

EL DISEÑO EXPERIMENTAL EN EL CONTEXTO DE LA GESTIÓN: NOTAS SOBRE SU IMPLEMENTACIÓN EN LA EVALUACIÓN DE POLÍTICAS PÚBLICAS. EXPERIENCIAS, ALCANCES Y LIMITACIONES

Adriana Fassio

María Gabriela Rutty

1. LA EVALUACIÓN DE POLÍTICAS SOCIALES Y LA EVALUACIÓN COMO PARTE DE LA GESTIÓN

Evaluación, desde nuestro punto de vista, no significa plantearse solamente los errores cometidos en procesos complejos como los de implementación de políticas sociales, sino por el contrario, nuestra visión es la de aprendizaje, es decir la de sistematizar las buenas prácticas y mejorarlas y de analizar lo que se considera que no resultó de acuerdo a lo esperado para superarlo. Es decir que la evaluación consistiría en un proceso a partir del cual se explora de manera sistemática y rigurosa el cumplimiento de las actividades y de la entrega de productos o servicios en función del logro de un cambio sostenible. La evaluación se constituye en una herramienta básica de la gerencia social adaptativa y orientada a resultados que intenta medir la generación de valor social, es decir, la mejora de la calidad de vida de la población sujeto de intervención. Es parte de un proceso integral de la gerencia de programas y organizaciones. Desde esta concepción la evaluación integral trascendería el paradigma meramente administrativo formal que solo indaga sobre el cumplimiento de las actividades y /o entrega de los productos y servicios (Mokate, 2003:2-9).

La información que se obtiene a través de la evaluación permite mejorar la efectividad cuantitativa de los programas, la calidad de los resultados, aumentar la eficiencia, adecuar los objetivos del servicio a las expectativas de la población objetivo, controlar la intervención de factores que dificultan la marcha o el logro y comprobar la relación entre el servicio

y los resultados buscados (Briones, 1999). Es decir, comprobar que los esfuerzos que se realizan están bien orientados y reajustar las acciones de ser necesario (función de retroalimentación de la evaluación).

Desde la década de los 90 en adelante el rol tradicional de la evaluación perseguía la indagación por los resultados, ocasionalmente los impactos, la rendición de cuentas (*accountability*) y las lecciones aprendidas.

Las lecciones aprendidas, se refiere al conocimiento ganado por medio de la reflexión sobre una experiencia o proceso a partir de la evaluación puesta en práctica. Permiten la identificación de factores de éxito, de deficiencias, y la resolución de problemas a través de nuevos cursos de acción. A su vez, mejorar la toma de decisiones futuras y de esta manera contribuir a la realización de Buenas Prácticas, entendidas como soluciones eficientes para resolver un problema.

Las tendencias actuales en evaluación a nivel mundial han planteado un cambio importante en las prioridades, los modelos de evaluación tradicionales de tipo experimental (*randomized scientific experimental trial*) ya no son vistos como los únicos apropiados. Hoy se enfatiza la necesidad de una mayor flexibilidad y métodos menos restrictivos para captar la complejidad de las iniciativas.

Se reconoce la existencia de un incremento en el conjunto de consumidores de la evaluación, lo que implica la atención de distintos intereses, no solo el de los financiadores de los proyectos. Entre los nuevos intereses se privilegia, además del logro de los resultados, la efectividad de las acciones y la evaluación como un proceso de *empowerment*. Desde esta mirada la evaluación es vista como un camino a través del cual grupos marginales y sin poder son capaces de adquirir habilidades e influencia a través de su compromiso en ella (Conlin y Stirrat, 2008).

Luego de muchos años, en los cuales los especialistas intentaron convencer a los distintos auditorios sobre la necesidad de la evaluación, hoy se observa el advenimiento de una corriente bastante crítica, donde algunos de los autores señalan como necesario, frente a una tarea de

evaluación, volver a emplear la estrategia de la estimación previa de su *evaluabilidad* (Dahler-Larsen, 2007). La principal cuestión en la estimación de la *evaluabilidad* no es si se puede hacer una evaluación, sino, si es *racional* hacerlo bajo las circunstancias del momento a la luz de las mejoras esperadas de ella.

La *evaluabilidad* hace referencia a si existe una descripción clara del programa (si no existe, es mejor usar los esfuerzos para clarificar el programa), si está bien implementado el programa (si ya se sabe que no es el caso, la evaluación será siempre negativa), si existe una buena teoría del programa (si no la hay, mejor clarificar el marco teórico subyacente), si existen metas plausibles y bien escritas (si no existen, el resultado de la evaluación es predecible), si hay datos relevantes accesibles (si no hay, los recursos de la evaluación podrían gastarse mejor en alternativas a la evaluación), si se han identificado oportunidades para mejorar el programa (si los usuarios no son capaces o no utilizan los resultados de la evaluación, no tiene sentido realizarla). La idea que alimenta la *evaluabilidad* es que la evaluación debe ser utilizada solamente cuando sea conveniente y cuando la información que produzca sea utilizada para "marcar una diferencia" en provecho del programa.

Uno de los aspectos importantes respecto de la estimación de la *evaluabilidad* es la comprobación de la existencia de disponibilidad de información para la realización de la evaluación. Este aspecto es vital a la hora de utilización de métodos experimentales en dicho proceso.

2. DISEÑOS EXPERIMENTALES

Tanto la evaluación de procesos (monitoreo) como la evaluación de impacto (en el curso del proceso o al finalizar el mismo) de los proyectos/políticas sociales pueden desarrollarse a partir de la implementación de métodos experimentales (Cohen y Franco, 1997:120-127).

Se ha discutido bastante la viabilidad de su implementación, fundamentalmente a partir de la selección de dos poblaciones equivalentes, por lo menos respecto de los factores considerados

pertinentes para la medición de los efectos del proyecto. En esta dirección la aleatorización de las muestras distribuiría de modo uniforme los efectos no controlados sobre los resultados obtenidos (tanto en el grupo semejante experimental como en el de control).

En el marco de los métodos experimentales la evaluación de impacto requiere por lo menos de dos mediciones (antes y después) y controlar los efectos netos del proyecto realizado. Es decir que la función de la evaluación es la de establecer que los efectos producidos en la realidad a partir de la implementación del proyecto están invariablemente relacionados con sus acciones. Para ello es necesaria la lógica de la prueba que muestra como la teoría, a través de su puesta en juego a partir de un proyecto, modifica la realidad a partir del principio de variación concomitante. Desde esta perspectiva los autores afirman que se trabaja sobre la causa eficiente, es decir la condición necesaria y suficiente para la aparición de un evento. Los diseños experimentales, a partir de la prueba establecen la covariabilidad de variables, es decir de las causas que producen determinados efectos.

Hay validez en el experimento cuando se mide lo que se tenía planificado medir, es decir cuando las diferencias en la medición de las variables del objeto de estudio reflejan diferencias en el atributo medido. Encontramos dos requisitos: a) validez interna: medir suficientes casos de mi unidad de análisis y b) validez externa: la posibilidad de generalizar los resultados obtenidos, ambas relacionadas con el N muestral y la construcción de una muestra probabilística. Hay confiabilidad si repetidas observaciones del fenómeno arrojan datos constantes. Validez y confiabilidad se construyen a partir de principios lógicos diversos, aunque en la práctica la alta validez (repeticiones de las mediciones) podría constituirse en una aproximación a la confiabilidad.

3. SU APLICACIÓN EN LA EVALUACIÓN DE PROGRAMAS/ PROYECTOS SOCIALES

Una de las razones fundamentales que dificulta la aplicación del modelo experimental es la aleatorización de los casos que conforman cada uno de los grupos, puesto que lograr esto las muestras deben ser mayores que

en otros diseños y por lo tanto más costosas. Por otro lado, en algunas oportunidades plantean dilemas éticos (quien dejar fuera y quien dentro; consideraciones respecto del respeto a los derechos humanos, el respeto al consentimiento informado, etc.), o no hay posibilidades de conformar ambos grupos.

En general lo más habitual es la puesta en juego de modelos cuasi experimentales de evaluación en los que los miembros de los grupos experimental y de control no son seleccionados aleatoriamente.

Si no existe una tendencia metodológica dominante, producto de la incertidumbre que caracteriza a las ciencias sociales y al campo de la acción social, a partir de los cuales conviven paradigmas diversos, en algunos casos hasta contradictorios, tampoco existe un único modelo de evaluación de los impactos o procesos de las acciones llevadas adelante. No se pueden trasladar modelos de evaluación provenientes de proyectos de inversión en obras de ingeniería (construcción o infraestructura) muy poco adaptables a los escenarios en los que los proyectos sociales tienen lugar y que no solamente han generado resistencias para su implementación sino que no han servido para medir el impacto de dichos proyectos de modo integral. A pesar de ello los organismos de financiamiento, en particular las agencias internacionales, los exigen (Niremberg, Brawerman y Ruiz, 2003^a:41).

Sin embargo, el experimento es viable y pertinente en un marco evaluativo integral en el que a partir de la triangulación se proponga a) la superación de la brecha entre lo cuali y lo cuantitativo; b) la identificación de indicadores pertinentes, en función del marco conceptual y la variedad de intereses y perspectivas asociadas a los distintos actores intervinientes; c) la definición y manejo de flujos de información originada en el proceso de evaluación y el incentivo para su uso por parte de los actores relevantes (Mokate, 2003:14).

4. ANÁLISIS DE CASOS: EL CASO DE LA EVALUACIÓN DE PROGRAMAS SOCIOEDUCATIVOS³

Cuando se aborda el diseño de una evaluación de programas socioeducativos, salvo que se trabaje con datos de mucha agregación, es necesario comprobar la disponibilidad de la información que se requiere antes de pensar en la implementación de estudios experimentales. Cuando se prevé la aplicación de una evaluación con la utilización de métodos experimentales, la experiencia de trabajo en el campo señala el empleo de una serie de recaudos previos. Una de las primeras cuestiones que se debe abordar es asegurarse que se posee la información suficiente para llevar a cabo esa tarea. Esta afirmación puede parecer sobranter, y no pertinente para evaluadores formados, pero es muy frecuente haber previsto contar con información que luego, cuando el evaluador llega al campo, no es posible relevar.

Muchos de los programas y proyectos de intervención en escuelas persiguen como objetivo lograr una mejora en la cobertura, la calidad del servicio educativo, o en la eficacia de la escuela⁴. En general en evaluación de programas educativos, no solo se pretende indagar sobre el logro de los objetivos, sino que también se pretende indagar si las estrategias puestas en marcha han sido las adecuadas, sobre los obstáculos encontrados, y la identificación de evidencia razonable sobre los resultados logrados por el programa.

Es frecuente que muchos de los programas que se implementan en el sistema educativo nacional y sistemas educativos provinciales se realicen con financiamiento mixto, (de origen nacional y externo). Es así que los organismos financiadores muchas veces son proclives a la utilización de técnicas de tipo experimentales para poder comprobar la incidencia del programa en la escuela, más aún, en muchos casos su utilización forma parte de los requisitos que integran los marcos contractuales exigidos para la evaluación.

³ Para la elaboración de este punto fue consultada la Lic. Lilia Toranzos, experta en evaluación.

⁴ Esto supone una mejora en los valores llamados de **rendimiento educativo**: *promoción, repitencia, retención, sobre edad, etc.*

Desde el planteo inicial pareciera posible y relativamente simple la utilización de grupo control a partir de la selección de escuelas similares que no hubieran participado del programa (*recibido el estímulo*), para comparar en tiempos distintos (momento 0 y momento 1 en un diseño antes -después) la ganancia de uno y otro grupo de escuelas y de esta forma evitar los problemas de *atribución*. Es decir, poder corroborar que efectivamente los cambios en la población destinataria se deben a la incidencia del programa y no debido a otro factor.

Sin embargo, a la hora de poner en práctica el procedimiento, el evaluador se enfrenta con algunos de los problemas característicos del campo de la educación que le dificultan la conformación de los grupos experimental y control:

a.- Las escuelas no se encuentran distribuidas homogéneamente en el territorio por lo tanto si se busca una escuela de características similares a una que se elige como escuela experimental, ya nos encontramos con la primera dificultad, las escuelas no pueden ser seleccionadas al azar.

b.- Si se selecciona la escuela más cercana, no necesariamente está integrada con la misma población, ni tiene el mismo tamaño, con lo cual no es posible desde el inicio pensar en una escuela "idéntica" a la experimental, aunque se puede inferir que una escuela cercana tenga al menos una población con características similares.

c.- Si los sujetos a evaluar son escuelas rurales, por ejemplo, el problema se agrava, porque las escuelas rurales se encuentran a bastante distancia una de otra, con lo cual lo único que se mantiene similar es la condición de ruralidad, pero no las otras variables de contexto que caracterizan la población.

d.- Si salvamos el tema de las distancias, de contexto, o tamaño y composición de la escuela y todavía mantenemos la idea de trabajar con grupo control, se le suma una nueva dificultad. Las escuelas públicas están atendidas por múltiples programas generados desde el estado y es raro que reciban el apoyo de un solo programa. Con lo cual el esquema de que un grupo de escuelas reciben el estímulo de un programa y otras no,

deja de ser exacto, pues todas reciben algún tipo de estímulo y en muchos casos múltiples.

e.- Si desestimamos esta contingencia, porque en resumidas cuentas si todas las escuelas reciben el estímulo y eso de alguna manera las homogeniza, nos encontramos con otra dificultad que tiene que ver con la información que se intenta recabar. Las escuelas poseen muchos registros: datos de los padres de los alumnos, de los alumnos, de la asistencia del alumno, de su desempeño escolar, etc., pero muchos de esos registros quedan en la escuela y luego no son utilizados. La mayoría de ellos se encuentran en formato papel. Es muy posible que la información que se precisa no esté disponible, pues la escuela la tiene pero nunca llegan al sistema de información que se encarga de recolectar las estadísticas escolares.

f.- Si al evaluador le interesa analizar el impacto positivo del programa en una mejora en los aprendizajes de los alumnos, puede recurrir a la información provista por los resultados de las pruebas ONE⁵. El Ministerio de Educación implementa periódicamente relevamientos nacionales cuyo objetivo es la evaluación de los aprendizajes de los alumnos y de los factores asociados. Estos relevamientos se realizan con el objeto de contar con información con el objeto de mejorar el desarrollo de políticas educativas. Sin embargo, a diferencia de lo que ocurre con las estadísticas educativas que se aplican a toda la población, estos datos no están disponibles para todas las escuelas (al menos en forma histórica), lo cual dificulta aún más, la tarea de constitución de los grupos control y experimental.

5. LIMITACIONES Y POTENCIALIDADES

A partir de la segunda mitad del siglo pasado se ha planteado una muy fuerte crítica al uso de modelos de evaluación a partir de la cual se

⁵ Estos relevamientos son llevados a cabo bianualmente por la Dirección Nacional de Información y Evaluación de la Calidad Educativa (DINIECE) del Ministerio de Educación de la Nación. Son pruebas que se aplican para conocer el grado de dominio que el alumno tiene sobre un recorte específico de contenidos y de capacidades cognitivas.

incorporan visiones más relativistas e inclusivas de los factores contextuales desde metodologías intersubjetivas. Por un lado, se ha dejado la concepción de la evaluación como una etapa final de la implementación de un proyecto social que conllevó a la utilización de modelos experimentales o cuasi experimentales puros que permitirían aislar los factores determinantes que produjeron los cambios esperados en la situación que motivó la intervención, y si esos cambios se debieron efectivamente a las acciones desplegadas (Mokate, 2003:3; Niremberg, Brawerman y Ruiz, 2003^a:40). Por el otro, el amplio desarrollo de la planificación estratégica y de las líneas participativas de investigación acción, dieron lugar a reconocer que las líneas de planificación no son lineales y que los actores involucrados en la planificación y la gestión reinterpretan sus acciones, por lo que se considera el amplio margen de incertidumbre de las intervenciones sociales.

A partir de ello, lo que proponemos es la incorporación de modelos experimentales y cuasi experimentales a partir de una lógica de triangulación que implique un abordaje multi-método, multi-actores y / o multi-técnicas que promueva el análisis de una línea de acción socioeducativa en la que se incorpore el experimento como uno de los ejes de la evaluación social, es decir ya no como "el diseño evaluativo " sino como una alternativa en función de la comprensión de los efectos de una acción dada, acompañada por la visión de los propios actores sobre el proceso en función del aprendizaje en el marco de la gerencia social.

La aplicación del método experimental resulta de esta manera como una instancia de enriquecimiento y profundización de los procesos de evaluación. La utilización de la técnica deberá observar una serie de recaudos para lograr una aplicación exitosa, entre ellos, un cuidadoso planeamiento de su implementación, la previsión de un tiempo adicional previo al proceso efectivo de la evaluación a los efectos de obtener un buen conocimiento del campo y la comprobación cierta de la información que se dispone y de la que no. Adicionalmente, si es necesario, prever alternativas de relevamiento de información ad hoc.

Como alternativas frente a situaciones donde se advierte la existencia de muchos obstáculos para la aleatorización, puede ser interesante, inclusive, la utilización de métodos experimentales con un N pequeño, con

muestras constituidas de manera intencional, pero más sólido en la constitución de las muestras en relación con las exigencias del experimento, que luego pueda combinarse con otros métodos descriptivos en más cantidad de sujetos.

En síntesis, frente a las severas críticas sobre la utilización de métodos experimentales, más que descartar su aplicación, resulta conveniente explorar las posibilidades valiosas que el método ofrece reivindicando su rol en la generación de conocimiento válido, confiable y oportuno para la gestión.

REFERENCIAS BIBLIOGRÁFICAS

Briones, G.(1999): "Evaluación Educativa" Formación de docentes en investigación educativa. Colombia, Convenio Andrés Bello.

Cohen E. y Franco R. (1997): *Evaluación de proyectos sociales*. Buenos Aires, Siglo XXI.

Dahler-Larsen, P. (2007): "¿Debemos Evaluarlo Todo? O De La Estimación De La Evaluabilidad A La Cultura De La Evaluación". *Evaluación De Políticas Públicas*. España, Mayo-Junio.. N.º 836 Ice. 93-104.

Feinstein, O. (2007): "Evaluación pragmática de políticas públicas" en *Revista ICE*. España, Mayo –junio 2007, nº 836.19-31.

Gómez Serra, M.(2004): *Evaluación de los servicios sociales*. España, Ed. Gedisa,.

Ministerio de Educación (2011): "Operativo Nacional de Evaluación 2010" en *Censo De Finalización.de la Educación Secundaria. Informe de Resultados*. Argentina.

Mokate, K. (2003): "Convirtiendo al monstruo en aliado: la evaluación como herramienta de la gerencia social" en *Documentos de trabajo del INDES*. Washington D.C.

Niremberg, O., Brawerman, J. y Ruiz, V. (2003a): *Evaluar para la transformación*. Buenos Aires, Paidós.

----- (2003b): *Programación y evaluación de proyectos sociales*. Buenos Aires, Paidós.