

Evaluación de programas: estudios de gran escala y de pequeña escala. Algunas generalizaciones

.....

Lorin W. Anderson
T. Neville Postlethwaite

5

serie Políticas educativas

A stylized illustration in shades of blue and white showing several children interacting with a globe. One child is standing and holding a vertical axis, another is sitting on the globe, and a third is reaching up. The globe is depicted with latitude and longitude lines.

Academia Internacional de la Educación



Instituto Internacional para la Planeación de la Educación

Serie •
políticas educativas • **5**

- Evaluación de programas:
• estudios de gran escala
• y de pequeña escala.
- Algunas generalizaciones
-

Lorin W. Anderson
T. Neville Postlethwaite

- Serie
- políticas educativas
- 5
- Evaluación de programas:
• estudios de gran escala
• y de pequeña escala.
- Algunas generalizaciones

Prefacio

Serie *Políticas educativas*

La Academia Internacional de la Educación y el Instituto Internacional para la Planeación de la Educación están publicando conjuntamente la serie de folletos *Políticas educativas*. El propósito de la serie es resumir lo que se sabe, con base en la investigación, de selectos temas de políticas en el campo de la educación.

La serie se diseñó para que pudieran consultarla rápidamente, “al vuelo”, los funcionarios de alto nivel que toman decisiones en los ministerios de Educación y que, por sus ocupaciones, rara vez tienen tiempo de leer informes de investigación extensos, asistir a conferencias y seminarios o enfrascarse en largos debates académicos con investigadores especializados en políticas educativas.

Los folletos a) se han centrado en temas de políticas que la Academia considera de alta prioridad en muchos ministerios de Educación, tanto en países desarrollados como en vías de desarrollo; b) están estructurados para lograr la mayor claridad posible, y contienen una introducción con una visión general del problema; una discusión, basada en investigaciones, de unas diez cuestiones clave que se consideran críticas para el tema del folleto, y referencias que proporcionan datos de apoyo y lecturas adicionales relacionadas con la discusión de los temas; c) se han restringido en cuanto a su extensión, ya que requieren entre 30 y 45 minutos de tiempo de lectura, y d) tienen el tamaño para poder llevarse en el bolsillo del saco, lo cual ofrece la oportunidad de una consulta fácil y accesible dentro o fuera de la oficina.

La Academia Internacional de la Educación escogió a los autores de la serie por sus conocimientos especializados sobre los temas de los folletos y también por su habilidad reconocida para comunicar resultados complejos de investigaciones en forma tal que puedan comprenderse y utilizarse fácilmente para elaborar políticas.

Los folletos aparecieron primero en inglés, y poco después en francés y español. Se está planeando traducirlos a otros idiomas.

Se publicarán cuatro folletos cada año y se podrán descargar gratuitamente del sitio de Internet del Instituto Internacional para la Planeación de la Educación. Poco después de la publicación electrónica se preparará una edición impresa limitada.

Incumbe cada vez más a los ministerios de Educación integrar la evaluación en los nuevos programas, especialmente aquellos en los que se gastan cantidades importantes de dinero. Por lo general, cada programa plantea preguntas que atañen al impacto y la efectividad del programa. Por ejemplo, es posible que los ministerios destinen más recursos a las aulas y a las escuelas. Pregunta: “¿Los recursos han llegado a los destinos deseados y han tenido algún efecto sobre el aprovechamiento de los alumnos?”. O tal vez se haya introducido un nuevo plan de estudios. Pregunta: “¿Cómo se introdujo el plan de estudios y qué problemas se presentaron cuando se implementó?”. O tal vez se haya desarrollado un nuevo programa para maestros en activo. Pregunta: “¿Los maestros aprendieron lo que debían aprender? Y de ser así, ¿tuvo esto algún efecto sobre lo que aprendieron los alumnos en términos de aprovechamiento, actitudes o comportamiento?”. Estos son algunos ejemplos para el caso de la educación escolar en general, pero lo mismo es válido para los nuevos programas en preescolar, en las escuelas para los minusválidos, en la educación vocacional, etcétera. Todos los programas educativos requieren que se les integre la evaluación, si ha de definirse el éxito del programa y si han de identificarse y corregirse los errores del mismo.

Cuando se introducen nuevos programas educativos no es fácil valorar si han tenido efecto en el aprendizaje de los alumnos. El tipo de diseño de investigación necesario para encontrar la verdadera ‘causa’ de los cambios en el aprendizaje de los alumnos variará de acuerdo con el tipo de aprendizaje que se especifica en las metas del programa. Al mismo tiempo, sin embargo, existen algunas similitudes en términos de diseños de evaluación sólidos y defendibles. Por ejemplo, siempre es importante que al principio del programa se haga una medición (o más de una) del nivel de aprendizaje de los alumnos. No se puede afirmar que los programas sean eficaces si no hay cambios mensurables en el aprendizaje de

los alumnos en el decurso. Igualmente, debe incluirse en el estudio algún grupo o grupos testigo de maestros y alumnos. Si hay cambios mensurables en el aprendizaje de los alumnos en el decurso, pero la magnitud de los cambios no difiere de los cambios que ocurren en los alumnos que no participaron en el programa, entonces no se puede afirmar que el programa haya sido efectivo.

Este folleto tiene que ver con los estudios de evaluación de programas en pequeña y gran escala. En muchos casos, se necesitan tanto los estudios de pequeña escala como los de gran escala en un proyecto de evaluación determinado. En este folleto, los autores sostienen que hacen falta ambas clases de estudios, siempre y cuando se llevan a cabo con apego a los estándares científicos.

Este folleto es el quinto en la serie *Políticas educativas* desarrollada por la Academia Internacional de la Educación y el Instituto Internacional para la Planeación Educativa. Cada folleto busca aportar las pruebas que arroja la investigación sobre temas importantes relacionados con la educación.

Lorin W. Anderson
y T. Neville Postlethwaite

Lorin W. Anderson es profesor emérito distinguido de la Universidad de Carolina del Sur, donde se desempeñó como profesor durante más de 30 años. Sus intereses en materia de investigación se han centrado en los campos de la enseñanza eficaz, el asesoramiento afectivo y el diseño de programas de estudio. Es editor senior de A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy (Una taxonomía para aprender, enseñar y evaluar: una revisión de la taxonomía de Bloom) (Allyn y Bacon, 2001) y es miembro de la Academia Internacional de la Educación.

T. Neville Postlethwaite ha sido muy activo en la evaluación de la educación durante los últimos 50 años y sobre todo en la evaluación de sistemas educativos. Desempeñó un papel central en el establecimiento de la Asociación Internacional para la Evaluación del Aprovechamiento Educativo (IEA, por sus siglas en inglés). Es profesor emérito de la Universidad de Hamburgo, Alemania, y sigue involucrado con una gama de proyectos de investigación transnacionales relacionados con el monitoreo de la calidad educativa. Es miembro de la Academia Internacional de la Educación.

Esta publicación ha sido producida por la Academia Internacional de la Educación (AIE) y el Instituto Internacional para la Planeación Educativa (IIEP). Puede reproducirse y traducirse a otros idiomas sin ninguna restricción. Favor de enviar una copia de cualquier publicación que reproduzca este texto completo o en parte a la AIE y al IIEP. Esta publicación está disponible en Internet en su forma impresa; diríjase a: <http://www.unesco.org/iiep>

Los autores son responsables de la elección y presentación de los datos que contiene esta publicación y de las opiniones que en ella se expresan, las cuales no son necesariamente las del IIEP (UNESCO) y no comprometen a la Organización. Los términos empleados y la presentación del material en esta publicación no implican la expresión de ninguna opinión del IIEP (UNESCO) con respecto al estatus legal de ningún país, territorio, ciudad o zona, ni de sus autoridades, ni con respecto a la delimitación de sus fronteras o límites territoriales.

*Publicado conjuntamente por:
El Instituto Internacional
para la Planeación de la Educación (IIEP)
7-9 rue Eugène Delacroix
75116 París
Francia*

*y
La Academia Internacional
de la Educación (AIE)
Palais des Académies
1, rue Ducale
1000 Bruselas
Bélgica*

1.	Introducción	1
2.	Diferencias entre los estudios de gran escala y de pequeña escala	4
3.	La utilización de los estudios de pequeña escala para diseñar herramientas de recopilación de datos y procedimientos para el manejo de datos en los estudios de gran escala	7
4.	Los estudios de pequeña escala son parte de todo un procedimiento de evaluación.	9
5.	Diseño de muestra y buenas herramientas para estudios de gran escala y de pequeña escala	12
6.	Presentación de estudios de evaluación de pequeña escala y de gran escala	14
7.	Interpretación de resultados	16
8.	Cómo redactar para distintos públicos	19
9.	Riesgos inherentes.	22
	Conclusión	25
	Referencias.	27

- Serie
- políticas educativas
- 5
- Evaluación de programas:
- estudios de gran escala
- y de pequeña escala.
- Algunas generalizaciones

Introducción

Los ministerios de educación tienen que evaluar los nuevos programas que introducen. Si no lo hacen, no tienen idea de si el programa se ha implementado tal como se diseñó y, si así fue, qué efecto ha tenido sobre el aprovechamiento, las actitudes y/o los comportamientos de los alumnos. La evaluación se vuelve especialmente importante cuando se invierten grandes sumas de dinero en un programa. Por lo general, esto significa integrar la evaluación desde el principio del estudio con el fin de ver si el programa ha tenido el efecto deseado o efecto alguno.

En la evaluación educativa se utiliza mucha jerga, por ejemplo “evaluación formativa y sumativa”, “diseño experimental”, “investigación de encuesta”, “estudios cuantitativos y cualitativos” y otros numerosos términos técnicos. Esta jerga puede ser muy confusa para el lector lego. En este folleto se han examinado los conceptos y la terminología de los estudios de evaluación de gran escala y de pequeña escala dentro del marco de la evaluación de programas. Normalmente, los estudios de pequeña escala abarcan hasta 20 o 30 escuelas, mientras que los estudios de gran escala pueden abarcar desde 100 hasta 4,000 escuelas y hasta 100,000 alumnos. Sin embargo, no se ha hecho ninguna distinción entre estudios cuantitativos y cualitativos porque éste será tema de otro folleto.

La evaluación programática se puede definir como la recopilación sistemática de información sobre las características, actividades y resultados de programas con el propósito de formarse juicios sobre su eficacia, y sobre cómo mejorar su eficacia y/o aportar datos para el desarrollo de programas futuros (Centros para el Control y la Prevención de Enfermedades, 1999). Como sugiere esta definición, la evaluación de un programa puede centrarse en sus características, sus actividades, sus resultados o una combinación de éstas.

Las “características” incluyen el contexto en que funciona el programa, los recursos utilizados para apoyar el programa y el personal que lo implementa. Las “actividades” incluyen aquellas acciones que deben tomarse para implementar el programa adecuadamente. Las estrategias de instrucción, las prácticas docentes, los comportamientos de los maestros y de los alumnos, las interacciones entre maestros y alumnos y el trabajo de éstos son sólo algunas de las actividades asociadas con los programas escolares. Los “resultados” se refieren al impacto del programa sobre aquellas personas a quienes se destina. Estos resultados se pueden especificar por adelantado (por ejemplo, como metas del programa) o pueden ser resultados no intencionales del diseño o la implementación del programa (que a menudo se conocen como “consecuencias”).

Además de las diferencias con respecto a las características, actividades y resultados, las evaluaciones de programas pueden diferir por sus propósitos. La evaluación puede centrarse en la efectividad del programa. A esto se le denomina típicamente como “evaluación sumativa” (Scriven, 1967). La evaluación también se puede centrar en el mejoramiento del programa. A esto se le ha llamado tradicionalmente “evaluación formativa”. Por último, la evaluación puede centrarse en la orientación de decisiones sobre el desarrollo futuro de programas. Esto implica típicamente tanto la evaluación formativa como la sumativa y le da más importancia a quien toma las decisiones que a quien hace la evaluación. Para este propósito, es absolutamente indispensable una comunicación clara y significativa entre quien evalúa y quien toma las decisiones.

Por último, aunque se les denomina en general “evaluación de programa”, los conceptos, principios y procedimientos de evaluación descritos en este folleto también conciernen a la evaluación de sistemas educativos. En muchos sentidos, se puede ver un sistema educativo como un conjunto de programas con metas complementarias que se organizan en varias unidades administrativas y son consistentes con (o facilitan) la misión general de un sistema educativo.

El propósito de este folleto consiste en describir el estado actual de los conocimientos respecto al campo de la evaluación de pro-

gramas. A lo largo del folleto se ha puesto el énfasis en conceptos y principios básicos y sobre las mejores prácticas que en lo general han sido aceptadas. Este folleto está organizado en torno a siete principios de una evaluación de programas sólida y defendible. Se incluye una bibliografía de publicaciones relevantes al final del folleto.

Diferencias entre estudios de gran escala y de pequeña escala

-
- *Los estudios de evaluación de gran escala y de pequeña escala difieren en sus propósitos y alcances, así como en el grado en que puedan generalizarse sus resultados.*
-
-
-
-
-
-

Cinco características fundamentales distinguen los estudios de evaluación de gran escala de los de pequeña escala. Si bien la diferencia más evidente es el tamaño, los estudios también se distinguen por su propósito, su costo, grado de generalización de los resultados y tipo y complejidad del análisis de datos.

- Tamaño de la muestra:* Los estudios de pequeña escala por lo general se hacen con muestras pequeñas. Pequeño normalmente significa no más de 30 escuelas o 60 grupos. Puesto que esto no es suficiente para cubrir las variaciones en todo un grado escolar en un país, los estudios de pequeña escala a menudo cubren sólo una región o un distrito y no todo el país. Los estudios de gran escala por lo general abarcan más de 100 escuelas y unos 20 a 30 alumnos en cada escuela. La muestra de escuelas proviene de todo el sistema educativo.
- Propósito del estudio:* Los estudios de pequeña escala tienden a utilizarse con propósitos exploratorios y se centran en el desarrollo de medidas, planes de estudio y libros de texto, nuevos métodos de enseñanza, etcétera. Los estudios de gran escala sirven para describir el sistema como un todo y las partes que lo integran (por ejemplo, regiones dentro del sistema). Las descripciones por lo general tienen que ver con logros en distintas materias, actitudes y comportamientos de alumnos y maestros y, frecuentemente, con las relaciones entre muchos de los factores que operan en el sistema. Los estudios de pequeña escala

suelen utilizarse en la evaluación formativa (véase más arriba), mientras que los de gran escala se emplean usualmente con propósitos de evaluación sumativa (véase más arriba).

- c) *Costo*: Los estudios de pequeña escala cuestan mucho menos que los de gran escala. Es muy costoso emprender la recopilación de datos en, digamos, 200 escuelas en todo el país, no sólo en términos de la recopilación misma de los datos, sino también por lo que respecta a su manejo (captura, depuración y análisis de datos). Por el mismo costo, un estudio de pequeña escala podría incluir muchas observaciones y variables, y así generar una descripción más detallada de un número limitado de escuelas.
- d) *Grado de generalización de resultados*: Si un estudio tiene como objetivo un solo grado escolar y una sola materia en un sistema educativo, y toma muestras de escuelas y/o grupos en ese grado escolar y esa materia dentro del sistema, los resultados pueden generalizarse para ese nivel y esa materia en todo el sistema con ciertas limitaciones (los errores estándar de muestreo). Para conocer estas limitaciones, debe informarse de los errores de muestreo para todas las estimaciones que se incluyan en el estudio. Cuando se emplean estudios de pequeña escala, es difícil, si no imposible, hacer tales generalizaciones debido a que los errores de muestreo son muy grandes.
- e) *Tipo y complejidad del análisis de datos*: En los estudios de gran escala con frecuencia se requieren análisis estadísticos más complejos. Por ejemplo, con grandes números de escuelas y maestros se puede examinar el efecto simultáneo de muchas variables sobre, digamos, el aprovechamiento. Este examen es posible incluso cuando las variables representan distintas unidades de análisis (por ejemplo, alumno, grupo y escuela). Estos tipos de análisis son importantes para identificar las interacciones de las variables y su efecto en el aprovechamiento. En los estudios de pequeña escala no es posible utilizar estos tipos de análisis por la cantidad tan pequeña de escuelas o alumnos. Los estudios de pequeña escala que incluyen cantidades adicionales de datos cualitativos a veces pueden volverse muy caros porque la recopilación de estos datos requiere de mucho trabajo.

Ambas clases de estudios tienen su lugar en la evaluación de programas, a veces de manera independiente y otras combinadas. La decisión respecto a cuál de ellas utilizar debe basarse primordialmente en las cinco características que se detallaron anteriormente.

La utilización de los estudios de pequeña escala para diseñar herramientas de recopilación de datos y procedimientos para el manejo de datos en los estudios de gran escala

-
- *Los estudios de gran escala y los de pequeña escala constituyen enfoques complementarios de evaluación más que enfoques diametralmente opuestos.*
-
-
-
-

Algunos evaluadores consideran incompatibles los estudios de gran escala y los de pequeña escala. De hecho, los estudios de pequeña escala son parte importante de un proceso de evaluación general porque pueden utilizarse para identificar lo que debe medirse, definir cómo medir lo que debe medirse, establecer los procedimientos para la recopilación y captura de datos y hacer pruebas preliminares de los factores ‘compuestos’ que se construyen.

- Identificación de lo que debe medirse:* Un estudio de pequeña escala puede usarse para observar grupos en varias escuelas distintas, para encontrar factores que marquen una “diferencia” entre escuelas/grupos/alumnos que logran puntuaciones altas en las pruebas y escuelas/grupos/alumnos que obtienen puntuaciones bajas. Esta lista de factores puede emplearse luego para generar hipótesis que puedan probarse en un estudio de gran escala.
- Definición de cómo medir:* Es necesario medir los factores importantes que se identifican en un estudio de pequeña escala. Los estudios de pequeña escala permiten hacer una exploración de bajo costo de enfoques alternativos para realizar la medición. Esto ofrece la oportunidad de probar la validez y confiabilidad de cada medición y decidir la mejor forma de medir (por ejemplo, encuestas, entrevistas, observaciones).
- Instrumentos de ensayo y procedimientos de recopilación de datos:* Es necesario ensayar en pequeña escala todos los instrumentos que se utilizarán en estudios de gran escala y luego modificarlos con base en el análisis de los datos de ensayo. Esto

permite cambiar los procedimientos de recopilación de datos para asegurar una mejor recopilación en el estudio principal. Además, pueden ensayarse los libros de códigos que se requieren para la captura de datos antes de que se utilicen en el estudio principal.

- d) *Construcción de factores compuestos*: En los estudios de gran escala a menudo se emplean como variables independientes clave factores compuestos que se construyen a partir de ciertas variables. Por ejemplo, el nivel de instrucción y el estatus laboral de los padres se combinan con frecuencia en un nuevo factor compuesto denominado “estatus socioeconómico”. Los estudios de pequeña escala permiten ensayar cabalmente todos los pasos en el proceso de construcción antes del estudio principal.

Los estudios de pequeña escala son parte de todo un procedimiento de evaluación

-
- *Un ejemplo del uso de estudios de pequeña escala para desarrollar y probar nuevos materiales de planes de estudio.*
-
-
-
-
-

Existen seis pasos principales en el desarrollo de materiales para planes de estudio.

- Establecimiento de metas y objetivos educativos:* Por lo general, los gobiernos establecen las metas de un sistema educativo. Sin embargo, estas metas a menudo son abstractas, y para orientar el desarrollo de materiales de planes de estudio se requiere de consultas con grupos interesados para trazar metas y objetivos más específicos (susceptibles de volverse operacionales). Por ejemplo, para la clasificación de metas y objetivos tal vez se empiece por consultar a pequeñas muestras de empleadores, académicos, maestros y el público en general. Estas consultas proporcionarán información con respecto a conocimientos, habilidades y actitudes valoradas. Además, podrán hacerse estudios en pequeña escala para identificar cómo han cambiado las situaciones y expectativas de distintas categorías de alumnos (basadas en género, antecedentes socioeconómicos, ubicación geográfica, etcétera) y las implicaciones que tendrán estos cambios para las modificaciones en los planes de estudio.
- Redacción de las unidades o módulos:* Este paso no requiere de un estudio de evaluación, si bien el establecimiento de los parámetros acordados (como por ejemplo la estructura, el contenido y el formato) tal vez exija un estudio de pequeña escala para determinar el grado de acuerdo.
- Ensayo de materiales en pequeña escala:* En este paso, por lo general se efectúa un estudio de pequeña escala en seis a ocho escuelas cerca del centro de diseño del plan de estudios de

manera que haya mucha interacción entre los desarrolladores del plan de estudios y los maestros en las escuelas “de ensayo”. Esta cobertura pequeña de escuelas permite la recolección de información “a fondo”. Por ejemplo, se desarrollan pruebas para cada módulo de unidad con el afán de identificar los objetivos que no se están alcanzando plenamente y determinar las razones de ese mal desempeño. A los maestros se les pregunta sobre los problemas que encuentran en cada parte de los materiales escritos y esta información se relaciona con los resultados reales de las pruebas.

- d) *Evaluación a gran escala de los nuevos materiales*: En este paso se lleva a cabo un estudio a gran escala en el que se elige una muestra de unas 60 escuelas que representan la gama total de escuelas a las cuales estará destinado el plan de estudios. Los maestros participantes reciben capacitación para que comprendan y puedan implementar el nuevo plan de estudios adecuada o “fielmente”. En general, los desarrolladores del plan de estudio necesitan saber si un objetivo se está alcanzando bien (alrededor de 80 por ciento de los alumnos contestan correctamente los temas relacionados), moderadamente (alrededor de 50 por ciento contesta correctamente los temas relacionados), o mal (menos de 30 por ciento contesta correctamente los temas relacionados). Se identifican las partes débiles del plan de estudios y se hacen las modificaciones necesarias. Además, se terminan y se documentan los materiales y procedimientos de capacitación de los maestros.
- e) *Implementación a gran escala (y a veces evaluación de los materiales de capacitación de los maestros)*: En este punto se capacita a los maestros y empieza la enseñanza de los nuevos materiales en las escuelas. En ocasiones los evaluadores integran un mecanismo de revisión para conocer la adecuación y eficacia de los materiales y procedimientos de capacitación de maestros en una muestra grande de los mismos. Estos materiales deben ayudar a los maestros a comprender el plan de estudios y adquirir las habilidades necesarias para implementarlo eficazmente tal como fue diseñado. En este momento debe evaluarse la eficacia

de la capacitación en términos de los cambios logrados en los conocimientos y habilidades de los maestros.

- f) *Estudio de evaluación 'global' de gran escala*: El último paso en el proceso se lleva a cabo dos o tres años después de que se haya implementado el plan de estudios en las escuelas. En este momento se realiza una evaluación de gran escala. Este estudio consiste en tomar una muestra de probabilidad de escuelas (véase más abajo) tal que no sean demasiado grandes los errores para estimar el aprovechamiento de los alumnos. Además, se identifican grupos testigo de maestros y alumnos y se hacen mediciones de valores de referencia de alumnos “dentro de programa” y “fuera de programa”.

Diseño de muestra y buenas herramientas para estudios de gran escala y de pequeña escala

-
- *Para diseñar estudios de evaluación de gran escala y de pequeña escala sólidos y defendibles es necesario poner atención a diversos asuntos técnicos.*
-
-
-
-
-

Los estudios de gran escala por lo general se llevan a cabo con muestras de escuelas y alumnos y no con todas las escuelas y todos los alumnos en una población objetivo. Los estudios de pequeña escala por lo general se basan en muestras de valor.

- a) *Muestras de probabilidad para estudios de gran escala:* Los estudios de gran escala deben contar con muestras de probabilidad tales que cada persona en la población tenga una probabilidad diferente de cero de formar parte de la muestra. Típicamente, los evaluadores educativos incluyen muestras de dos etapas, en las cuales las escuelas son las primeras en ser seleccionadas con probabilidad proporcional al tamaño de la población definida dentro de cada escuela. Luego se seleccionan los alumnos al azar dentro de cada escuela elegida. En ocasiones se utilizan muestras de tres etapas, en las que las regiones o distritos son la primera etapa y en consecuencia los primeros en ser seleccionados. Puesto que los marcos de muestreo y las tasas de respuesta rara vez son perfectas, deben calcularse ponderaciones de muestreo por la variación de probabilidades y la falta de respuestas. Por último, debido a que hay muestras de escuelas así como de alumnos, habrá cierto error de muestreo. Estos errores deben calcularse para cada estimado que se dé en el informe de evaluación.
- b) *Muestras de valor para estudios de pequeña escala:* En los estudios de pequeña escala por lo general se eligen muestras de valor. Las escuelas muestra normalmente se seleccionan de manera que proporcionen una buena cobertura de la variedad de

escuelas de un sistema escolar. Antes de sacar las muestras es importante que los evaluadores obtengan información relevante sobre las escuelas que tal vez se incluyan en el estudio. Si, por ejemplo, el propósito del estudio es ver las diferencias en el aprovechamiento en matemáticas, entonces se eligen las escuelas de manera que abarquen las diferencias conocidas en el aprovechamiento en matemáticas. Si el estudio tiene la intención de examinar las diferencias en los problemas de conducta en las escuelas, entonces la muestra debe basarse, por ejemplo, en la información de los inspectores respecto a la gravedad de los problemas de conducta. Con frecuencia se toman muestras “rectangulares” para poder probar el desempeño de los alumnos en los extremos de la distribución.

- c) Las pruebas, cuestionarios y cualesquiera otras herramientas utilizadas en todos los estudios de evaluación deben dar como resultado mediciones con un alto grado de confiabilidad y validez. Cuando se usa la observación, el programa de observación debe prepararse de tal forma que cualquier observador capacitado pueda utilizarlo para producir registros o codificaciones precisos basados en las observaciones. Además, debe existir un alto grado de confiabilidad entre observadores (es decir, consenso entre ellos). Cuando se espera que el observador describa lo observado, en lugar de hacer inferencias sobre ello, la confiabilidad entre observadores debe ser sumamente alta (no menos de 0.90). Cuando se requiere de la inferencia por parte de los observadores, la confiabilidad entre observadores por lo general es menor, pero no debe ser menos de 0.80.

Presentación de estudios de evaluación de pequeña escala y de gran escala

-
- *La calidad del informe debe apearse a cierto estándar, independientemente del tipo de estudio que se lleve a cabo.*
-
-
-
-

Una de las cuestiones más importantes en la elaboración de un estudio de evaluación es el informe de la investigación que se ha llevado a cabo. Es importante que este informe contenga suficiente detalle, de manera que otros evaluadores puedan duplicar el estudio. El informe sobre el estudio de evaluación debe tratar los siguientes temas:

- Propósitos del estudio:* Los propósitos del estudio y las preguntas de investigación relacionadas con él deben quedar muy claros y su relevancia debe ser evidente. Debido a que el énfasis en muchos estudios de evaluación de programas se pone en la efectividad del programa, esto significa diseñar un estudio que permita al evaluador atribuir los cambios en el aprendizaje de los alumnos (aprovechamiento, actitudes y/o conductas) al programa en sí, y no a toda una serie de factores ajenos (por ejemplo, diferencias en el aprendizaje anterior de los alumnos, diferencias en la calidad de los maestros, diferencias en la estructura organizacional de la escuela, etcétera).
- Cuestionarios:* Las preguntas que se incluyan en el cuestionario deben ensayarse plenamente. En caso de que se construyan escalas a partir de grupos de preguntas, deben existir pruebas claras de que la confiabilidad de la escala es alta y de que la escala presenta una valoración válida de la hipótesis que se está considerando.
- Pruebas:* Hay que definir claramente los objetivos cognitivos específicos antes de empezar a redactar las preguntas. Las preguntas deben redactarse, revisarse y luego ensayarse, y deben desecharse las malas. La estructura final de la prueba (en su totalidad o en sus partes) debe concordar con las preguntas de investigación.

- d) *Muestreo*: La población objetivo del estudio debe describirse con precisión. Si se requiere de una muestra de valor, debe quedar claro si se trata de una distribución rectangular. Si se requiere de una muestra de probabilidad, debe quedar clara la tolerancia de error (típicamente, no debe ser mayor de 5 por ciento en el caso de un porcentaje). Además, es necesario proporcionar información sobre la forma en que se obtuvo la muestra y cómo se calcularon las ponderaciones de la muestra (si es que hubo). Deben establecerse las reglas con que se decidirá la exclusión del estudio de alumnos que pertenecen a la población definida, y no debe permitirse la exclusión de la población deseada de más de 5 por ciento de los alumnos pertenecientes a la población definida.
- e) *Realización del estudio*: Deben hacerse explícitos los pasos para producir los instrumentos de recopilación de datos, contactar las escuelas, seguirle la pista a las escuelas, los maestros y los alumnos, seleccionar y capacitar a los recolectores de datos, administrar los instrumentos y devolverlos a un lugar central.
- f) *Captura y depuración de datos*: Deben quedar claros la preparación del programa de captura de datos, la capacitación de quienes van a capturar los datos, los pasos involucrados en la captura y depuración de los datos y las reglas utilizadas para tomar decisiones respecto a los datos.
- g) *Análisis de los datos*: Los distintos tipos de análisis de datos que se llevarán a cabo deben ser claros y apropiados para responder a las preguntas de investigación. Además, los tipos de análisis de datos deben especificarse en una etapa temprana del proceso de planeación. Se pueden practicar análisis de datos adicionales según los justifique la situación.
- h) *Diseminación de resultados*: Es muy útil producir distintos informes. Uno es el mismo informe de investigación, que lo explica todo. Otro es un informe breve para los elaboradores de políticas de mayor jerarquía. El tercero, para el público en general, no es fácil de redactar porque tiene que escribirse con sencillez, pero sin distorsionar los resultados. Los resultados de algunos de los estudios de evaluación de mayor éxito se han difundido mediante la televisión y, cada vez más, por Internet.

Interpretación de resultados

-
- *Los resultados de los estudios no hablan por sí mismos; deben ser interpretados para beneficio de los elaboradores de políticas y otras personas interesadas.*
-
-
-
-
-
-

La interpretación es el proceso de darle sentido a los datos que se recopilaron, analizaron y reportaron. Con demasiada frecuencia los datos se reúnen y analizan, pero el siguiente paso (poner los resultados dentro de un contexto y darles un sentido antes de difundirlos) no se da. Por ejemplo, ¿qué significa que el 45 por ciento de los adolescentes que respondieron a un cuestionario creen que beber alcohol y consumir drogas es dañino para su salud? ¿Eso es bueno? El porcentaje, ¿está aumentando o disminuyendo? ¿Cómo comparar esto con otras escuelas, agencias de educación locales o países? ¿Qué significa en términos de salud y seguridad? ¿Qué significa en términos de la eficacia de un programa para prevenir el consumo de drogas y alcohol? Cuando se atienden las inquietudes respecto a la interpretación, deben seguirse algunos lineamientos generales.

- a) *Paradigmas*: La interpretación siempre requiere que se examinen los datos a través de una lente, un marco de referencia, algún modelo o paradigma. Los datos que se recopilan durante las observaciones que se hacen en las aulas pueden interpretarse desde la perspectiva del marco de referencia del modelo proceso-producto (en el que se cree que las conductas del maestro influyen directamente en el aprovechamiento del alumno) o del modelo de mediación del proceso (en el que se cree que las percepciones y las conductas de los alumnos median o intervienen entre las conductas del maestro y el aprovechamiento del alumno). Si bien los datos son los mismos, las interpretaciones

de esos datos serán distintas, de acuerdo con el modelo (lente) a través del cual se ven o se examinan los datos.

- b) *Uso de comparaciones*: La interpretación casi siempre implica algún tipo de comparación. Los datos de referencia, los grupos testigo o de comparación, los estándares de desempeño esperado predefinidos y los estándares que indican un significado estadístico son todos ejemplos de comparaciones potencialmente relevantes. El uso de valores de referencia significa que los datos se recopilan antes de adoptar e implementar un programa nuevo o modificado. Los datos que se obtienen posteriormente a la implementación se comparan con los valores de referencia para determinar el cambio en el decurso. Los grupos testigo o de comparación son grupos de alumnos que no participan en el programa que se está evaluando, pero que son similares en muchos aspectos a los que participan en el programa (por ejemplo, en género, raza y aprovechamiento anterior). La comparación, digamos, de las puntuaciones que obtienen en las pruebas los alumnos que participan en el programa con las puntuaciones obtenidas por los que no participan permite llegar a una interpretación del impacto que tiene el programa en el aprovechamiento de los alumnos. Los estándares de desempeño esperado predefinidos son niveles de desempeño que indicarán el éxito o eficacia del programa. Por ejemplo, se podrá afirmar que un programa vocacional técnico ha sido “exitoso” si 80 por ciento de los alumnos están trabajando en un empleo relacionado con el programa dentro de los seis meses posteriores a la graduación. Por último, los estándares de significado estadístico se utilizan conjuntamente con los valores de referencia y los grupos testigo o de comparación. El significado estadístico indica si la diferencia entre los valores de referencia y los datos posteriores a la implementación o entre los alumnos del programa y los de comparación es lo suficientemente grande que no pueda atribuirse a acontecimientos aleatorios o “al azar”.
- c) *Relevancia estadística y relevancia sustantiva*: Cuando se hacen evaluaciones cuantitativas es importante recordar que relevancia estadística no siempre significa relevancia sustantiva. Ganar

la batalla con respecto a la relevancia estadística no es lo mismo que ganar la guerra de políticas públicas o de relevancia práctica. Una reducción de uno o dos por ciento en la tasa de analfabetismo no resultará tan persuasivo para los elaboradores de políticas como un cambio de 10 por ciento, independientemente del nivel de relevancia estadística asociado con los resultados.

- d) *Fuentes múltiples de datos:* La interpretación mejora cuando existen y se toman en cuenta fuentes múltiples de datos. Si, por ejemplo, un programa se centra en disminuir el número de alumnos que desertan antes de terminar sus estudios, los datos adicionales a las tasas de deserción mejorarán nuestra comprensión respecto a la eficacia del programa. Ejemplos de estos datos tal vez incluyan el desempeño académico del alumno, su sentido de pertenencia, su nivel de participación en la vida escolar y sus patrones de amistad dentro y fuera de la escuela.
- e) *Grado de generalización:* Deben discutirse los alcances y límites del grado de generalización de los resultados de evaluación. Una evaluación rigurosa, libre de grandes errores y que revela resultados importantes, puede ayudar a demostrar que un programa ha funcionado en un contexto particular. Desafortunadamente, ninguna evaluación por sí sola puede demostrar que el programa funcionará igualmente bien en otro contexto.
- f) *Perspectivas distintas:* Como la interpretación depende de la lente o marco de referencia de la persona que hace la interpretación, por lo general redundante en una mayor comprensión cuando varias personas examinan y comentan los datos. Realmente se consigue un resultado muy valioso cuando personas que trabajan desde distintas perspectivas llegan a la misma interpretación.

Cómo redactar para distintos públicos

-
- *En la elaboración de los informes de evaluación, hay que poner atención al público al que va destinado el informe.*
-
-
-
-

Si bien los informes de evaluación los redactan evaluadores, no son éstos típicamente los destinatarios. En general, los informes de evaluación van dirigidos a tres públicos: educadores, políticos y el público en general (incluyendo los integrantes de los medios noticiosos). Con pocas excepciones, los integrantes de estos públicos no poseen una comprensión de las cuestiones técnicas ni están familiarizados con la terminología que emplean los que realizan y difunden las evaluaciones. Cuando se escribe para estos públicos, deben seguirse estos lineamientos.

- Públicos destinatarios:*** El informe debe redactarse para un público destinatario de lectores comunes que tal vez no estén familiarizados con los problemas y los temas materia de la investigación. Para aumentar la probabilidad de alcanzar esta meta, hay que incluir a “lectores comunes” en el grupo de personas entre quienes circulan los borradores del informe. Además, los resultados deben presentarse en un lenguaje accesible, directo y conciso. Evítase el uso de terminología técnica hasta donde sea posible, y si es necesario usar esta terminología, tómesese el tiempo para definir los conceptos.
- Contextos:*** Hay que asegurarse de incluir información sobre el contexto dentro del cual se llevó a cabo la evaluación. La información contextual es necesaria para ayudar a los públicos a interpretar la evaluación. El contexto deberá describirse con suficiente detalle para que los integrantes de los diversos públicos puedan determinar el probable impacto del contexto sobre la implementación y la eficacia del proyecto.

- c) *Objetivos y propósitos*: Hay que vincular los resultados con el propósito y los objetivos para los cuales se realizó la evaluación. Con frecuencia el formato de preguntas y respuestas es útil en este sentido; las preguntas se derivan del propósito y de los objetivos planteados y las respuestas de los datos que se recolectaron, analizaron e interpretaron.
- d) *Equilibrio entre formatos informativos*: Debe evitarse la dependencia total de la narrativa o de tablas y cuadros. Hay que buscar el equilibrio entre ambos formatos. Cuando se incluyen tablas y/o cuadros, conviene utilizar la narrativa para “llevar de la mano” al lector en la revisión de los datos que se incluyan o se resuman en las tablas y cuadros para mejorar su comprensión.
- e) *Honradez*: La mayoría de las evaluaciones producen una mezcla de resultados positivos y negativos que deben presentarse en forma equilibrada. Cuando los resultados son contradictorios o inconsistentes con los de evaluaciones similares, hay que dar explicaciones razonables de tales contradicciones e inconsistencias.
- f) *Vinculación de resultados y recomendaciones*: Hay que resaltar el vínculo entre las recomendaciones propuestas y los resultados de la evaluación. Siempre que sea posible, antes de emitir el informe final conviene revisar las recomendaciones propuestas con aquellos que serán responsables de implementarlas.
- g) *Resumen ejecutivo*: Inclúyase siempre un resumen ejecutivo. Los políticos, sobre todo, a menudo no tienen el tiempo para leer el informe completo. El resumen ejecutivo debe incluir, como mínimo, los propósitos del estudio y las preguntas de investigación relacionadas con ellos, los métodos utilizados para realizar el estudio, los principales resultados del estudio y las recomendaciones que se derivan de los resultados para la elaboración de políticas, para la práctica y para la investigación.
- h) *Conflictos en la interpretación*: Si elabora el informe un equipo de evaluadores, hay que aclarar cualesquier diferencias que surjan entre ellos en cuanto a la interpretación o el punto de vista del informe, y esas diferencias o bien deben resolverse o incluirse en el informe como comentarios en el texto o en una sección

especial. Sin embargo, sólo una persona debe ser responsable de la calidad y exhaustividad del informe escrito final.

- i) *Pulimento y revisión:* Asegúrese de revisar y de “pulir” el borrador final del informe antes de difundirlo. Es común que se requieran varios lectores para pescar todos los errores y mejorar la redacción en general del informe. Aunque es tediosa, la corrección de pruebas contribuye en mucho a dar la impresión de que se trata de un estudio y de un informe de calidad. Muchos de los que desean criticar el estudio, el informe y/o las recomendaciones a menudo buscarán cualquier razón, sin importar lo trivial que sea, para justificar sus críticas.

-
- *Existen riesgos inherentes a todos los estudios de evaluación.*
-
-
-
-
-

Los estudios de evaluación con frecuencia cuestionan los procedimientos y las prácticas establecidos; por lo tanto, es posible que se vean negativamente o con gran suspicacia. Esto puede dar pie a ciertos riesgos que quizá impidan que se utilice el informe de evaluación para los fines que lo motivaron. A continuación se enumeran algunos de los riesgos inherentes que se presentan con mayor frecuencia en los estudios de evaluación.

- Supresión de resultados:* Puede suceder que algunos burócratas, que están actuando en nombre de los gobiernos, decidan qué aspectos del estudio deben o no deben hacerse públicos. Se ha sabido de casos en que una sola persona lee y tacha con tinta roja aquellas partes del borrador que no deben incluirse en el informe final. La razón de esta práctica es que la persona teme que un resultado en particular podría incomodar al gobierno. También puede suceder que un gobierno impida la publicación de los resultados de un estudio de evaluación y dé una serie de justificaciones de su actitud. Entre las razones que dan están las siguientes: (i) el estudio deja sin responder algunas dudas importantes; (ii) muchas de las pruebas no son concluyentes; (iii) las cifras se prestan a múltiples interpretaciones; (iv) ciertas hallazgos son contradictorios, y (v) se han puesto en tela de juicio algunas de las conclusiones y recomendaciones principales (véase Lynn y Jay, 1985). Otra forma más de supresión consiste en decir que es necesario duplicar los resultados antes de que se puedan publicar los resultados del estudio actual.
- Jerarquización (ranking) de países y regiones o escuelas:* Cuando se han incluido todas las escuelas de un país en los estudios

de evaluación, a veces se acostumbra clasificar las escuelas o regiones en orden de resultados, y estos resultados se hacen del conocimiento público. Esto se basa en el fundamento de que “los padres de familia desean conocer los resultados”. Es fácil para los maestros de una escuela en una zona socioeconómica alta lograr que los alumnos aprendan mucho y se desempeñen bien. Es mucho más difícil para los maestros de escuelas en zonas socioeconómicas muy pobres lograr que sus alumnos alcancen este mismo nivel de desempeño. En consecuencia, algunos países producen clasificaciones sólo después de tomar en cuenta (o “controlar”) las diferencias socioeconómicas. Tiene poca utilidad ordenar jerárquicamente a las escuelas en estudios donde sólo se tomó en cuenta una muestra de escuelas. Sin embargo, cuando en un estudio se han incluido todas las regiones y si las estimaciones, digamos, de aprovechamiento son válidas y precisas, es razonable utilizar los datos para clasificar las regiones. Desde luego, la conveniencia de hacerlo se convierte en una cuestión política.

El lugar que le corresponde a un país en un estudio internacional de aprovechamiento es meramente resultado de qué otros países se incluyeron en el estudio. Mucho más importante que simplemente clasificar los países es comprender las razones de las diferencias entre ellos. Con esta información se hace posible mejorar los sistemas educativos de los países que tienen un bajo nivel de aprovechamiento.

- c) *Basar generalizaciones sobre unas cuantas escuelas:* Es muy posible que se dé el caso en que se lleve a cabo un estudio muy bueno sobre, digamos, 12 escuelas. Sin embargo, no es posible que 12 escuelas representen una buena muestra de todas las escuelas, salvo en un país pequeño en el que todas las escuelas son muy similares en términos de varios factores demográficos (por ejemplo, tamaño, estatus socioeconómico de los alumnos, etcétera). Debido a que éste no es por lo general el caso, hay que tener mucho cuidado de NO generalizar los resultados de un estudio que contenga muy pocas escuelas con respecto a toda una región o todo un país.

- d) *Estudios de evaluación que incluyan pruebas de alto impacto (high stakes testing)*: Las pruebas de alto impacto son aquellas cuyos resultados se utilizan para tomar decisiones que cambian la vida. Las decisiones pueden referirse a una escuela, a una maestra o incluso a un alumno. Por ejemplo, en algunos sistemas escolares es posible que se clausure un plantel o que se reemplace a todo su personal si tiene muy pobres resultados de aprovechamiento durante tres o cuatro años. De manera semejante, tal vez se determine que hay que ponerle atención especial a una maestra que ha tenido invariablemente muy pobres resultados con sus alumnos durante un periodo similar. Cuando se realizan pruebas de alto impacto (high stakes) (i), todos los alumnos de todas las escuelas en un grado escolar determinado deben incluirse en el estudio y (ii) los resultados del estudio deben ser consistentes a lo largo de un periodo razonable (por ejemplo, de tres a cinco años).
- e) *Confusión entre correlación y causalidad*: El propósito de la mayoría de las evaluaciones de programas es vincular la implementación del programa con alguna medición de aprendizaje estudiantil. Sin embargo, en la mayoría de las circunstancias existen factores independientes del programa que podrían influir en el aprendizaje de los alumnos. Si la implementación del programa es uno de varios factores asociados con el aumento en el aprendizaje de los alumnos, puede afirmarse que tal implementación se correlaciona con el aprendizaje de los alumnos. Sin embargo, a menos que se eliminen los otros factores, no puede afirmarse que la implementación provocó el aumento en el aprendizaje de los alumnos. Los estudios de evaluación que no están bien diseñados pueden conducir a una confusión entre correlación y causalidad. Es necesario contar con estudios de evaluación bien diseñados para desenredar adecuadamente la correlación de la causalidad.

Conclusión

En este folleto se ha hecho hincapié en que los estudios de evaluación de programas, ya sean de gran escala o de pequeña escala, deben planearse desde el principio del programa y es imprescindible que se realicen con rigor científico. Ambos tipos de estudio por lo general forman parte de todo un paquete de evaluación de un programa en particular. Se dio un ejemplo del uso de ambos tipos de estudio en un programa para desarrollar materiales de planes de estudio. Sin embargo, sería igualmente cierto con referencia a la evaluación de cualquier programa de gran escala.

Los estudios de pequeña escala se utilizan para identificar variables o factores importantes que deben incluirse en un estudio grande, para ensayar elementos de pruebas y preguntas de cuestionarios, para decidir la mejor manera de elaborar factores o hipótesis y para contar con algunos ejemplos detallados de puntos generales que tal vez surjan de un estudio grande. Los estudios de pequeña escala también se usan para ensayar procedimientos que se emplearán en estudios de mayor escala. Sin embargo, es imposible generalizar a partir de estudios de pequeña escala. Las generalizaciones sólo pueden hacerse a partir de estudios de gran escala basados en muestras de probabilidad.

El rigor científico que debe aplicarse en ambos tipos de estudio quedó ejemplificado con los buenos procedimientos de muestreo, ya sea para muestras de probabilidad o para muestras de valor. Se hicieron algunas sugerencias para la presentación de los estudios de evaluación y para el cuidado que debe tenerse en la interpretación de los resultados.

Cualquiera que sea el tipo de estudio que se realiza, los informes a menudo se redactan para distintos tipos de públicos. La redacción de estos informes debe hacerse con cuidado. Se ofrecieron consejos basados en las experiencias de los dos autores. Por

último, existen algunos riesgos inherentes cuando se realizan estudios de evaluación. Se dieron algunos ejemplos de los tipos de riesgos y se sugirieron maneras de enfrentarlos.

Referencias

- Centers for Disease Control and Prevention. (1999). *Framework for program evaluation in public health* (Marco para la evaluación de programas en salud pública). Atlanta, GA: Author.
- Lewy, A. (Ed.). (1977). *Handbook of curriculum evaluation* (Manual de evaluación de planes de estudio). Paris: IIEP-UNESCO.
- Lewy, A. (1991). *National and school-based curriculum development* (Desarrollo de planes de estudio nacionales y basados en la escuela). Fundamentals of Educational Planning Series Núm. 40. Paris: IIEP-UNESCO.
- Lynn, J. and Jay, A. (1985). *Yes, Prime Minister* (Sí, Primer Ministro). London: BBC.
- Kish, L. (1987). *Statistical design for research* (Diseño estadístico para la investigación). New York: John Wiley & Sons.
- OECD (2005). *PISA 2003 Technical report* (Informe técnico). Paris: OECD.
- OECD (2005). *PISA 2003 Data Analysis Manual* (Manual de análisis de datos). Paris: OECD.
- Patton, M.Q. (1997). *Utilization-focused evaluation: The new century text* (Evaluación centrada en la utilización: texto para el nuevo siglo). Thousand Oaks, CA: Sage.
- Payne, D.A. (1994). *Designing educational project and program evaluations: A practical overview based on research and experience* (Diseño de evaluaciones de proyectos y programas educativos: una visión general práctica basada en la investigación y la experiencia). Boston: Kluwer Academic Publishers.
- Schubert, W. H. (1986). *Curriculum: perspective, paradigm, and possibility* (Planes de estudio: perspectiva, paradigma y posibilidad). New York: MacMillan.
- Scriven, M. (1967). The methodology of Evaluation (La metodología de la evaluación), in R. W. Tyler, R. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Perspectivas de la evaluación de planes de estudio) (AERA Monograph Series on Curriculum Evaluation, Núm. 1. Chicago: Rand McNally.

Wolf, R. M. (1990). *Evaluation in education, 3rd edition* (Evaluación en la educación, 3ª. ed.). Westport, CT: Praeger Publishers.

Payne, D.A. (1994). *Designing educational project and program evaluations: A practical overview based on research and experience* (Diseño de evaluaciones de proyectos y programas educativos: una visión general práctica basada en la investigación y la experiencia). Boston: Kluwer Academic Publishers.

La Academia Internacional de la Educación



La Academia Internacional de la Educación (AIE) es una asociación científica sin fines de lucro que promueve la investigación educativa, su difusión y la aplicación de sus resultados. Fundada en 1986, la Academia se dedica a fortalecer las aportaciones de la investigación, resolver los problemas críticos de la educación en todo el mundo y proporcionar una mejor comunicación entre los responsables de elaborar las políticas, los investigadores y los profesionales de la educación. La sede de la Academia se encuentra en la Academia Real de Ciencia, Literatura y Artes, en Bruselas, Bélgica, y su centro de coordinación está en la Universidad Tecnológica Curtin, en Perth, Australia.

El propósito general de la Academia consiste en fomentar la excelencia académica en todos los campos de la educación. Para lograr este fin, la Academia proporciona síntesis oportunas de resultados de investigaciones de importancia internacional. También hace evaluaciones de investigaciones, de sus bases probatorias y de sus aplicaciones a las políticas públicas.

Los miembros de la Junta de Directores de la Academia son:

- Erik De Corte, Universidad de Leuven, Bélgica (Presidente);
- Barry Fraser, Universidad Tecnológica Curtin, Australia (Director Ejecutivo);
- Monique Boekaerts, Universidad Leiden, Países Bajos;
- Jere Brophy, Universidad Estatal de Michigan, Estados Unidos de Norteamérica;
- Eric Hanushek, Instituto Hoover, Stanford, Estados Unidos de Norteamérica;
- Denis Phillips, Universidad Stanford, Estados Unidos de Norteamérica;
- Sylvia Schmelkes, Departamento de Investigaciones Educativas, México.

Los siguientes individuos, todos ellos miembros de la Academia, pertenecen al Comité Editorial para la serie *Políticas educativas*.

- Lorin Anderson, Universidad de Carolina del Sur, Estados Unidos de Norteamérica;
- Eric Hanushek, Instituto Hoover, Stanford, Estados Unidos de Norteamérica;
- T. Neville Postlethwaite, Universidad de Hamburgo, Alemania;
- Kenneth N. Ross, Instituto Internacional para la Planeación Educativa, París, Francia.

<http://www.smeccurtin.edu.au/iae/>

El Instituto Internacional para la Planeación de la Educación (IIEP) se estableció en París en 1963 bajo los auspicios de la UNESCO y con ayuda financiera inicial del Banco Mundial y la Fundación Ford. El gobierno francés proporcionó recursos para el edificio y el equipo del IIEP. En años recientes, el IIEP ha recibido apoyo de la UNESCO y de una amplia gama de gobiernos y organismos.

El IIEP es parte integral de la UNESCO y lleva a cabo actividades de investigación y capacitación que atienden las prioridades del programa general de educación de la UNESCO. Goza de autonomía intelectual y administrativa y funciona de acuerdo con sus propios estatutos. El IIEP tiene su propia Junta de Gobierno, que determina la orientación general de las actividades del Instituto y aprueba su presupuesto anual.

La misión del IIEP es el desarrollo de competencias en la planeación y administración educativas. Para lograr este fin, el IIEP utiliza diversas estrategias: la capacitación de planificadores y administradores educativos; el apoyo a instituciones de capacitación e investigación nacionales; el fomento de un ambiente que favorezca y apoye el cambio educativo, y la cooperación con países en el diseño de sus propias políticas y planes educativos.

La sede del IIEP en París está encabezada por un director, que cuenta con la asistencia de alrededor de cien profesionales y personal de apoyo. Sin embargo, éste es apenas el núcleo del Instituto. A lo largo de los años, el IIEP ha desarrollado con éxito asociaciones con redes de individuos e instituciones regionales e internacionales, tanto en países desarrollados como en vías de desarrollo. Estas redes apoyan al Instituto en sus distintas actividades de capacitación y también ofrecen oportunidades para extender el alcance de sus programas de investigación.

<http://www.unesco.org/iiep/>



Con motivo de la próxima celebración de la Asamblea General de la Academia Internacional de Educación, que se realizará en nuestro país en septiembre de 2006, cinco instituciones mexicanas acordaron traducir y publicar en español la serie *Políticas educativas*, editada originalmente por la Academia Internacional de Educación y el Instituto Internacional para la Planeación de la Educación.

El Centro Nacional de Evaluación para la Educación Superior, A.C. (Ceneval) ofrece servicios de evaluación a escuelas, universidades, empresas, autoridades educativas, organizaciones de profesionales del país y otras instancias particulares y gubernamentales del país y del extranjero.

El Departamento de Investigaciones Educativas (DIE) del Centro de Investigación y de Estudios Avanzados (Cinvestav) es un órgano descentralizado del gobierno federal que tiene como principales actividades la investigación, la enseñanza de posgrado y la difusión.

El Consejo Mexicano de Investigación Educativa (Comie) es una asociación civil cuyo propósito es impulsar y consolidar la actividad de grupos de investigadores en el campo educativo. Organiza cada dos años el Congreso Nacional de Investigación Educativa y edita la *Revista Mexicana de Investigación Educativa*.

El Instituto Nacional para la Evaluación de la Educación (INEE) tiene como tarea ofrecer a las autoridades educativas y al sector privado herramientas idóneas para la evaluación de los sistemas educativos, en lo que se refiere a educación básica (preescolar, primaria y secundaria) y media superior.

La Universidad Pedagógica Nacional (UPN) es una institución pública de educación superior cuyos objetivos son contribuir al mejoramiento de la calidad de la educación y constituirse en institución de excelencia para la formación de los maestros.

