

INVESTIGACIÓN: "El valor añadido en educación y la función de producción educativa"

Milagros Muñoz Martín.
Presidenta de ADIDE-Madrid.
Investigación conjunta de la Inspección de Educación
de la Comunidad de Madrid y Universidad Complutense de Madrid.

La Inspección de Educación de la Comunidad de Madrid (Subdirección General de Inspección) y la Universidad Complutense (Departamento de Métodos de Investigación y Diagnóstico en Educación) están llevando a cabo una investigación cuya referencia y contenido incluimos a continuación.

Proyecto de investigación: "El valor añadido en educación y la función de producción educativa"

Grupo de Investigación 'Evaluación y Medida', del Departamento de Métodos de investigación y Diagnóstico en Educación de la Universidad Complutense de Madrid. Grupo de Investigación de la Subdirección General de Inspección de la Comunidad de Madrid.

Director del proyecto: José Luis Gaviria Soto, Catedrático de Métodos de Investigación en Educación de la Universidad Complutense de Madrid.

Introducción

Entre los objetivos científico-tecnológicos prioritarios que se señalan en el Programa Nacional de Socioeconomía, se menciona el 'contribuir a la eficiencia de las instituciones y de las administraciones; (...) establecer la viabilidad de un modelo de desarrollo sostenible, requisito imprescindible de calidad de vida para todos los ciudadanos.'

En las primeras fases de desarrollo e institucionalización de nuestra democracia se consideraba como un elemento imprescindible de la misma el aumento del gasto en educación. El gasto en educación en España por ejemplo ha pasado del 3,1% del PIB en 1975 al 5,9% en 1994. La enseñanza obligatoria, primaria y secundaria en ese mismo año representaba el 4,1% del PIB, mientras que la media de la OCDE el 3,8%.

Esto supone un enorme esfuerzo económico, pues dado el aumento del propio PIB, y del estancamiento relativo de la población, el aumento del gasto por alumno se puede considerar espectacular. Sin prejuzgar los niveles en los que el gasto en educación debe establecerse, es indiscutible que nunca en la historia de la humanidad dicho gasto ha sido más elevado.

Este incremento responde al convencimiento generalizado de que la formación de capital humano que favorece la educación beneficia no sólo a aquellos sujetos que son beneficiarios directos de la acción formativa, sino que se generan unos efectos indirectos que constituyen un beneficio para todo el sistema social en su conjunto.

Serían estas externalidades las que justificarían la intervención del Estado en el sector.

En las fases de crecimiento cuantitativo del sistema educativo, aquellas en las que el objetivo primordial es extender la escolaridad más básica a la mayor cantidad posible de población, el incremento en el gasto está directamente ligado al crecimiento de las infraestructuras esenciales y a la formación del personal que ha de servir esas infraestructuras. Eso hace que las alternativas en la asignación del gasto sean pocas, y los retornos marginales a esas inversiones sean altos.

Pero cuando las infraestructuras básicas del Sistema Educativo no van a crecer más, la asignación del gasto presenta muchas alternativas distintas, y no todas ellas tienen el mismo retorno marginal. El modelo dominante de asignación del gasto afirma que los recursos deben asignarse de modo que el producto marginal por unidad monetaria invertida en cada input se iguale.

Además el gasto público en educación no puede crecer ni indefinida ni asistemáticamente. En un momento en el que el estado de bienestar parece haber llegado al límite de su expansión, la atribución de más recursos al sistema educativo implica necesariamente la detracción de los mismos de otros fines sociales valiosos.

Por otra parte nuestra sociedad se encuentra con cambios en su estructura demográfica que determinan necesidades educativas cambiantes. Así, tenemos que nos encontramos con un estancamiento o una disminución en el tamaño de las cohortes de edad que llenan el sistema educativo, debido a la baja tasa de natalidad. Pero al mismo tiempo, y como producto combinado de la baja tasa de natalidad nacional y de la aparición de un diferencial económico con países de nuestro entorno menos desarrollados, nos encontramos con unos flujos migratorios caracterizados por altas tasas de natalidad y necesidades educativas específicas que imponen nuevas exigencias al sistema educativo.

Por ello es posible que el gasto en educación no haya terminado su crecimiento relativo, pero indudablemente es necesaria la determinación de las prioridades a las que los limitados recursos disponibles deben hacer frente.

En este sentido se impone la necesidad de evaluar los resultados de las distintas políticas educativas en términos de eficacia absoluta, es decir, en términos de elevación de la cantidad y calidad de los productos educativos esperables del sistema, como de su eficiencia, es decir, de la relación entre la cantidad de recursos asignados para el funcionamiento del sistema y los logros obtenidos por el mismo.

La función de producción educativa

Resulta evidente que es preciso establecer la relación entre los recursos asignados a cada input y el aumento correspondiente del producto educativo. Esta relación se establece a través de la 'función de producción educativa' (Hanushek, 1988,1995).

La máxima cantidad de resultados para una cantidad dada de recursos, tomando las limitaciones impuestas por el proceso técnico subyacente como restricciones fácticas, es la función de producción. En educación el proceso técnico subyacente es el proceso pedagógico latente.

La función de producción toma el 'proceso técnico' subyacente como una condición, como lo dado, y pretende determinar qué combinación de los elementos del input maximizará el output.

La investigación pedagógica tratará de determinar cómo para una estructura de gasto determinada se puede maximizar el output.

La función de producción depende de los procesos técnicos subyacentes, por tanto es 'local' o contextual. Esto quiere decir que allá donde el proceso técnico, en este caso pedagógico, se vea modificado, habrá una variación en la función de producción. Por tanto una asignación homogénea de recursos puede resultar altamente ineficiente, dado que no todos los contextos se producirá la deseada maximización del producto obtenido.

Así, para un centro determinado la función de producción en la que tendrá la forma siguiente:

$$y_{it} = \lambda y_{it-1} + \alpha S_t + \beta F_{it} + \varepsilon_{it}$$

donde

- y_{it} es el rendimiento del alumno i el año t ,
- y_{it-1} es el rendimiento de ese mismo alumno en el curso anterior,
- S_t son las características de la escuela analizada.
- F_{it} son las características de background familiar
- ε_{it} es el residuo no explicado para ese alumno

Algunas variaciones incluyen el desdoblamiento de las características de la escuela analizada en dos componentes: el primero tiene que ver con los recursos que la escuela recibe sin controlar ni su cantidad ni su clase, el segundo con los recursos que la escuela controla directamente (Ladd y Walsh, 2002)

Este tipo de funciones son susceptibles de encontrar alguno de los cuatro siguientes problemas: especificaciones simplistas, multi-colinealidad, insuficiente varianza en los inputs, y diferente sensibilidades entre los inputs (Pritchett y Filmer, 1997).

El producto educativo y el valor añadido

Un determinante fundamental para la especificación de la 'función de producción' es la medida del producto. En España no ha habido una medida adecuada del producto educativo. "La escasez de datos ha propiciado que la literatura empírica sobre producción educativa sea bastante escasa en España" (San Segundo, 2001). Algunos intentos en esta dirección encontramos en Modrego y San Segundo (1988), San

Segundo (1991), Ipiña, Narvaiza y Ugarte (1994), García Díez (1997), Castillo (1998) y García (2000).

Algunos de estos trabajos utilizan el banco de datos que se creó en la evaluación de la Reforma de las Enseñanzas medias. En este banco de datos los rendimientos están medidos por medio de pruebas objetivas construidas al efecto. Sin embargo la nota por medio de pruebas objetivas en datos *cross-sectional* no tienen en cuenta el estado inicial de las unidades de análisis. Por eso el principal problema de este tipo de medida tiene que ver con la alta correlación entre los resultados y el estatus socioeconómico de los alumnos (Clotfelter y Ladd, 1996).

En otros se utilizan 'outputs' directos como número de egresados, y por lo tanto se atienen estrictamente a criterios cuantitativos. Estas medidas del producto escolar se basan en la proporción de alumnos de la escuela que pasan de un año al siguiente de los fractiles de rendimiento inferiores a los medios o superiores. El mayor inconveniente de este sistema consiste en que no tiene en cuenta las diferencias en composición de los alumnos de una escuela de uno a otro año (Ladd y Walsh, 2002).

Es necesario hacer caso de criterios cualitativos, lo que implica una medida precisa de los resultados educativos. Cuando en alguno de estos estudios se intenta obtener una medida no sólo cuantitativa de los resultados, sino una cualificación de la calidad del producto obtenido, se acude a la nota asignada por el profesor. Pero en ese caso nos encontramos con el problema de que las puntuaciones de distintos centros están en escalas distintas, y sin un proceso previo de equiparación las comparaciones resultan cuando menos cargadas de incertidumbre (Gaviria, 2005).

Sólo hay un intento de realizar un estudio longitudinal (Marchesi y Martín, 2002), con resultados muy interesantes desde el punto de vista educativo. Sin embargo no es posible su utilización en la determinación de una función de producción dado que no incluyen datos económicos referidos a los inputs.

En un reciente trabajo Poveda (2002) intenta establecer una relación entre los datos de evaluación educativa producidos por el INCE (1998) con los datos disponibles de gasto educativo de las distintas comunidades autónomas en los años anteriores a la toma de datos. Por ser la del INCE la primera evaluación de la educación secundaria que, dada la metodología utilizada, permitía comparaciones con evaluaciones posteriores, este tipo de estudios es muy prometedor e interesante. Desgraciadamente el procedimiento de muestreo utilizado por el INCE no permite el análisis de la evolución de los aprendizajes de los alumnos individuales en dos puntos temporales distintos.

Por eso es necesario el uso de datos longitudinales, medidos con pruebas objetivas validadas con los métodos psicométricos más sofisticados de modo que pueda medirse el valor añadido.

En el plano de la eficacia escolar es crucial la medida adecuada y precisa de los resultados para la legitimidad y deseabilidad de cualquier sistema de evaluación que pretenda llegar a la rendición de cuentas.

Algunos indicadores utilizados para la determinación del producto educativo de los centros escolares han sido la puntuación media en las pruebas de rendimiento allá donde

estas pruebas existen, o bien las proporciones de alumnos que promocionan con todas las asignaturas positivamente evaluadas.

Otro tipo de medidas, basadas en las calificaciones otorgadas por los profesores se asumen el supuesto no contrastado, y de hecho bastante dudoso, de que las calificaciones escolares son comparables, esto es, son medidas que están en la misma escala (Gaviria, 2005).

Sobre la definición y medida del valor añadido

El crecimiento y mejora académica de los niveles de aprendizaje está vinculado teóricamente al concepto de valor añadido. Un buen entendimiento del término del valor añadido es fundamental para mejorar la escolaridad pública. La idea básica del concepto de valor añadido (en el aprendizaje del alumno) es la diferenciación y el aislamiento de los factores puramente escolares que influyen en los resultados escolares, separándolos de los factores individuales y sociales que irremediamente también intervienen. El concepto de valor añadido habla conjuntamente de estatus o nivel actual de conocimientos y destrezas y de incremento del aprendizaje o ganancia a través del tiempo debido a la acción escolar. Supone el establecimiento de la relación del punto de partida y del punto de llegada (real, deseado o regulado). Además, las medidas de valor añadido tienen dos elementos clave: enfoque en el progreso del aprendizaje de los alumnos en periodos sucesivos; enfoque en rendimientos de los alumnos individuales más que en indicadores genéricos de la escuela.

Este término de valor añadido es llamativo y fuente de polémica dentro de la discusión sobre evaluación de los sistemas educativos. Sus orígenes se enraízan con las nociones de bienes, riqueza y prosperidad en economía política (Sanders, 1999). Una referencia temprana del uso de este término en conexión con evaluaciones de intervenciones en un proceso de crecimiento educativo se puede encontrar en Bryk y Weisberg (1976). Su uso moderno transmite la idea de "progreso relativo", tal y como señala Fitz-Gibbon (1997). Una interpretación más apolítica y estadística reconoce que, independientemente de la forma que adopte, los análisis de valor añadido se centran en modelos para "comparaciones ajustadas" (Goldstein y Spiegelhalter, 1996).

Sin embargo, la atención sobre la medida del valor añadido es relativamente nueva, si bien los procedimientos y técnicas para medirlo están bien definidas en la literatura científica. ¿Qué se entiende operativamente por una medida de valor añadido? Son expresiones empíricas de lo que se considera ganancia, crecimiento, o incremento, expresiones de medida de la distancia existente entre el nivel actual de conocimientos, en relación con el nivel previo, más la consideración valorativa sobre la distancia deseable de crecimiento.

Y son muchas las definiciones operativas a través de las cuales se puede definir una medida de ganancia. Quizá la más sencilla sea el concepto de *ganancia bruta o absoluta*, que es una medida directa de la distancia entre .

el nivel actual de conocimientos y destrezas (tiempo=1, Y_{i1})
y el nivel inicial (tiempo=0, Y_{i0}) ($Y_{i1} - Y_{i0}$)

Esta operacionalización se puede refinar a través del ajuste de covariables en el nivel actual o post-test, considerando que la principal covariable es el nivel inicial o rendimiento previo. La distancia queda entonces definida de esta forma:

$$Y_{i1} - E(Y_{i1} | Y_{i0}),$$
 y se denomina *ganancia residual*.

Una alternativa más ajustada consiste en considerar tanto el nivel inicial como el actual como medidas de resultados, situándonos en el marco de las medidas repetidas. La *ganancia estimada* queda expresada de la siguiente manera:

$$E(Y_{i1}) - E(Y_{i0})$$

, que supone una mejora notable en la concepción del crecimiento, puesto que no se ve afectado por artefactos del diseño.

Hay dos recientes desarrollos en investigación educativa que ayudan a tratar metodológicamente el problema planteado. Son el desarrollo de las medidas de crecimiento o ganancia y los modelos jerárquicos lineales.

Así, la primera gran aportación es la línea de los trabajos de entre otros Bryk y Weisberg (1976), Rogosa et al. (1982), Rogosa and Willet (1985), Rogosa (1995), que disipan la noción de que las medidas de ganancia o crecimiento son inherentemente poco fiables, al tratar las medidas de crecimiento dentro del amplio marco de los modelos de crecimiento para comprender el cambio y la evolución individual.

La segunda aportación es conocida como modelos jerárquicos lineales, modelos multinivel o modelos de efectos mixtos, que aportan el complemento metodológico para un tratamiento adecuado del crecimiento o del cambio en diseños de muestras estratificadas que son los más comunes en investigación educativa (medidas repetidas de estudiantes en el tiempo, estudiantes anidados en escuelas, distritos, etc.), absolutamente necesarios puesto que respetan la naturaleza anidada de los datos, reconociendo expresamente las distintas fuentes de variabilidad. La investigación para el estudio del cambio ha estado llena de inadecuaciones en su conceptualización, medida y diseño (Raudenbush y Bryk, 2002). La preocupación conceptual es que se necesitan modelos que guíen la investigación sobre el fenómeno a estudiar, puesto que raramente se identifica un modelo explícito para el cambio individual dentro de organizaciones como las escolares (Raudenbush, 2001). En relación a la medida, los estudios de cambio típicamente utilizan instrumentos que fueron diseñados para discriminar entre individuos en puntos fijos en el tiempo. La adecuación de tales medidas para distinguir diferencias entre tasas de crecimiento entre sujetos es raramente considerada. Finalmente, el problema del diseño es quizá el más importante, ya que la

mayoría de los estudios recogen datos sólo en dos momentos distintos, siendo inadecuados para estudiar el crecimiento individual (Bryk y Weisberg, 1977; Rogosa et al., 1982; Bryk y Raudenbush, 1987).

El desarrollo de los modelos jerárquicos lineales ha creado un potente conjunto de técnicas para la investigación del cambio individual dentro de organizaciones como las escolares. Cuando se aplican junto a medidas válidas para diseños con múltiples puntos temporales, estos modelos ofrecen una herramienta adecuada para el estudio de la estructura y la predicción del cambio individual. Una ventaja de esta aportación es que es flexible con respecto a la estructura temporal de las observaciones. Igualmente importante es que los modelos multinivel facilitan el estudio de los efectos de las covariables en cada nivel de anidamiento (Laird y Ware, 1982; Bryk y Raudenbush, 1992; Goldstein, 1995; Snijders y Bosker, 1999).

Muchos sistemas de evaluación emplean alguna forma de los modelos multinivel (Willms, 1992; Willms y Raudenbush, 1989; Snijders y Horn, 1994; Webster y Mendro, 1997; Gray et al., 1995; Harker y Nash, 1996; Mayer, 1996; Bryk et al., 1998). Otra contribución positiva es que los modelos multinivel ayudan a evitar informaciones engañosas derivadas de la agregación de puntuaciones en escuelas, distritos o comunidades autónomas, centrándose en el nivel de cambio o mejora del alumno (Meyer, 1996).

En evaluación educativa, la noción de eficacia diferencial de un profesor o de una escuela está a menudo equiparada con análisis de ganancias residuales basados en sucesivas evaluaciones del rendimiento de los alumnos, como por ejemplo el sistema de evaluación de Dallas (Webster y Mendro, 1997; Meyer, 1996) o la estrategia usada por Karol Fitz-Gibbon (1997) en el Reino Unido. Desde nuestro punto de vista, también compartido por otros investigadores (Thum, 2002) la alternativa para la medida del valor añadido es la propuesta desde los modelos de efectos mixtos con medidas repetidas (de los cuales, los modelos de crecimiento son un caso especial). Los ejemplos de esta alternativa también son profusos y extensamente probados, como las investigaciones de Sanders y Horn (1995) con el sistema de evaluación del valor añadido de Tennessee (*Tennessee Value-Added Assessment System* o TVAAS) o los análisis de Bryk et al. (1988) que usan evaluaciones longitudinales de los estudiantes de las escuelas primarias de Chicago. Ambas experiencias estiman la verdadera ganancia basándose en las puntuaciones individuales de los alumnos, sin considerar ganancias brutas. Lo que hace que el estudio del crecimiento o de las ganancias emplee mejor la información disponible y permita aislar la estructura de varianza-covarianza de las pruebas.

Antecedentes y resultados previos

Hay antecedentes de investigación referidos a la viabilidad de sistemas nacionales de evaluación de los resultados educativos. Entre estas investigaciones podemos contar los estudios iniciados por el INCE (1997) en cuyo desarrollo participaron la mayoría de los miembros de este equipo de investigación. En el ámbito internacional, se puede mencionar como estudios con la misma intención el Proyecto PISA 2000 (OCDE, 2000), PISA 2003 (OCDE, 2003), las evaluaciones de TIMMS (Martin et al. 2000), la evaluación sistemática que se viene desarrollando en las comunidades autónomas de Navarra y La Rioja, entre otras (Tourón y Gaviria, 2000

www.educarioja.com/evaluacion/primaria/index.html; Tourón y Gaviria, 2000
<http://www.educarioja.com/evaluacion/primaria/index.html>; Tourón y Gaviria, 2001
<http://www.educarioja.com/evaluacion/primaria/index.html>).

Sin embargo, todos estos proyectos adolecen de la falta, en primer lugar, de una adecuada medida de valor añadido y de seguimiento del crecimiento en los niveles de aprendizaje de los alumnos, y, en segundo lugar, de la consideración y recolección de datos de carácter económico. Ahora bien, la medida del valor añadido y los sistemas de rendición de cuentas tienen su impacto tanto en iniciativas de carácter político administrativo, como los sistemas desarrollados en California (Public School Accountability Act, 1999), en Tennessee (Tennessee Value-Added Assessment System, Bratton, Horn, y Wright, 1999 ;Sanders y Horn, 1995) o en Chicago (Grade Productivity Index, Bryk et al. 1998) como en trabajos de investigación (Thum 2002a, 2002b; Thum y Castro, 1999; Thum y Castro, 2000) dedicados a la reflexión, desarrollo y prueba de distintas medidas de valor añadido. Por otro lado, el trabajo de Poveda (2002) pone de manifiesto la ausencia palpable de datos y de estudios sistemáticos sobre la función de producción educativa.

Finalidad del proyecto

El objetivo fundamental de este proyecto consiste en diseñar una medida del crecimiento de los logros escolares basada en el concepto de valor añadido en educación.

Para ello se desarrollará una evaluación longitudinal de dos dimensiones fundamentales: el aprendizaje matemático y la comprensión lectora en el último ciclo de la educación primaria y a lo largo de toda la educación secundaria obligatoria. Del mismo modo se tomarán datos de las variables de background socioeconómico de los alumnos, las familias y el entorno escolar y de organización de los centros.

Con toda esta información se construirá y validará un modelo de valor añadido.

Objetivos específicos

El objetivo general de este proyecto de investigación es doble, e implica diseñar una medida del crecimiento de los logros escolares basada en el concepto de valor añadido en educación, y posibilitar de esta forma la futura formulación de una función de producción educativa que establezca la relación entre esos logros y los recursos asignados a la institución escolar. Esta finalidad está directamente vinculada con la cuarta línea prioritaria del Programa Nacional de Socioeconomía, puesto que repercute en la evaluación de políticas públicas como es la política educativa.

Los objetivos específicos en los cuales se despliega esta finalidad principal son los siguientes:

- 1. Determinar la viabilidad de un sistema de evaluación educativa basado en el valor añadido.**

1. Definir y medir el valor añadido de los resultados escolares.
2. Diseñar indicadores fiables de crecimiento o progreso del aprendizaje escolar basado en la evaluación del rendimiento del alumnado a lo largo del tiempo.
3. Construir un modelo estadístico para describir y explicar el crecimiento de los centros escolares que permita seguir el progreso de unidades múltiples y anidadas (subgrupos de estudiantes, escuelas, distritos, etc.).
4. Determinar la capacidad predictiva del rendimiento futuro que tienen los factores personales y contextuales considerados en el modelo.

2. Establecer la productividad de los recursos asignados al sistema educativo en términos de incremento de los aprendizajes básicos de los alumnos. Lo que supone:

1. Determinar la productividad media de cada uno de los recursos empleados en el proceso educativo.
2. Determinar la productividad marginal de los factores más importantes.
3. Evaluar la productividad total de los factores de la producción educativa.
4. Especificar la importancia de la naturaleza del centro como factor de diferenciación de la eficiencia.
5. Establecer la importancia del estatus socio-económico y cultural de las familias y del centro escolar como factor de diferenciación de la eficiencia.

3. Sentar las bases para la evaluación futura de las políticas de gasto educativo en relación a su susceptibilidad de incremento del capital humano.

1. Determinar cuáles son los factores que facilitan la integración de los grupos minoritarios en el sistema educativo y de qué forma ayudan al incremento del capital humano de estos grupos.
2. Definir medidas de producción escolar que sean sensibles a los resultados relativos a la integración social de las minorías.

Modelo teórico

Se trata de un modelo general lineal multivariado de efectos mixtos. Los modelos lineales de crecimiento individual pueden utilizarse para: a) la estimación de la curva media de crecimiento y la extensión del modelo de la variación individual; b) la evaluación de la fiabilidad de las medidas empleadas para el estudio del nivel actual y de crecimiento; c) la estimación de la correlación entre el nivel actual de dominio y la tasa de cambio en el aprendizaje; y d) la modelización de las relaciones de los predictores del nivel individual (estudiantes) y grupal (escuela) y las medidas del nivel actual y de cambio (Raudenbush y Bryk, 2002).

El primer nivel recoge la estimación del rendimiento del alumno (medidas repetidas).

Sea $k = 1, 2, \dots, K$ puntuaciones de un alumno (i) en $t = 1, 2, \dots, T$ ocasiones en una determinada escuela (j). La puntuación, Y_{tijk} , para, comprensión lectora y matemáticas, por ejemplo, se puede representar como sigue:

$$Y_{tijk} = \pi_{0j} + \pi_{1j}(\text{Time}_{tijk} - 1) + \pi_{2j} \text{Leng}_{tijk} + \pi_{3j} \text{Mat}_{tijk} + e_{tijk}$$

Donde

π_{0j} estima el nivel inicial en comprensión lectora y

π_{1j} es la ganancia general para todas las materias conjuntamente.

Esperamos que los estudiantes varíen de alguna forma en este nivel inicial y en la tasa de ganancia, por tanto parece razonable que el modelo permita la variación aleatoria en el siguiente nivel. Los efectos de las materias de lengua y matemáticas están representados por

$$\pi_{2j} \text{ y } \pi_{3j}$$

y son ajustes al nivel inicial que toma como referente la comprensión lectora. Por ejemplo, si quisiéramos conocer el efecto del nivel inicial en matemáticas no tendríamos más que calcular

$$(\pi_{0j} + \pi_{3j})$$

o si quisiéramos conocer la tasa de ganancia en lenguaje habría que calcular

$$(\pi_{1j} + \pi_{2j}).$$

El término

e_{tijk} representa el error que para cada unidad son variables independientes y (σ^2) normalmente distribuidos, con media 0 y varianza constante

Este primer modelo permite recoger las estimaciones para el nivel inicial y las ganancias y describe la variabilidad entre estudiantes y entre escuelas.

El segundo nivel representa al estudiante, incluyendo covariables

$$(X_{pqj})$$

que afectan a la variación individual del aprendizaje de los alumnos, como por ejemplo el género del alumno o su nivel socioeconómico. Tales relaciones aparecen en un modelo de regresión múltiple y multivariada como el que sigue:

$$\pi_{p ij} = \beta_{p0j} + \sum_{q=1}^{Q_q} \beta_{pqj} X_{qij} + r_{p ij}$$

Así

β_{p0j} es la media general de rendimiento en comprensión lectora para un alumno de la escuela j

β_{pqj} son los efectos diferenciales de las características de los alumnos

(X_{pqj}) para cada covariable Q

En nuestro caso concreto las covariables especificadas en este nivel se pueden encontrar en el epígrafe de variables en este mismo apartado. Por último,

r_{ij} es la variación residual en este nivel, distribuida normalmente con media 0 y (τ) varianza constante

El tercer nivel representa al centro escolar incluyendo factores que afectan (W_j) sistemáticamente a la escuela

La influencia de tales efectos viene representada por

$$\gamma_{pqs}$$

empleando un modelo de regresión múltiple y multivariado entre centros escolares, representado como sigue:

$$\beta_{pqj} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sj} + v_{pqj}$$

$$\beta_{pqj} = \gamma_{pq0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sj} + u_{pqj}$$

Los subíndices p , q y s están referidos a los efectos intra-estudiantes (variación en la puntuaciones de un mismo estudiante, p), los efectos entre-estudiantes (diferencias entre distintos alumnos, q) y los efectos de la variación entre-escuelas (variación entre centros de enseñanza, s).

Así, en el nivel de centro, hay un conjunto de covariables (W_{sj})
 numeradas S_{pq} .

De modo análogo a los niveles anteriores

γ_{pq0} representa el rendimiento medio inicial en cada escuela en comprensión lectora. Mientras que

γ_{pqs} representa la aportación de cada covariable.

El término

u_{pqj} representa la variación residual entre escuelas, una vez controlados todos (ψ) los factores individuales y de centro incluidos en el modelo, que está normalmente distribuido, con media 0 y varianza constante

Diseño

Se propone un diseño **multinivel, longitudinal (medidas repetidas) y multivariado**.

La caracterización como *multinivel* es debida a la estructura anidada de los datos, siendo necesario respetar la estructura de la variabilidad en cada uno de los niveles de agregación definidos: las puntuaciones en cada materia están anidadas en cada punto temporal (nivel 1: medidas repetidas), cada conjunto de puntuaciones están anidadas en el alumno (nivel 2: alumnos), alumnos que a su vez están agrupados en distintos centros educativos (nivel 3: escuelas).

Estudio longitudinal: se tomará como línea base las mediciones del curso académico 2004-05, siguiendo la serie temporal con las del curso académico 2005-06, para observar el crecimiento y la evolución de la cohorte incluida en la muestra.

Las mediciones se realizarán al comienzo y al final del curso escolar, recogiendo un

total de 4 puntuaciones para cada alumno en cada una de las dimensiones incluidas en el estudio.

La unidad de análisis es el estudiante individual, y las unidades del sistema de evaluación serán las escuelas y los distintos grupos de interés que se establezcan (materias, distintos grupos específicos de alumnos, etc.).

El carácter multivariado hace referencia directa a la variable de respuesta que estará compuesta conjuntamente por las puntuaciones de los alumnos en las dimensiones seleccionadas.

Variables

Variables de respuesta: las puntuaciones obtenidas en las pruebas elaboradas ad-hoc aplicadas a los alumnos de los centros muestreados, escolarizados durante los cursos académicos 2005-06 y 2006-07, midiendo al comienzo y al final del curso escolar (mayo). Por tanto se realizará un total de 4 mediciones para cada alumno en cada una de las materias de estudio.

Los valores de las variables de respuesta, conceptualizadas en el contexto de la Teoría de Respuesta al Ítem, serán estimados mediante procedimientos de Máxima Verosimilitud Marginal, como los incluidos en programas como BILOG-MG, (Muraki, 1994) o ICL (Hanson, 2002). Los estimadores de los parámetros de los ítems de las pruebas también serán calculados con el mismo programa.

Las puntuaciones serán estimadas en escala estandarizada (sobre los datos de la propia muestra) y posteriormente reescaladas a una media de 250 con desviación típica de 50. Se trata por tanto de una escala arbitraria que puede ser modificada fácilmente mediante cualquier transformación lineal.

Algunas de las variables predictoras más importantes en los distintos niveles, alumno y centro escolar, serán las que se relacionan a continuación.

Nivel 2: Alumno

1. Género
2. Grupo cultural de referencia
3. Dominio del español
4. Repetición
5. Nivel de aspiraciones académicas
6. Participación en actividades extraescolares
7. Horas semanales dedicadas al estudio

8. Horas semanales dedicadas a actividades extraescolares
9. Horas semanales dedicadas a ver la TV
10. Horas semanales dedicadas a otros ocios
11. Nivel socio-cultural de la familia
12. Recursos culturales del hogar
13. Acceso a recursos informáticos e Internet en el hogar
14. Situación laboral del padre
15. Situación laboral de la madre
16. Profesión del padre
17. Profesión de la madre
18. Nivel de estudios del padre
19. Nivel de estudios de la madre
20. Índice de posesiones del hogar

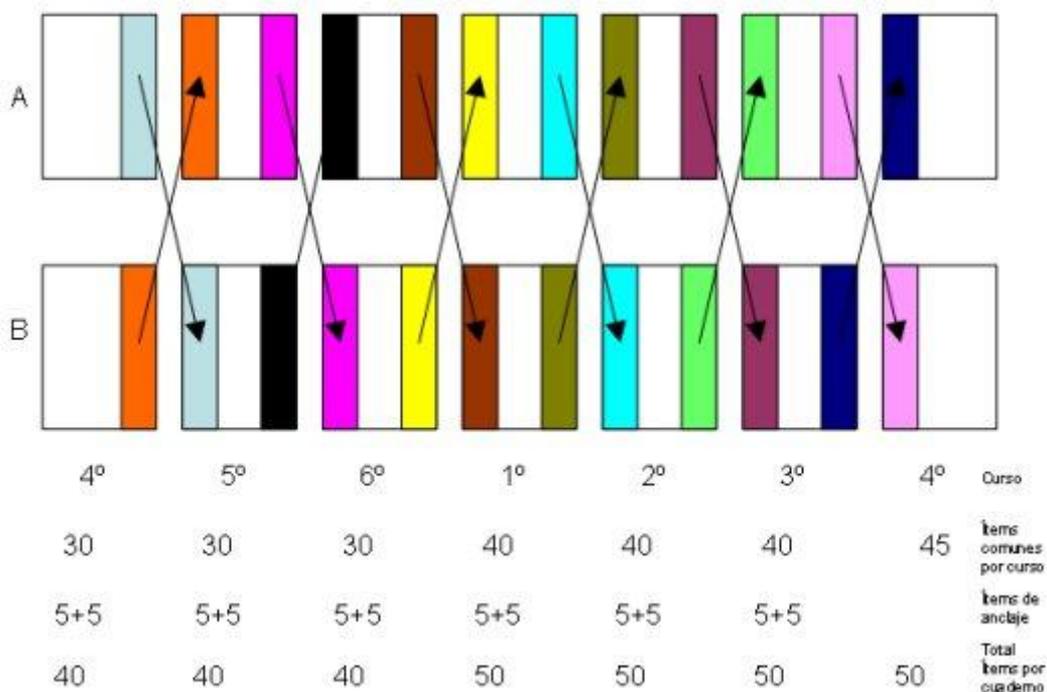
Nivel 3: Centro escolar

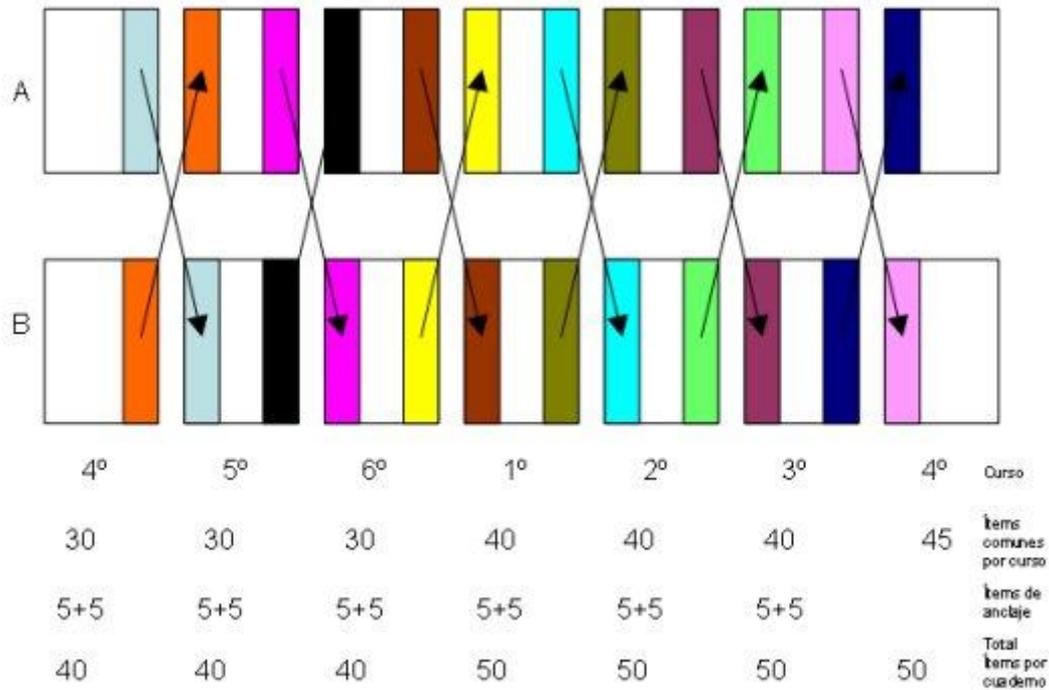
1. Titularidad del centro
2. Proporción de alumnos inmigrantes
3. Número de alumnos en el aula
4. Ratio media profesor/alumnos del centro
5. Carga horaria de cada profesor del grupo
6. Nivel formativo del profesor de cada materia
7. Experiencia del profesor
8. Recursos didácticos disponibles
9. Acceso a medios informáticos e Internet en el centro
10. Recursos bibliográficos
11. Actividades extraescolares organizadas por el centro

Instrumentos de medida

- Especialmente importante es todo lo que atañe a la medida de las variables dependientes, es decir lo que afecta a los resultados escolares. Las pruebas serán elaboradas por expertos en cada materia a los que se les proporcionará formación técnica necesaria para realizar dicha función.
- Una consideración de suma importancia es la necesidad de que las puntuaciones de rendimiento educativo que se tomarán en cada materia a lo largo de la duración del estudio estén en la misma escala. Dado que deseamos determinar las variaciones que se producen en el aprendizaje de los alumnos, es imprescindible que los cambios medidos puedan ser atribuidos a cambios realmente producidos, y no a variaciones en la dificultad de las pruebas aplicadas.
- Esta invarianza de escala puede lograrse mediante la utilización de los adecuados diseños de equiparación de las pruebas combinadas con las técnicas de estimación basadas en Teoría de respuesta al Ítem.
- En el diseño de equiparación nos encontramos con algunos condicionantes importantes.
- En primer lugar, como hemos mencionado si queremos medir el cambio todas las puntuaciones deben estar en la misma escala.
- Para ello las pruebas de dos cursos consecutivos deben contener algunos ítems comunes, denominados ítems de anclaje.
- Sin embargo, un mismo alumno no debe enfrentar dos veces el mismo ítem, dado que se produciría efecto aprendizaje.
- Como consecuencia es necesario crear para cada curso dos formas paralelas que deberán compartir el núcleo principal de ítems, y que cada una de ellas contendrá un grupo de ítems variantes que serán comunes con una de las formas del curso siguiente en diseño cruzado.

En el siguiente gráfico puede verse el esquema que se utilizará para la materia de matemáticas.





- En Comprensión Lectora se utilizará una variante de este diseño, combinando cuatro formas distintas que serán aplicadas en dos cursos consecutivos. Cada alumno contestará en total 4 formas distintas desde el comienzo del estudio hasta el final del mismo.
- Para las variables contextuales, se elaborará un cuestionario individual de recogida de información de cada alumno, diseñado para recoger las covariables mencionadas en el nivel de alumno.

Muestreo

Se trata de crear una base de datos longitudinal de estudiantes a lo largo de los cursos académicos 2005-06 y 2006-07. Dada la previsible mortalidad experimental propia de los estudios de series temporales y que la unidad de muestreo es el centro, la muestra ha de ser necesariamente grande.

Nos vamos a referir a la población objetivo, a los dominios, a los estratos, los conglomerados, la correlación intra-clase, y el tamaño de la muestra (Blalock, 1972; Kalton, 1983; Kish, 1987; Som, 1996).

La población objetivo

Población incluida: todos alumnos que entre octubre de 2005 y junio de 2007 realicen alguno de los cursos comprendidos entre 5º primaria y 4º de la E.S.O.

Población explícitamente excluida: los alumnos con necesidades educativas especiales

tales que impidan contestar la prueba en igualdad de condiciones espaciales y temporales con los demás alumnos del grupo.

Se instruirá a los aplicadores para que apliquen a estos alumnos las pruebas igual que al resto de sus compañeros, y sólo a posteriori identificarán dichas pruebas para su exclusión.

Los dominios

Las divisiones administrativas constituidas por las direcciones territoriales en la Comunidad Autónoma de Madrid se convierten en los dominios de referencia.

Los estratos

En cada uno de los dominios se definen distintos estratos. Los estratos están definidos por las redes de centros públicos, centros privados concertados y centros privados.

Los conglomerados y la correlación intra-clase

Dado que las unidades primarias de muestreo son los centros escolares, se impone un muestreo polietápico, de forma que dentro de los estratos definidos se muestrean centros, que constituyen los conglomerados. En este tipo de muestreo es fundamental la consideración del efecto diseño, que viene determinado por el valor de la autocorrelación. Por la experiencia de muestreos similares en otros estudios (INCE, 1998) puede afirmarse que la autocorrelación no será mayor que 0,25. Este valor nos permitirá determinar el tamaño de la muestra. A modo de ejemplo, en la evaluación de la secundaria realizada por el INCE en 1997, la aplicación de la prueba piloto permitió obtener una estimación de los valores de la correlación intra-clase para los grupos de 14 y de 16 años, siendo esos valores de 0,13 y 0,22 respectivamente.

En el muestreo por conglomerados el efecto diseño tiene un efecto negativo. Esto quiere decir que cuanto mayor sea la correlación intra-clase, mayor deberá ser el tamaño de la muestra final respecto de lo que hubiera sido si se tratase de muestreo aleatorio simple.

Que la muestra sea representativa supone que todos los sujetos de la población objetivo, o en su caso del dominio, tengan la misma probabilidad de pertenecer a la muestra. Es decir, el muestreo se realizará con probabilidad proporcional al tamaño del centro.

Referencias

Blalock, H. M. (1972) Social Statistics, New York: McGraw-Hill

Bratton, S.E. Jr.; Horn, S.P. y Wright, S.P. (1996). Using and Interpreting Tennessee's Value-Added Assessment System: A Primer for Teachers and Principals, pp. 15-18.

Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101 (1), 147-158.

Bryk, A. S. & Raudenbush, S. W. (1992). Hierarchical linear models: applications and data analysis methods. Newbury Park, CA.: Sage.

Bryk, A. S. and Weisberg, H. I. (1976). Value-added Analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics*, 1, 127-155.

Bryk, A. S., Thum, Y. M., Easton, J. Q., and Luppescu, S. (1998). Assessing School Academic Productivity: The Case of Chicago School Reform. *Social Psychology of Education*, 2, 103-142.

Castillo, I. (1998) Rentabilidad educativa: un análisis comparativo de los centros de la Comunidad Autónoma de Cantabria para los niveles de E.S.O. y B.U.P.-C.O.U., en I. Castillo y otros(coord..) 1998: VII Jornadas de la Asociación de Economía de la Educación, A.E.D.E.-universidad de Cantabria, Santander, pp. 373-386.

Coltfefer, C. y Ladd, H. F. (1996) Recognizing and rewarding success in public schools. En H. Ladd, *Holding schools accountable: performance-based reform in education*, Washington, D.C.: Brookings Institution.

Corwin.

Fitz-Gibbon, C. T. (1997). The Value-added National Project: Final Report. Feasibility studies for a national system of value-added indicators. London: SCAA.

García Díez, M. M. (1997) Estimaciones de funciones de producción educativas: la enseñanza de la economía a nivel universitario, en A. Olaya Iniesta y C. Selva Sevilla (coord..) (1997): *Actas de las V Jornadas de la Asociación de Economía de la Educación*, Ediciones de la Universidad de Castilla-La Mancha, Cuenca, pp. 255-270.

Gaviria, J.L. (2005) La equiparación del expediente de bachillerato en el proceso de selección de alumnos para el acceso a la Universidad. *Revista de Educación*, 337, pp. 351-387.

Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A*, 159, 384-443.

Gray, J., Jesson, D., Goldstein, H., Hedger, K., and Rabash, J. (1995). A multi-level analysis of school improvement: Changes in school's performance over time. *School Effectiveness and School Improvement*, 6, 97-114.

Hanushek, E. A. (1988) *Production Functions in Education*, en Husén, T. y Postlethwaite, T. N. (eds.) (1988): *The International Encyclopedia of Education. Research and Studies*, reprinted, Pergamon Press, Oxford, vol. 7, pp. 4059-4069.

Hanushek, E. A. (1995) *Education Production Functions*, en Carnoy, M. (ed.) (1995b): *International Encyclopedia of Economics of Education*, second edition, Pergamon

Press, Oxford, pp. 277-283.

Harker, R. and Nash, R. (1996). Academic outcomes and school effectiveness: Type "A" and Type "B" effects. *New Zealand Journal of Educational Studies*, 32, 143-170.

INCE (1998) 2. Los Resultados escolares. Ministerio de Educación y Cultura. Madrid

Ipiña, A. Narvaiza, J. L. Y Ugarte, J. V. (1994) Acerca de la producción educativa. Estudio de los Centros Educativos de la CAPV, Departamento de Educación Universidades e Investigación de la Comunidad Autónoma Vasca, Vitoria.

Kalton, G. (1983) *Introduction to Survey Sampling*, Sage, London

Kish, L. (1987) On the future of survey sampling. En N. K. Namboordi (Ed.) *Survey sampling and measurement*. Academic Press. N. Y.

Ladd, H. F. y Walsh, R. P. (2002) Implementing value-added measures of school effectiveness: getting the incentives right, *Economics of Education review*, 21, pp. 1-17.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrika*, 65, 581-590.

Marchesi, A. y Martín, E. (compiladores) (2002) *La evaluación de la educación secundaria. Fotografía de una etapa polémica*. Editorial SM. Fundación Santa María. IDEA

Martin, M.O. et al (2000). *Effective schools in science and mathematics*. IEA's third international mathematics and science study. Boston College, MA:TIMMS International Study Center.

Meyer, R. H. (1996). Value-added indicators of school performance. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's Schools: The role of incentives* (pp. 197-223). Washington, DC: National Academic Press.

Modrego, A. y San Segundo, M. J. (1988) Ecuaciones de rendimiento escolar para la evaluación de la Reforma de las Enseñanzas Medias, *Revista de Educación*, 287, pp.147-179.

OCDE (2000). *Measuring student knowledge and skills: The Pisa 2000 assessment of reading, mathematical and scientific literacy*. Paris: OCDE.

Orden Hoz, A. de La (1993) *La escuela en la perspectiva del producto educativo. Reflexiones sobre la evaluación de centros docentes*. *Bordon*, 45(3) 263-270.

Orden, A. de la (1995) *Hacia un modelo para evaluar la calidad universitaria Ponencia en el 'Seminario sobre evaluación de la calidad universitaria'* Centro Anáhuac de Investigación y Servicios Educativos México

Poveda, F. M. (2002): *La Economía de la Educación: fundamentación histórica y situación actual. Una aplicación práctica*. Tesis Doctoral inédita. Universidad de

Navarra, Pamplona.

Pritchett, L. y Filmer, D. (1997) What Education Production Functions Really Show: A Positive Theory of Education Expenditures, World Bank Working Paper 1795.

Raudenbush, S.W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, 52, 501-525.

Raudenbush, S.W. and Bryk, S.A. (2002). Hierarchical linear models. Applications and data analysis methods. London: SAGE.

Rogosa, D. R. (1995). Myths and methods: "Myths about longitudinal research" plus supplemental questions. In J. M. Gottman (ed.), *The Analysis of Change*. Mahwah, NJ: LEA.

Rogosa, D. R. and Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, 50, 203-228.

Rogosa, D. R., Brand, D. and Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 90, 726-748.

San Segundo, M. J. (1991) Evaluación del sistema educativo a partir de datos individuales, *Economía Industrial*, 278, Marzo-abril, pp. 23-37.

Sanders, W.L. & Horn, S.P. (1995). The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment. In Shrinkfield, A.J., & Stufflebeam, D. (Eds.), *Teacher Evaluation: Guide to Effective Practice* (pp. 337-350). Boston, MA: Kluwer.

Saunders, L. (1999). A brief history of educational "Value Added": How did we get to where we are? *School Effectiveness and School Improvement*, 10, 233-256.

Snijders, T. and Bosker, R. (1999). *Multilevel Analysis*. Thousand Oaks, CA: Sage.

Som, R. K. (1996) *Practical Sampling Techniques*. Marcel Dekker, Inc. N.Y.

Thum, Y.M. (2002a). Measuring Progress towards a goal: estimating teacher productivity using a multivariate multilevel model for value-added analysis. Milken Family Foundation Report.

Thum, Y.M. (2002b). Measuring Student and School Progress with the California API. Los Angeles, Ca: CRESST Report.

Tourón, J., Repáraz, C., Peralta, F., Gaviria, J. L., Fernández, R., Ramos, J. M. y Reyero, M. (2000) La validación del SCAT (School and College Ability Test) en Navarra. Resultados del estudio piloto En Almeida, L.; Oliveira, E. P. y Melo, A. S. Alunos sobredotados. Contributos para a sua identificação e apoio. ANEIS, Braga: Portugal

Webster, W. J. and Mendro. R. L. (1997). The Dallas value-added accountability

system. In J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp. 81-99 Thousand Oaks, CA:

Willms, D. (1992). *Monitoring School Performance: A Guide for Educators*. Washington, D. C.: Falmer Press.

Willms, D. J. and Raudenbush, S. W. (1989). A longitudinal hierarchical linear models for estimating school effects and their stability. *Journal of Educational Measurement*, 26, 209-232.