. & ALTMAN D. (Ed.). *Systematic Reviews in Health Care. Meta-analysis in context*. London: BMJ, p. 122-139.
636. SHERWOOD-FABRE L, GOLDBERG H. & BODROVA V. (2002) The Impact of an Integrated Family Planning Program in Russia. *Evaluation Review*, 26 (2): 190-212.
637. SHERWOOD-FABRE, L., GOLDBERG, H. & BODROVA, V. (2002). The Impact of an Integrated Family Planning Program in Russia. *Evaluation Review*, 26 (2): 190-212.
638. SHINKAI N. (2000) Security and Income Affect the Living Arrangements of the Elderly? Evidence from Reforms in Mexico and Uruguay. Research Department Working paper series ; 432. Base de datos electrónicos BIF.
639. SHIRLEY, M., COLIN, X. & ZULUAGA, A. (2000). Reforming urban water supply: the case of Chile. World Bank Policy Research Working Paper N° 2294. World Bank, Development Research Group, Regulation and Competition Policy, Washington D.C.
640. SHU-HONG, Z. (1999). A method to obtain a randomized control group where it seems impossible A Case Study in Program Evaluation. *Evaluation Review*, 23 (4): 363-377.
641. SIANESI, B. (2003) An Evaluation of the Swedish System of Active Labour Market Programs in the 1990s. The Institute For Fiscal Studies WP02/01
642. SIANESI, B. (2003). Differential Effects of Swedish Active Labour Market Programmes for Unemployed Adults During the 1990s .The Institute For Fiscal Studies WP01/25

643. SINHA, N. (2003). Fertility, child work and schooling consequences of family planning programs: evidence from an experiment in rural Bangladesh. Economic Growth Center Yale University. Center Discussion Paper N°. 867
644. SKOUFIAS E. & DI MARO V. (2006). Conditional Cash Transfers, Adult Work Incentives, and Poverty. World Bank Policy Research Working Paper 3973.
645. SKOUFIAS E. & MCLAFFERTY B. (2001) Is progresas working? summary of the results of an evaluation by IFPRI. Internacional Food Policy Research Institute. Washington, DC.
646. SKOUFIAS, E. (2001). PROGRESA and its impacts on the human capital and welfare of households in rural Mexico: A synthesis of the results of an evaluation by IFPRI. International Food Policy and Research Institute, Washington D.C.
647. SKOUFIAS, E. (2003). Importancia de la evaluación de los programas sociales: el ejemplo de Oportunidades: Conferencia impartida en el Instituto Tecnológico Autónomo de México, México, D.F., 25 de abril.
648. SLAVIN, R. (1987). Best-evidence Synthesis: Why Less is more. *Educational Researcher*, 16 (4): 15-16.
649. SMART, R. & MANN, R. (2000). The impact of programs for high-risk drinker on population levels of alcohol problem. *Addiction*, 95 (1): 37-52.
650. SMITH EA., SWISHER JD, & VICARY JR. (2004) Welfare Reforms, Family Resources, and Child Maltreatment Evaluation of Life Skills Training and Infused-Life Skills Training in a rural setting: Outcomes at two years. *J Alcohol Drug Educ.* Vol. 48 No1, 51-70.
651. SMITH, J. (2000). A critical survey of empirical methods for evaluating active labour market policies. *Swiss Journal of Economics and Statistics*, 136 (3): 1-22.
652. SNODGRASS, DR. & SEBSTAD, J. (2002) Clients in context: The impacts of microfinance in three countries. Assessment the Impact of Microenterprise Services (AIMS). Washington.
653. SNOW, R. (1979). Diseños representativos y cuasirepresentativos para la investigación en la enseñanza. En F. ALVIRA, M.D. AVIA; R. CALVO Y F. MORALES (Eds.). Los dos métodos de las Ciencias Sociales. Madrid: Centro de Investigaciones Sociológicas.
654. SOARES F. & SOARES. Y. (2005) The Socio-Economic Impact of Favela-Bairro: What do the Data Say? OVE. Working Paper – 08.
655. SOLARTE, L. (2002). La evaluación de programas sociales en el Estado Liberal. VII Congreso del CLAD sobre la Reforma y de la Administración Pública, Lisboa, Portugal, 8-11 de Octubre.
656. SOMMER JG, AQUINO R, FERNÁNDEZ C, GOLAY FH, SIMMONS E, CASTILLO GH. & ROCO CH. (1982). AID Project Impact Evaluation Report N° 28.
657. STADLER J. & DELANEY S. (2004). The ‘healthy brothel’: the context of clinical services for sex workers in Hillbrow. Reproductive Health Research Unit, CHBH, Soweto.

658. STAKE, R. (1975a). Program evaluation: particularly responsive evaluation. *Occasional Paper*, 5. University of Western Michigan.
659. STAME, N. (2003). Evaluation and the policy context: the European experience. *Evaluation Journal of Australasia*, 3 (2): 36 - 43
660. STANTON BF, LI X, KAHIHUATA J, FIRZGERALD AM, NEUMBO S, KANDUUOMBE G, RICARDO IB, GALBRAITH JS, TERRERI N, GUEVARA I, SHIPENA H, STRIJDOM J, CLEMENS R. & ZIMBA RF. (1998). History Workshop, African Studies Seminar.
661. STEELE F, CURTIS SL. & CHOE M. (1999) The Impact of Family Planning Service Provision on Contraceptive-Use Dynamics in México. *Studies in Family Planning*, vol. 30, 1: 28-42.
662. STEELE, F, AMIN S. & NAVED RT (2001) Savings/credit group formation and change in contraception. *Demography*, Vol. 38-Nº 2, 267- 282.
663. STEINBERG DI, MORROW RB. & DONG-IL K.(1980) Korean Irrigation. *AID Project Impact Evaluation Report Nº 12*. Base de datos electrónica USAID
664. STEVENS FI. (1995). Preliminary Impact Evaluation Report of NGO Educare Training in South Africa. Center for Human Capacity Development United States Agency for International Development. Base de datos electrónica USAID
665. STIFEL D. & ALDERMAN H. (2003) The “Glass of Milk” Subsidy Program and Malnutrition in Peru. World Bank Policy Research Working Paper 3089. Base de datos electrónicos BIF.
666. STROMQUIST NP, KLEES S. & MISKE S. (1999) Improving girls’ education in Guatemala. CDIE Impact Evaluation, PN-ACA-919. Base de datos electrónica USAID
667. STUFFLEBEAM, D. & SHINKFIELD, A. (1987). Evaluación sistemática. Guía teórica y práctica. Barcelona: Paidós/MEC. España.
668. STUFFLEBEAM, D. (1966). A depth study of the evaluation requirement. *Theory into Practice*, 5 (3): 121-134.
669. STUFFLEBEAM, D. (1989). Evaluación sistemática. Madrid. Paidós.
670. STUFFLEBEAM, D. et al. (1971). Educational evaluation and decision making. Itasca, III, Peacock.
671. STUFFLEBEAM, D., & WEBSTER, W. (1980). An analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis*, 2 (3), 5-19.
672. SUÁREZ, J.; ALIAGA, F.; ORELLANA, N.; SALAVERT, L.; SÁEZ, A.; BELLOCH, C. & BO, R. (1995). La utilización de los procedimientos de medida del cambio en la investigación educativa en España. En AIDIPE, Estudios de investigación Educativa en intervención Psicopedagógica. Valencia: AIDIPE.
673. SUBBARAO, K.; EZEMENARI, K.; RANDA, J. & RUBIO, G. (1999). Impact Evaluation in FY98 Bank Projects: A Review. World Bank Poverty Reduction and Economic Management Network, Processed, January.
674. SULBRANDT, J. (1993). La evaluación de los programas sociales: una perspectiva crítica de los modelos actuales. En: Kliksberg, B. (comp.), Pobreza,

- un tema impostergable: nuevas propuestas a nivel mundial, México, CLAD/FCE/ Programa de Naciones Unidas.
675. SUPOVITZ, J. (2005). Systemic Education Evaluation Evaluating the Impact of Systemwide Reform in Education. *American Journal of Evaluation*, 26 (2): 204-230
676. SUPOVITZ, J. (2005). Systemic Education Evaluation Evaluating the Impact of Systemwide Reform in Education. *American Journal of Evaluation*, 26 (2): 204-230.
677. SUTTON M, TIETJEN K, BAH A & KAMANO P. (1999) Promoting primary education for girls in Guinea. CDIE Impact Evaluation, PN-ACA-915. Base de datos electrónica USAID
678. TALLMADGE, G. (1982). An empirical assessment of norm-referenced evaluation methodology. *Journal of Educational Measurement*, 19: 97-112.
679. TALMAGE, H. (1982). Evaluation of programs. En MILTZEL, H. (Ed.). *Encyclopaedia of educational research*. Macmillan 4, pp. 592-661.
680. TASCHEREAU, S. (1998). Evaluating the Impact of Training and Institutional Development Programs, a Collaborative Approach. Economic Development Institute of the World Bank, enero.
681. TERUEL G. & DAVIS B. (2000). Una evaluación del impacto de los apoyos en efectivo de Progresía sobre las transferencias privadas entre los hogares. Washington, D.C.: International Food Policy Research Institute (IFPRI)
682. THE JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION. (1994). The program evaluation standards. The Evaluation Center. Western Michigan University. Kalamazoo, Michigan. Thousand Oaks, CA. Sage Publications Inc.
683. THE PEOPLE'S COMMITTEE OF THUA THIEN HUE PROVINCE. Summary of Evaluations Of Norad Financed Projects.
684. THE WORLD BANK. (1996). Participation sourcebook. Washington. D.C.
685. THISTLETHWAITE, D., & CAMPBELL, D. (1960). Regression-Discontinuity Analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology*, 51: 309-17.
686. THOMPSON, M. (1980). Benefit-Cost Analysis for Program Evaluation. Sage.
687. TILLEY, N. (2000). Realistic Evaluation: An Overview. Ponencia presentada en el Founding Conference of the Danish Evaluation Society, Septiembre. Disponible en internet: [http://www.danskevalueringsselskab.dk/Materiale_fra_DES.asp]. Acceso el 18 de diciembre de 2005.
688. TILNEY JS & RIORDAN JT (1988). Agricultural Policy Analysis and Planning: A summary of two recent analyses of a.i.d.-supported projects worldwide.AID. Evaluation Special Study N°. 55. Base de datos electrónica USAID
689. TORERO, M. & PASCO-FONT, A. (2001). The Social Impact of privatization and the Regulation of Utilities in Peru." Discussion Paper No.

- 2001/17. United Nations, United Nations University, World Institute for Development Economics Research, New York.
690. TORRES G, ISAZA L, & CHÁVEZ L. (2004). Evaluación del Impacto en las Instituciones Escolares de los Proyectos Apoyados por el Instituto para la Investigación Educativa y el Desarrollo Pedagógico "IDEP", de Bogotá. *Revista Digital Umbral 2000*, 15.
691. TORRES, R., QUISPE, E. & SERRIE, N. (2006). Convirtiendo promesas en evidencia. Programa de Caminos Rurales II y Programa de Transporte Rural Decentralizado. Ponencia presentada en el Taller organizado por el Banco Mundial en Cuernavaca, Mayo de 2006.
692. TOULEMONDE, J. (2000). Evaluation culture(s) in Europe: differences and convergence between national practices. Paper for Vierteljahrshefte zur Wirtschaftsforschung. Berlín.
693. TROCHIM, W. & CAPPERLLERI, J. (1992). Cutoff Assignment Strategies for Enhancing Randomized Clinical Trials. *Controlled ClinicalTrials*, 13:190-212.
694. TROCHIM, W. & DAVIS, J. (1986). Computer simulation for program evaluation. *Evaluation Review*, 5 (5): 609-634.
695. TROCHIM, W. (1984) Research design for program evaluation. Beverly Hills: Sage.
696. TROCHIM, W. (1986). Advances in Quasi-Experimental Design and Analysis. San Francisco: Jossey-Bass.
697. TRULSON, M. (1986). Martial arts training: A novel "cure" for juvenile delinquency. *Human Relations*, 39(12): 1131-1140.
698. TUIJMAN, A. & KEEVES, J. (1997). Path analysis and Linear Structural Relations Analysis. En J. KEEVES (Ed.), Educational Research, Methodology and Measurement. An International Handbook. 2ª Edición. Londres: Pergamon.
699. TUIRÁN, A. & MEDINA, A. (2001). El MIC estrategia para mejorar las estimaciones por línea de pobreza y elaboración de recomendaciones para enfrentarla. Material mimeografiado.
700. TYLER, R.W. (1942). General statement on evaluation. *Journal of Educational Research*, 35: 492-501.
701. U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT (1994) A Synthesis of tour legacy/impact studies of USAID assistance to Cameroon. Camerú. Base de datos electrónica USAID
702. UNDERWOOD C, HACHONDA H, SERLEMITSOS E. & BHARATH U. (2001). Impact of the Heart Campaign: Findings from the youth surveys, 1999-2000. Baltimore: Johns Hopkins School of Public Health, Center for Communication Programs.
703. UNITED STATUS GENERAL ACCOUNTING OFFICE (1998) Performance measurement and evaluation: Definitions and relationships. Abril. GAO/GGD-98-26.
704. URWIN P, JACK G. & LISSENBURGH S. (2006) The impact of the National Minimum Wage in low-wage sectors: does the Earnings Top-up Evaluation study add to our understanding? .Industrial Relations Journal. Vol. 37 N°. 3: 259-277.

705. USDIN S, SCHEEPERS E, GOLDSTEIN S. & JAPHET G. (2005) Achieving social change on gender-based violence: A report on the impact evaluation of Soul City's fourth series. *Social Science & Medicine*, 61: 2434–2445.
706. VALADEZ, J. & BAMBERGER, M. (1994) *Monitoring and Evaluation Social Programs in Developing Countries*. Washington: The World Bank.
707. VALDIVIA M. (2004) Poverty, Health Infrastructure and the Nutrition of Peruvian Children. Latin American Research Network Red de Centros de Investigación Research Network Working Paper #R-498.
708. VAN DE WALLE, D. & CRATTY, D. (2005). Do Donors Get What They Paid For? Micro Evidence on the Fungibility of Development Project Aid. World Bank Policy Research Working Paper N° 3542. World Bank, Washington D.C.
709. VAN STEENWYK, N. (1984) *Impact Evaluation; LAC/ Honduran Training Program*. Base de datos electrónica USAID
710. VARA-HORNA, A. (2006). La lógica de la investigación en ciencias sociales. *Manual de Investigación y Estadística Avanzada para Científicos Sociales*. Tomo I. Lima: Asociación por la Defensa de las Minorías. Libro electrónico disponible en internet: [http://www.aristidesvara.com/libros/libro_a.htm] Acceso el 23 de junio de 2006.
711. VASQUEZ E. & FIGUEROA C. (2000). Documento base de discusión para el diseño de una Estrategia de Seguridad alimentaria en el Perú. 2000-2005. Universidad del Pacífico. Lima.
712. VASQUEZ, E. & MENDIZABAL, E. (2002). Los niños... primero? El gasto público social focalizado en niños y niñas en el Perú 1990-2000. Centro de Investigación de la Universidad del Pacífico y Save The Children. Lima.
713. VASQUEZ, E., CORTEZ, R. & RIESCO, G. (2000). Inversión social para un buen gobierno en el Perú. Centro de Investigación de la Universidad del Pacífico. Lima.
714. VEDUNG, E. (1997). *Public Policy and Program Evaluation*. Transaction Publishers.
715. VELA, R. (2003). Hacia un Nuevo enfoque de la evaluación de impacto de proyectos de desarrollo rural. *Cuadernos de Desarrollo Rural*, 50: 125-142.
716. VENTOSA, V. (1992) *Evaluación de la animación sociocultural*. Guía de orientación para animadores. Madrid, Popular.
717. VERMEERSCH, C. (2006). Sesión III: Diseño de regresiones en discontinuidad. Ponencia presentada en el Taller organizado por el Banco Mundial en Cuernavaca, Mayo de 2006.
718. VERMEERSCH, C. (2006b). Sesión V: Variables instrumentales. Ponencia presentada en el Taller organizado por el Banco Mundial en Cuernavaca, Mayo de 2006.
719. VERMEERSCH, CH. & KREMER, M. (2004). School Meals, Educational Achievement, and School Competition: Evidence from a Randomized Evaluation. Policy Research Working Paper N° 3523. World Bank, Washington D.C

720. VERSTRAETE, L. (1993). Propuesta metodológica para la evaluación ex post y el informe de término de los proyectos de inversión. Documento de la Dirección de Proyectos y Programación de Inversiones del ILPES.
721. VISSER, R. & DE LEEUW, J. (1984). Maximum Likelihood Analysis for a Generalized Regression-discontinuity Design. *Journal of Educational Statistics*, 9: 45-60.
722. VISSER, R. (1985). Analysis of longitudinal data in behavioural and social research. An expository survey, Leiden, DSWO Press.
723. VIVEROS AM. & BECERRA AM. (1981) Peru: CARE OPG Water Hdth Services Project. AID Project Impact Evaluation Report N°. 24. Base de datos electrónica USAID.
724. VIVO, S. (2006). Simulación de análisis de evaluación de impacto. Sesiones aplicadas del I al V. Ponencias presentadas en el Taller organizado por el Banco Mundial en Cuernavaca, Mayo de 2006.
725. WAISSBLUTH, M. (2002). La reforma del Estado en América Latina: Guía abreviada para exploradores en la jungla. Programa Latinoamericano de Gerencia Pública. Universidad de Chile.
726. WALKER, H., GOLLY, A., ZOLNA, J. & KIMMICH, M. (2005). The Oregon First Step to Success Replication Initiative: Statewide Results of an Evaluation of the Program`s Impact. *Journal of Emotional and Behavioral Disorders*; 13 (3): 163-172.
727. WALKER, I., CID, R., ORDONEZ, R. & RODRÍGUEZ, F. (1999). Ex-Post Evaluation of the Honduran Social Investment Fund (FHIS 2). Elaborado por ESA Consultants, Honduras, para el World Bank, Latin American and Caribbean Region (LCSHD).
728. WALKER, J. & EVERS, C. (1997). Research in Education: Epistemological Issues. En J. Keeves (Ed.) *Educational Research, Methodology and Measurement. An International Handbook*. 2ª Edición. Londres: Pergamon.
729. WASHINGTON STATE INSTITUTE FOR PUBLIC POLICY (2001) Foundations for Learning: Safe and Civil Schools Project. Disponible en www.wsipp.wa.gov.
730. WASTAFF, A. & SHENGCHAO, Y. (2005) Do Health Sector Reforms Have Their Intended Impacts? The World Bank`s Health VIII Project in Gansu Province, China. World Bank Policy Research Working Paper 3743.
731. WEISS, C. (1983). The stakeholder approach to evaluation: origins and promise. En BRYK, A. (Ed.). *Stakeholder-based evaluation* (pp. 3-14). San Francisco: Jossey - Bass.
732. WEISS, C. (1998). Writing the report and disseminating results. En: WEISS, C. *Evaluation* (2ªEd.). Upper Saddle River, N.J.: Prentice-Hall, pp. 294 – 319.
733. WEISS, D. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6: 473-492.
734. WEISS, R. & REIN, M. (1972). The evaluation of broad-aim programs: Difficulties in experimental design and an alternative. En C. WEISS (Ed.),

- Evaluationaction programs: Readings in social action and education. Boston: Allyn & Bacon.
735. WHITE, H. (2006). Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank. Washington, D.C.: Banco Mundial.
736. WHOLEY, J. (1992). ¿What can we actually get from program evaluation? *Policiy Sciencie*, 3, p 361-369. Citado por: MENY, Yves; THOENIG Jean C. *Las políticas públicas*. Barcelona: Editorial Ariel. P. 201
737. WIENERT, F. (1997). Translating Research into Practice. En J.P. KEEVES (Ed.), *Educational Research, Methodology and Measurement. An International Handbook*. 2ª Edición. Londres: Pergamon.
738. WILKINSON JL, MCKEAN C, MEYER RE, NUNBERG BS, WEIL B. & MARTINEZ H. (1984). Perú: Improved Water and Land Use in the Sierra. A.I.D. Project Impact Evaluation Report No. 54. Base de datos electrónica USAID
739. WILSON SJ. & LIPSEY MW. (2000) Wilderness challenge programs for delinquent youth: a meta- analysis of outcome evaluations *Evaluation and Program Planning*, 23: 1-12
740. WOLF PJ, PETERSON PE, WEST MR. (2001) Results of a School Voucher Experiment: The Case of Washington, D.C. After Two Years . Kennedy School of Government, Faculty Research. RWP02-022
741. WOOTEN J, JANSEN W, KOHASHI WARREN M.(1982) Project Impact: A low-cost alternative for universal primary education in the philippines. Project Impact Evaluation Report No. 38. Base de datos electrónica USAID
742. WORLD BANK. (2000). Nicaragua Ex-Post Evaluation of the Emergency Social Investment Fund. Economic Report N° 20400-NI. World Bank, Washington D.C.
743. WORTMAN, P.; REICHARDT, C. & PIERRE, R. (1976). The first year of the education voucher demostration. *Evaluation Quarterly*, 2, 193-214.
744. YAMADA, G. & PEREZ, P. (2005). Evaluación de impacto de proyectos de desarrollo en el Perú. Centro de Investigaciones de la Universidad del Pacífico. Serie Apuntes de Estudio N° 61. Lima.
745. YAMANO, T., ALDERMAN, H. & CHRISTIAENSEN, L. (2003). Child Growth, Shocks, and Food Aid in Rural Ethiopia. World Bank Policy Research Working Paper N° 3128. World Bank, Washington D.C.
746. YANOVITZKY, I.; ZANUTTO, E. & HORNIK, R. (2005). Estimating causal effects of public health education campaigns using propensity score methodology. *Evaluation & Program Planning*, 28 (2): 209-220.
747. YAP, Y., SEDLACEK, G. & ORAZEM, P. (2002). Limiting Child Labor Through Behavior-Based Income Transfers: an experimental evaluation of the PETI Program in rural Brazil. En: ORAZEM, P., SEDLACEK, G. & TZANNATOS, Z. (Eds.), *Child labor in Latin America*. Washington, DC: World Bank and Inter-American Development Bank.
748. ZAJONC, R. & MARCUS, H. (1975). Birth order and intellectua development. *Psychological Review*, 82: 74-88.