

Invarianza de la estructura de covarianzas de las medidas de rendimiento académico en estudios longitudinales en la transición de Educación Primaria a Secundaria

Invariance of the covariance structure for the means of achievement variables in longitudinal studies dealing with the transition period from Primary Education to Secondary Education

José Luis Gaviria Soto, María Chantal Biencinto López y Enrique Navarro Asencio

Universidad Complutense. Facultad de Educación. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). Madrid, España

Resumen

Dados los enormes cambios de carácter tanto psicobiológicos de los estudiantes como en la estructura organizativa y curricular del sistema educativo que experimentan en la transición entre Educación Primaria y Educación Secundaria, en el presente artículo se analiza la posibilidad de mantener la misma estructura longitudinal de base en los estudios de valor añadido con alumnos de Educación Primaria y con los de Educación Secundaria. Para ello, se estudia la estructura de varianza-covarianza de las variables de rendimiento en un estudio longitudinal realizado con 6.709 estudiantes agrupados en 153 centros educativos de la Comunidad de Madrid escolarizados entre 5º de Educación Primaria y 4º de ESO en los cursos académicos 2005-06 y 2006-07. Para el análisis de la invarianza de la estructura de varianza-covarianza del rendimiento en Matemáticas y Comprensión Lectora, se ha empleado, por un lado, el contraste mediante el estadístico M de Box y, por otro, el análisis de un modelo de ecuaciones estructurales de relaciones entre variables observadas, esto es, comenzar con un modelo que establezca la línea base de las relaciones en el tiempo. Los resultados obtenidos apoyan sólo parcialmente

la idea del mantenimiento de dicha estructura en las distintas cohortes o grupos, ya que existe una mayor semejanza en la estructura de Educación Primaria y primer ciclo de Educación Secundaria que entre los dos ciclos de secundaria. Esta variación asociada al tiempo se mantiene en los dos campos disciplinares que se han estudiado, Comprensión Lectora y Matemáticas, aunque cada uno de ellos tenga características diferenciadas.

Palabras clave: valor añadido, rendimiento, invarianza factorial, análisis de ecuaciones estructurales.

Abstract

Given the huge changes that both the students and the structure of the educational system experience over the transition period from Primary Education to Secondary Education, this article studies the hypothesis of invariance of the same basic longitudinal structure on added value studies involving students from both Primary and Secondary Education levels. In order to do this, the covariance structure of achievement variables is studied in a longitudinal study carried out with 6,709 students grouped in 153 educational centres (Primary and Secondary Schools) in the Community of Madrid, from 5th level of Primary Education to 4th level of Secondary Education, in the academic years of 2005-06 and 2006-07. The invariance of the covariance structure of achievement variables in Mathematics and Reading literacy has been tested by means of, on one hand, the statistical M of Box and, on the other, the analysis of a structural equations model of relations between observed variables. Results obtained only partially support the idea of the maintenance of this structure in different cohorts or groups, since there is a stronger similarity between Primary Education and the first stage of Obligatory Secondary Education, than between the two stages of Obligatory Secondary Education. Reading literacy and Mathematics present their own characteristic temporal variation.

Key Words: Added value, achievement, factorial invariance, analysis of structural equations.

Introducción

La transición de los alumnos de la etapa de Educación Primaria a la Educación Secundaria Obligatoria está acompañada por muchos cambios de carácter tanto psicobiológicos de los estudiantes como en la estructura organizativa y curricular del sistema educativo. Esto hace de ese paso un momento crítico que se manifiesta, entre otros indicadores, por un descenso del rendimiento escolar.

En esa misma transición, se produce también un cambio organizativo del proceso docente muy importante. Los alumnos pasan de tener un solo maestro que les acompaña a lo largo de casi toda la jornada escolar a tener a un grupo de distintos especialistas que se van turnando en la tarea de dar clases.

Tampoco es indiferente el hecho de que, conforme avanzan los alumnos en el sistema educativo, la complejidad de las materias estudiadas, a pesar de compartir título, aumenta significativamente. Ese aumento de complejidad curricular puede venir, y probablemente viene, acompañado de una mayor complejidad estructural de los factores latentes (Lizasoain y Joaristi, 2009).

Como consecuencia de ese aumento de la complejidad curricular, los modelos que sólo incorporan un factor como causa inmediata de las intercorrelaciones presentan en el ámbito de lo observado patrones correlacionales que difieren de los esperables en el nivel latente de las variables estructuradas (Blanco, González y Ordóñez, 2009).

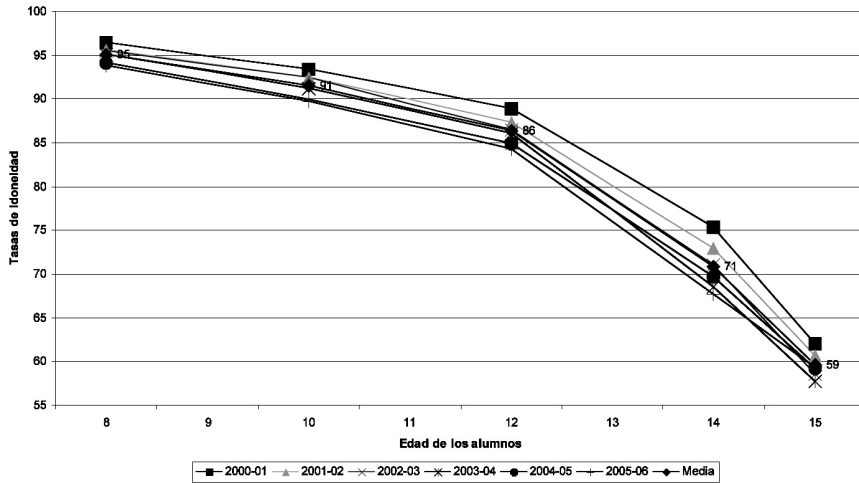
Por otra parte, la construcción de variables escaladas en una métrica común exige algunos diseños metodológicos y estructuras de las pruebas que faciliten la posterior equiparación y comparabilidad de los resultados.

Uno de estos indicadores sensibles a los problemas transicionales es la tasa de idoneidad. Su evolución refleja claramente la situación referida. La tasa de idoneidad se define como el porcentaje de alumnos que se encuentra matriculado en el curso que corresponde a su cohorte de edad (Instituto de Evaluación, 2006). Según las últimas estadísticas de la Educación en España disponibles (MEC, 2006), la tasa de idoneidad media para toda España a los 12 años se sitúa en un 84,3%. Los alumnos de la Comunidad de Madrid se sitúan ligeramente por encima de la media con una tasa bruta de 84,9%.

Sin embargo, la evolución de las tasas de idoneidad a partir de los 12 años muestra un importante descenso. El Gráfico I refleja la evolución de las tasas de idoneidad en cinco edades clave (desde los 8 a los 15 años) en la Comunidad de Madrid, que se corresponde con el momento de transición entre la primaria y la secundaria.

La investigación que se presenta en este artículo utiliza datos provenientes de un estudio longitudinal en el que tres cohortes de alumnos han sido medidos en cuatro ocasiones consecutivas, al inicio y al final de dos cursos académicos, en las dos materias instrumentales básicas de Matemáticas y Comprensión Lectora. El objetivo es tratar de comprobar si la estructura de varianzas covarianzas y de correlaciones entre las medidas de rendimiento es invariante en las tres cohortes analizadas.

GRÁFICO I. Tasas de idoneidad en la Comunidad de Madrid de 2001-02 a 2005-06



Fuente: elaboración propia a partir de las series de las estadísticas de enseñanza no universitaria del MEC

Estudio empírico

Objetivo del estudio

Este estudio pretende dar una respuesta a las siguientes preguntas:

- ¿La matriz de varianzas-covarianzas formada por las cuatro variables de respuesta del estudio longitudinal se mantiene estable entre las tres cohortes definidas, tanto para Comprensión Lectora como para Matemáticas?
- Si no es invariante, ¿entre qué grupos existe una mayor diferencia?
- Como consecuencia de lo anterior, ¿es posible utilizar el mismo modelo de valor añadido para primaria y secundaria o, por el contrario, se necesitan modelos diferentes o, incluso, un modelo distinto para cada cohorte?

Muestra y población. Estructura de las cohortes

La población objeto de estudio fue el conjunto de alumnos (tanto de centros públicos, como privados y privados concertados) de la Comunidad de Madrid escolarizados en el

último ciclo de Educación Primaria y en Educación Secundaria Obligatoria, que asciende a un total de 343.746 repartidos en 1.798 escuelas en el curso académico 2005-06. El dominio específico de este estudio es la Comunidad Autónoma de Madrid, constituida por toda la red de centros públicos, privados concertados y privados, que componen subdominios específicos dentro de la misma. En este dominio, se definieron cinco estratos correspondientes con las cinco áreas territoriales (Capital, Norte, Sur, Este y Oeste) de la Comunidad.

Para el cálculo inicial del tamaño necesario en cada una de las cohortes, se utilizó un muestreo aleatorio simple utilizando el tamaño de la población de referencia, un nivel de confianza del 95% y un error muestral máximo de $0,1\sigma$. La segunda parte del proceso de consistió en un muestreo por conglomerados, éste tiene un efecto negativo, es decir, el tamaño de la muestra aumentará, con respecto al muestreo aleatorio simple, en función del tamaño del conglomerado y de la autocorrelación.

Para llevar a cabo esta corrección, se tomaron los valores de 25 como tamaño medio del conglomerado y una correlación intraclase de 0,2 (Kiplinger, 2004; Zvoch y Stevens, 2003; Navarro y Redondo, 2007). Finalmente, se introdujo la modificación debida a la mortalidad experimental que se produce en este tipo de estudios longitudinales.

La muestra final está compuesta por 174 centros repartidos en tres grados, seleccionados con probabilidad proporcional a su tamaño y estratificada por la titularidad y el área territorial. El total de alumnos implicados llegó hasta 13.000, medidos en cuatro ocasiones a lo largo de los cursos académicos 2005-06 y 2006-07.

Para llevar a cabo este estudio, se han utilizado sólo los datos de los sujetos con puntuaciones en las cuatro aplicaciones; es decir, los de aquéllos de los que se tenía registro completo porque participaron en las cuatro evaluaciones llevadas a cabo. La muestra definitiva queda configurada como se muestra en la Tabla I.

TABLA I. Configuración definitiva de las muestras

	MATEMÁTICAS		COMPRESIÓN LECTORA	
	ALUMNOS	CENTROS	ALUMNOS	CENTROS
Cohorte 1	2751	92	2739	109
Cohorte 2	2152	61	2128	65
Cohorte 3	1806	59	1887	

Características del diseño

El diseño longitudinal para esta investigación supone la incorporación de cuatro ocasiones de medida a lo largo de dos cursos académicos. De esta forma, se mide a los

alumnos al inicio y al final del curso académico, durante dos años consecutivos (2005-06 y 2006-07). Se definieron tres segmentos educativos de interés: el último ciclo de Educación Primaria y los dos ciclos de Educación Secundaria Obligatoria.

Cada uno de estos segmentos educativos constituye una cohorte. Para los alumnos que en 2005-06 comenzaron en 5º de primaria (cohorte 1) se han recogido datos en octubre de 2005 (primera medición) y junio de 2006 (segunda medición). En el curso 2006-07, cuando estos alumnos se encontraban en 6º de primaria, se recogió información en noviembre de 2006 (tercera medición) y junio de 2007 (cuarta medición). Con las dos cohortes restantes correspondientes al primer ciclo de secundaria (cohorte 2) y segundo ciclo de secundaria (cohorte 3), se ha procedido de manera análoga. La Tabla II muestra la estructura de la recogida de datos longitudinal asociada a esta investigación.

TABLA II. Estructura de recogida longitudinal de datos

Cohorte	Curso	2005-06									2006-07											
		2005			2006						2006			2007								
1	5º EP	Oct	Nov	Dic	Ene	Feb	Marz	Abr	May	Jun												
	6º EP										Jul	Agos	Sept	Oct	Nov	Dic	Ene	Feb	Marz	Abr	May	Jun
2	1ºESO	Oct	Nov	Dic	Ene	Feb	Marz	Abr	May	Jun												
	2ºESO										Jul	Agos	Sept	Oct	Nov	Dic	Ene	Feb	Marz	Abr	May	Jun
3	3ºESO	Oct	Nov	Dic	Ene	Feb	Marz	Abr	May	Jun												
	4ºESO										Jul	Agos	Sept	Oct	Nov	Dic	Ene	Feb	Marz	Abr	May	Jun

Este tipo de diseño de recogida de información longitudinal nos permite obtener medidas del rendimiento de los alumnos al final de tres de las etapas clave del Sistema Educativo Español (final de la Educación Primaria y de los dos ciclos de Educación Secundaria Obligatoria).

Al mismo tiempo se evitó el grave problema que para el diseño de investigación supone que una cohorte cambie de ciclo durante el período de la recogida de datos. En efecto, los alumnos que comenzaron en 5º de Educación Primaria se encontraban al final de 6º de Educación Primaria en el mes de junio de 2007 cuando se recogieron los últimos datos de cada cohorte. Si se hubiese comenzado tomando datos a alumnos de 6º de Educación Primaria, en el segundo año del estudio estos alumnos habrían cambiado de etapa y, en muchos casos, de centro. La consecuencia de esto último hubiera sido una gran pérdida de datos longitudinales, dado que muchos alumnos habrían cambiado a otros centros no incluidos en la muestra estudiada.

Esta ventaja señalada implica también una importante limitación para este tipo de estudios, ya que, aunque tenemos datos de una cohorte de Educación Primaria y dos

de Educación Secundaria, no disponemos de datos de una cohorte en la que se incluya la transición de una etapa a la otra. No se escapará al lector la dificultad de recabar este tipo de información.

Esto hace todavía más importante la necesidad de comprobar la constancia de la estructura de los datos analizados en las etapas del Sistema Educativo.

Procedimientos y modelos de análisis

Para contrastar las hipótesis planteadas se utilizaron dos procedimientos. En el primero, se utilizó la prueba M de Box para comprobar la igualdad de las matrices de covarianzas entre cohortes. El estadístico M puede aproximarse por una distribución F de Snedecor si se cumplen los supuestos del mismo, el más importante de los cuales es el de normalidad de las distribuciones implicadas. El cumplimiento de este supuesto se estudió mediante las pruebas de Kolmogorov-Smirnov y Shapiro y Wilks.

El segundo procedimiento consistió en el contraste de un modelo de ecuaciones estructurales multigrupo.

La estructura longitudinal del estudio impone analizar el modelo compuesto por las cuatro variables extraídas de las puntuaciones obtenidas por los alumnos en las cuatro aplicaciones, tanto en Comprensión Lectora como en Matemáticas. El estudio de la variación de esta estructura entre cohortes permite comprobar si el cambio que se refleja en algunos indicadores en los alumnos en la transición de primaria a secundaria supone también variaciones en los patrones de correlaciones entre las medidas consecutivas tan importante que impliquen modelos de valor añadido distintos.

La prueba M de Box (Box, 1949) contrasta la hipótesis nula de igualdad de las matrices de varianzas-covarianzas poblacionales.

El estadístico M de Box toma la siguiente forma:

$$M = (n - g) \log |S| - \sum_{j=1}^g (n_j - 1) \log |S_j| \quad (1)$$

Donde S es la matriz de varianzas-covarianzas combinada y S_j la de cada grupo j , n es el número total de casos, n_j hace referencia al número de casos de cada grupo j y g es el número de grupos. El estadístico M no tiene distribución muestral conocida; no obstante, puede aproximarse por un estadístico con distribución F de Snedecor. Por tanto, un valor de dicha F significativo indicará que la estructura de covarianzas no es invariante entre cohortes.

El test M de Box es sensible a la falta de normalidad multivariante, es decir, matrices iguales pueden aparecer como significativamente diferentes si no existe normalidad (Pardo y Ruiz, 2002).

Esta especial sensibilidad a las violaciones de los supuestos de normalidad y dado el importante tamaño de la muestra con que se contaba, que hacía más probable el rechazo de la hipótesis de nulidad, hizo aconsejable complementar el estudio con otra estrategia de análisis algo más robusta y con índices de ajuste más adecuados a la naturaleza de los datos manejados.

Esta estrategia es el análisis de modelos de ecuaciones estructurales multigrupo (Arbuckle, 2006). Este tipo de análisis permite imponer restricciones sucesivas al modelo planteado; en nuestro caso, como muestra la Figura III, sólo consta de varianzas y covarianzas entre puntuaciones; por tanto, la única restricción es la fijación de dicha matriz.

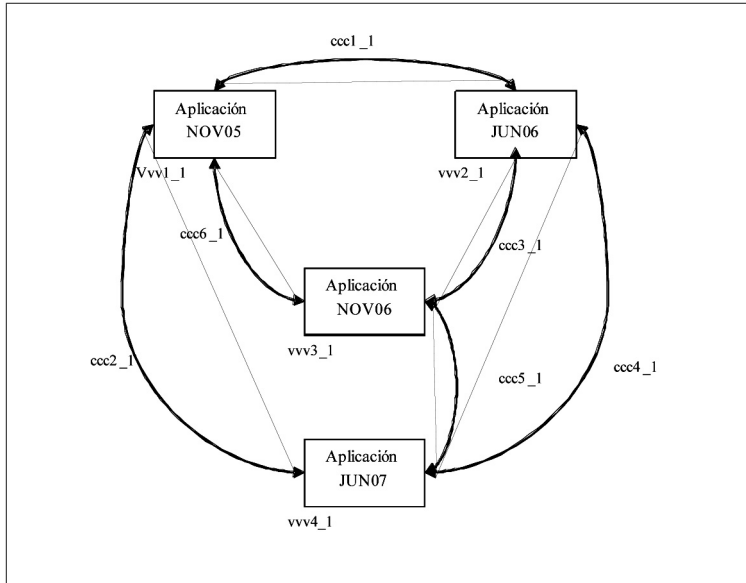
Es importante señalar en este punto que nuestro objetivo no es probar un modelo de rendimiento académico, ya que éste sería mucho más complejo e implicaría la introducción de distintos predictores con el objetivo de maximizar la varianza explicada. Como ya se ha señalado anteriormente, un estudio longitudinal supone el comenzar con un modelo que establece la línea base, es decir, cómo es la relación de las medidas dependientes con el tiempo o, dicho de otro modo, de qué manera se produce la sucesión de las variables dependientes cuyo progreso habrá que explicar con los predictores adecuados.

El modelo sometido a prueba por tanto es el que aparece en la Figura I, con la restricción de que las correlaciones permanecen invariantes en las tres cohortes.

Para el contraste de este modelo, se utilizó el programa AMOS 7.0 (Arbuckle, 2006). Para llevar a cabo el proceso de análisis, se han comparado diferentes modelos que varían en las restricciones impuestas. El modelo saturado asume que todos los parámetros posibles son significativamente distintos de cero; se trata de un modelo de ajuste perfecto y que, por tanto, refleja completamente toda la información de los datos. En el otro extremo, se encuentra el modelo fijo; éste asume que todos los parámetros son iguales entre cohortes y, por tanto, un buen ajuste de este modelo indicaría la semejanza entre los diferentes grupos estudiados. Entre estos dos modelos extremos, se sitúan otros que permiten establecer comparaciones entre pares de cohortes.

Todas las estimaciones se han realizado por el procedimiento de máxima verosimilitud, tal y como está implementado en el mencionado programa AMOS. Para evaluar el ajuste del modelo, se han utilizado, aparte del estadístico de bondad de ajuste χ^2 , otros índices que juzgan la adecuación del modelo (Bollen y Long, 1993), ya que, en ocasiones, valores altos de χ^2 pueden producirse por el incumplimiento de los

FIGURA I. Modelo de contraste



supuestos de normalidad (Batista y Coenders, 2000; Baldwin y Caldwell, 2003; Schumacker y Lomax, 2004). Se ha utilizado para tal caso el índice de bondad de ajuste GFI (Jöreskog y Sörbom, 1989) y su ajustado AGFI. Valores cercanos a uno en estos índices suponen un buen ajuste. Por encima de 0,90 son aceptables. También se ha analizado la raíz media cuadrática del error de estimación (RMSEA) (Browne y Cudeck, 1993) y su intervalo de confianza. Valores inferiores a 0,05 señalan un correcto ajuste de los datos. Algunos autores consideran que valores entre 0,05 y 0,08 también son aceptables (Hair, Anderson, Tatham y Black, 1999).

Resultados

Comenzamos presentando los resultados correspondientes a Comprensión Lectora y seguiremos con los de Matemáticas. Se presentan, en primer lugar, los resultados de la prueba M de Box, seguidamente la prueba de normalidad de las variables de respuesta y, por último, los del análisis multigrupo.

Resultados en Comprensión Lectora

En la Tabla III, podemos ver la matriz de varianzas-covarianzas entre las medidas de Comprensión Lectora. En dicha matriz, podemos observar cómo los patrones que se producen en una cohorte se asemejan mucho a los que se producen en cualquiera de las otras dos.

TABLA III. Matriz de Varianzas - Covarianzas

Cohorte 1		CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05	,669	,385	,394	,372
CLJUN06		,657	,401	,383	
CLNOV06			,691	,415	
CLJUN07				,695	
Cohorte 2		CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05	,701	,389	,406	,373
	CLJUN06		,734	,475	,409
	CLNOV06			,837	,465
CLJUN07				,752	
Cohorte 3		CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05	,642	,323	,292	,321
	CLJUN06		,638	,328	,329
	CLNOV06			,527	,358
CLJUN07				,731	

Si nos fijamos en las diferencias en las varianzas y covarianzas de las distintas aplicaciones, vemos que los patrones en ningún caso reflejan una mayor distancia entre la cohorte de primaria y secundaria que entre los dos ciclos de secundaria. Como se puede observar en la Tabla IV, existe mayor diferencia entre el primer ciclo de secundaria y el segundo, es decir, entre la cohorte 2 y 3, que entre la cohorte de primaria y los dos ciclos de secundaria. Por ejemplo, la diferencia en la covarianza de la puntuación CLNOV05 y CLJUN06 es de 0,004 entre la cohorte de primaria y la de primer ciclo de secundaria; en cambio, esta misma diferencia de covarianzas tienen un valor de 0,066 entre los dos ciclos de secundaria. En cualquier caso es evidente que se trata de mínimas diferencias.

Por otra parte, observando la matriz de correlaciones (Tabla V) podemos comprobar la intensa relación existente entre las aplicaciones en cada una de las tres cohortes. Prácticamente la totalidad de las correlaciones están por encima de 0,50 y resultan significativas al 99%.

Como cabía esperar, las correlaciones entre dos aplicaciones sucesivas presentan siempre el mayor valor. Conforme transcurre más tiempo entre aplicaciones el valor de la correlación se amortigua. Este fenómeno se produce de igual modo en las tres

TABLA IV. Diferencias en las varianzas y covarianzas entre cohortes

Diferencia 1-2		CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05	-0,032	-0,004	-0,012	-0,001
CLJUN06		-0,077	-0,074	-0,026	
CLNOV06			-0,146	-0,05	
CLJUN07				-0,057	
Diferencia 1-3		CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05	0,027	0,062	0,102	0,051
	CLJUN06		0,019	0,073	0,054
	CLNOV06			0,164	0,057
CLJUN07				-0,036	
Diferencia 2-3	2	CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05	0,059	0,066	0,114	0,052
	CLJUN06		0,096	0,147	0,08
	CLNOV06			0,31	0,107
	CLJUN07				0,021

cohortes analizadas. Es interesante comprobar que los valores absolutos de las correlaciones disminuyen proporcionalmente a la edad media de la cohorte. En efecto, las correlaciones de la cohorte 3 son en términos absolutos más bajas que las de la 2 y las de la 1. Pero en los tres casos se reproduce un patrón similar de amortiguación.

TABLA V. Matriz de Correlaciones

Cohorte 1		CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05		,581 ^(***)	,579 ^(***)	,546 ^(**)
CLJUN06			,594 ^(***)	,566 ^(**)	
CLNOV06				,598 ^(**)	
CLJUN07					
Cohorte 2		CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05		,542 ^(***)	,530 ^(***)	,513 ^(**)
	CLJUN06			,606 ^(***)	,551 ^(**)
	CLNOV06				,586 ^(**)
CLJUN07					
Cohorte 3		CLNOV05	CLJUN06	CLNOV06	CLJUN07
	CLNOV05		,505 ^(***)	,502 ^(***)	,468 ^(**)
	CLJUN06			,566 ^(***)	,481 ^(**)
	CLNOV06				,577 ^(**)
CLJUN07					

** La correlación es significativa al nivel 0,01 (bilateral).

En la Tabla VI, aparecen los resultados de la prueba M de Box. Como ya hemos comentado en párrafos anteriores, el estadístico de Box contrasta la hipótesis nula de igualdad entre las matrices de covarianzas poblacionales.

Los resultados de la prueba M de Box reflejan, inicialmente, diferencias significativas entre las matrices de covarianzas de las diferentes cohortes. Dado el tamaño muestral con el que se ha realizado el estudio, resulta interesante comprobar hasta qué punto distintas muestras aleatorias con tamaño menor reflejan esos mismo resultados. Para ilustrar esto se seleccionaron submuestras aleatorias de 3.000, 1.000, 300 y 100 sujetos.

Con una selección aleatoria de 100 sujetos, la prueba deja de mostrar diferencias al nivel de 0,01 bajando la significatividad hasta 0,045. Esto es debido a la sensibilidad de esta prueba al tamaño de la muestra. En la Tabla VI, podemos ver la evolución del estadístico y su probabilidad asociada en función de la reducción muestral llevada a cabo.

TABLA VI. Resultados de la prueba M de Box

M de Box	Muestra completa (6754 sujetos)	3000 sujetos	1000 sujetos	300 sujetos	100 sujetos
	186,029	120,644	45,229	42,118	34,067
F Aprox.	9,293	6,020	2,248	2,063	1,594
Sig.	,000	,000	,001	,003	,045

A la vista de estos resultados, cabe preguntarse si la violación de los supuestos de la prueba puede influir en los valores obtenidos. La prueba de Kolmogorov-Smirnov y la de Shapiro y Wilk nos indican hasta qué punto nuestros datos son consistentes con la hipótesis de normalidad multivariante. En el caso de Comprensión Lectora, las cuatro variables carecen de normalidad en cada una de las tres cohortes estudiadas.

La segunda parte del proceso de análisis consiste, como ya hemos comentado en el epígrafe anterior, en poner a prueba el modelo de interrelación de las matrices de covarianzas entre las cohortes mediante un análisis multigrupo en el contexto de análisis de ecuaciones estructurales. El resumen de los resultados más relevantes se presenta en la Tabla IX. Los modelos comparados en dicha tabla son los siguientes:

TABLA VIII. Modelos de comparación entre cohortes

MODELO	DEFINICIÓN	PARÁMETROS ESTIMADOS
Saturado	Todos los parámetros libres	4x3 varianzas + 6x3 covarianzas= 30
1	Con covarianzas fijas entre cohortes 1 y 2	4x3 varianzas + 6x2 covarianzas (las de las cohortes 1 y 3; recordemos que las de la cohorte 2 son iguales a las de la 1)=24
2	Con covarianzas fijas entre cohortes 1 y 3	4x3 varianzas + 6x2 covarianzas (las de las cohortes 1 y 2; recordemos que las de la cohorte 3 son iguales a las de la 1)=24
3	Con covarianzas fijas entre cohortes 2 y 3	4x3 varianzas + 6x2 covarianzas (las de las cohortes 1 y 2; recordemos que las de la cohorte 3 son iguales a las de la 2)=24
Fijo	Todos los parámetros fijos	4 varianzas + 6 covarianzas=10

Los valores de χ^2 correspondientes a estos modelos no indican un buen ajuste de los mismos, siendo valores altos debido muy probablemente a la no-normalidad de las variables y al gran tamaño de la muestra. En cambio, el resto de índices muestran unos valores que indican en términos generales que estamos trabajando con modelos muy plausibles. El modelo con todos los parámetros fijos, es decir, el que nos mostraría la igualdad en la matriz de varianzas-covarianza, con valores de GFI y AGFI de 0,987 y 0,980, respectivamente, y con un valor de RMSEA de 0,035, también su intervalo de confianza se sitúa por debajo de 0,05, señala en esa dirección. Podemos concluir, por tanto, que la estructura de las variables de respuesta en comprensión lectora es semejante en primaria y secundaria.

Si nos centramos ahora en las comparaciones por pares de cohortes (Tabla IX), podemos destacar la mayor similitud existente entre primaria y primer ciclo de secundaria (Modelo 1), con valores de ajuste muy buenos, que entre los dos ciclos de Educación Secundaria (Modelo 3).

Por tanto, la calidad y ajuste de los diferentes modelos son muy satisfactorios, como se puede comprobar en la Tabla IX.

TABLA IX. Resumen de la estimación de los modelos de comparación entre cohortes para Comprensión Lectora

Modelos	NPAR	CMIN	DF	P	CMIN/DF	GFI	AGFI	RMSEA	LO 90	HI 90
Modelo saturado	30	,000	0			1,000				
Modelo 1	24	17,699	6	,007	2,950	,999	,993	,017	,008	,026
Modelo 2	24	25,874	6	,000	4,312	,998	,991	,022	,014	,031
Modelo 3	24	42,332	6	,000	7,055	,997	,984	,030	,022	,039
Modelo fijo	10	185,996	20	,000	9,300	,987	,980	,035	,031	,040

Estos resultados nos llevan a pensar que la posible discontinuidad se produce no tanto en el inicio de la Educación Secundaria, sino más bien al avanzar en esta última. Es posible que las causas de este salto se presenten desde un principio en el cambio entre Educación Primaria y Educación Secundaria, pero sus efectos se hacen especialmente visibles al acercarse el último tramo de la Educación Secundaria.

Como consecuencia de este resultado, podemos preguntarnos si el cambio implica una variación en la importancia que la primera aplicación de cada cohorte tiene en la predicción de los rendimientos posteriores. Para tratar de responder a esta cuestión, se sometieron a contraste tres modelos para cada materia. En ellos, se comparan los ajustes cuando las correlaciones de la primera aplicación con las demás se mantienen constantes en las sucesivas cohortes, en un patrón de comparaciones similar al anterior.

Para llevar a cabo este paso, se ha seguido el mismo procedimiento: primero la estimación del modelo saturado y del modelo fijo (en este caso, es el que mantiene fijas las covarianzas sólo de la primera aplicación con el resto entre cohortes) y, a continuación, la estimación de diferentes modelos con restricciones entre pares de cohortes. Los modelos estimados son los siguientes:

TABLA X. Modelos

MODELO	DEFINICIÓN	PARÁMETROS ESTIMADOS
Saturado	Todos los parámetros libres	4x3 varianzas + 6x3 covarianzas= 30
1	Con las covarianzas de la primera aplicación con el resto fijas, entre la cohortes 1 y 2	4x3 varianzas + 3x3 covarianzas (son las cov. entre las 3 últimas aplicaciones en las tres cohortes) + 3x2 covarianzas (las de la primera aplicación con el resto en la cohorte 1 y 3; recordemos que las de la cohorte 2 son las mismas que las de la 1, las fijas)=27
2	Con las covarianzas de la primera aplicación con el resto fijas, entre la cohortes 1 y 3	4x3 varianzas + 3x3 covarianzas (son las cov. entre las 3 últimas aplicaciones en las tres cohortes) + 3x2 covarianzas (las de la primera aplicación con el resto en la cohorte 1 y 2; recordemos que las de la cohorte 3 son las mismas que las de la 1, las fijas)=27
3	Con las covarianzas de la primera aplicación con el resto fijas, entre la cohortes 2 y 3	4x3 varianzas + 3x3 covarianzas (son las cov. entre las 3 últimas aplicaciones en las tres cohortes) + 3x2 covarianzas (las de la primera aplicación con el resto en la cohorte 1 y 2; recordemos que las de la cohorte 3 son las mismas que las de la 2, las fijas)=27
Fijo	Con las covarianzas de la primera aplicación con el resto fijas	4x3 varianzas + 3x3 covarianzas (son las cov. entre las 3 últimas aplicaciones en las tres cohortes) + 3 covarianzas (las de la primera aplicación con el resto en una cohorte) =24

El modelo fijo muestra unos buenos índices de ajuste (Tabla XI) con un valor de GFI de 0,999 y un RMSEA de 0,017. Este modelo sólo es superado por el que mantiene fijas las covarianzas de la primera aplicación entre las cohortes 1 y 2, con un ajuste casi perfecto.

TABLA XI. Importancia de la primera aplicación como predictor (Comprensión Lectora)

Modelos	NPAR	CMIN	DF	P	CMIN/DF	GFI	AGFI	RMSEA	LO 90	HI 90
Modelo saturado	30	,000	0			1,000				
Modelo 1	27	,465	3	,927	,155	1,000	1,000	,000	,000	,006
Modelo 2	27	24,293	3	,000	8,098	,998	,982	,032	,021	,045
Modelo 3	27	24,960	3	,000	8,320	,998	,982	,033	,022	,045
Modelo fijo	24	17,699	6	,007	2,950	,999	,993	,017	,008	,026

Podemos afirmar que la primera aplicación tiene la misma importancia como predictor en las tres cohortes, encontrando mayores diferencias entre los dos ciclos de secundaria que entre cualquiera de éstos con la cohorte de primaria.

Resultados en Matemáticas

La Tabla XII permite observar las varianzas y covarianzas de las variables de respuesta en las tres cohortes.

TABLA XII. Matriz de Varianzas - Covarianzas

Cohorte 1		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	0,776	0,547	0,358	0,337
	MATJUN06		0,811	0,412	0,391
	MATNOV06			0,358	0,261
	MATJUN07				0,331
Cohorte 2		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	0,733	0,524	0,392	0,347
	MATJUN06		0,739	0,42	0,388
	MATNOV06			0,424	0,306
	MATJUN07				0,406
Cohorte 3		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	0,735	0,418	0,322	0,292
	MATJUN06		0,802	0,441	0,381
	MATNOV06			0,534	0,326
	MATJUN07				0,421

De la misma forma que ocurría en Comprensión Lectora, las diferencias brutas reflejan una mayor discrepancia entre las covarianzas de los dos ciclos de secundaria que entre las cohortes de primaria y primer ciclo de secundaria (tabla 13). Podemos tomar como ejemplo el valor de la diferencia entre la covarianza de la primera aplicación (MATNOV05) y la última toma de datos (MATJUN07). Esta última tiene un valor absoluto de 0,01 entre la cohorte de primaria y primer ciclo de secundaria; llega a la cantidad de 0,045 entre la muestra de primaria y la de segundo ciclo de secundaria para finalmente ascender a 0,055 entre las dos cohortes de secundaria (Tabla XIII).

TABLA XIII. Diferencias en las varianzas y covarianzas entre cohortes

Diferencia 1-2		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	0,043	0,023	-0,034	-0,01
	MATJUN06		0,072	-0,008	0,003
	MATNOV06			-0,066	-0,045
	MATJUN07				-0,075
Diferencia 1-3		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	0,041	0,129	0,036	0,045
	MATJUN06		0,009	-0,029	0,01
	MATNOV06			-0,176	-0,065
	MATJUN07				-0,09
Diferencia 2-3		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	-0,002	0,106	0,07	0,055
	MATJUN06		-0,063	-0,021	0,007
	MATNOV06			-0,11	-0,02
	MATJUN07				-0,015

Los valores de las correlaciones entre las puntuaciones de las cuatro aplicaciones en Matemáticas son superiores, en las tres cohortes, a los obtenidos en la materia de Comprensión Lectora. Recordemos que en Comprensión Lectora las correlaciones variaban en torno a 0,5 y, en este caso, en las dos primeras cohortes llegan a 0,765 entre la segunda y tercera aplicación (Tabla XIV). En la tercera cohorte, los valores son ligeramente más bajos; no obstante, son correlaciones significativas.

TABLA XIV. Matriz de Correlaciones

Cohorte 1		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	1	,689 ^(***)	,679 ^(***)	,665 ^(***)
MATJUN06			1	,765 ^(***)	,754 ^(***)
MATNOV06				1	,758 ^(***)
MATJUN07					1
Cohorte 2		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	1	,712 ^(***)	,704 ^(***)	,637 ^(***)
MATJUN06			1	,750 ^(***)	,709 ^(***)
MATNOV06				1	,737 ^(***)
MATJUN07					1
Cohorte 3		MATNOV05	MATJUN06	MATNOV06	MATJUN07
	MATNOV05	1	,544 ^(***)	,514 ^(***)	,525 ^(***)
MATJUN06			1	,674 ^(***)	,656 ^(***)
MATNOV06				1	,688 ^(***)
MATJUN07					1

** La correlación es significativa al nivel 0,01 (bilateral)

La prueba M de Box muestra cómo, de forma similar a la materia de Comprensión Lectora, la matriz de varianzas-covarianzas no es invariante entre las cohortes cuando se utiliza toda la muestra. No obstante, con la progresiva reducción muestral, disminuye de modo muy importante el valor de F (ver Tabla XV).

TABLA XV. Resultados de la prueba M de Box

M de Box	Muestra completa (6689 sujetos)	3000 sujetos	1000 sujetos	300 sujetos	100 sujetos
	668,962	307,757	152,349	86,549	36,821
F Aprox.	33,418	15,357	7,572	4,240	1,728
Sig.	,000	,000	,000	,000	,023

A la luz de estos resultados, es conveniente analizar, de la misma forma que hicimos en Comprensión Lectora, la normalidad de las cuatro variables que componen la estructura longitudinal del estudio. Los datos de las dos pruebas realizadas reflejan la falta de normalidad en la distribución de las variables. No obstante, podemos excluir

de esta tendencia a la variable correspondiente a la tercera aplicación (MATNOV06) en la cohorte de primer ciclo de Educación Secundaria Obligatoria (cohorte 2).

En la segunda fase del análisis, tenemos los resultados obtenidos a través del modelo de ecuaciones estructurales en su versión multigrupo. El resumen de la información podemos encontrarlo en la Tabla XVI. No es necesario volver a definir cada modelo porque son iguales a los estimados con Comprensión Lectora.

Los índices de ajuste empeoran ligeramente si los comparamos con los obtenidos en Comprensión Lectora. Los valores de GFI y AGFI son de 0,949 y 0,924, respectivamente, para el modelo fijo; en cambio, el valor de RMSEA con un valor de 0,07 no refleja un buen ajuste.

De la misma forma que ocurría en Comprensión Lectora, el modelo que mejor ajusta es aquél que mantiene fijas las covarianzas entre la primera y segunda cohorte, es decir, entre Educación Primaria y primer ciclo de Educación Secundaria Obligatoria. Las cohortes que muestran mayores diferencias son la uno y la tres como indican los resultados del modelo 2.

TABLA XVI. Resumen de la estimación de los modelos de comparación entre cohortes para Matemáticas

Modelos	NPAR	CMIN	DF	P	CMIN/DF	GFI	AGFI	RMSEA	LO 90	HI 90
Modelo saturado	30	,000	0			1,000				
Modelo 1	24	48,168	6	,000	8,028	,996	,982	,032	,024	,041
Modelo 2	24	106,408	6	,000	17,735	,992	,960	,050	,042	,059
Modelo 3	24	52,715	6	,000	8,786	,996	,980	,034	,026	,043
Modelo fijo	10	668,673	20	,000	33,434	,949	,924	,070	,065	,074

Para el análisis de la importancia de la primera aplicación como predictor en la materia de Matemáticas, los modelos estimados muestran ajustes muy buenos. En este caso, el modelo 3, es decir, el que mantiene fijas las covarianzas entre las cohortes 2 y 3 (los dos ciclos de Educación Secundaria) es el que obtiene mejores índices de ajuste con un valor de GFI igual a 0,999, AGFI 0,984 y RMSEA 0,023.

TABLA XVII. Importancia de la primera aplicación como predictor (Matemáticas)

Modelos	NPAR	CMIN	DF	P	CMIN/DF	GFI	AGFI	RMSEA	LO 90	HI 90
Modelo saturado	30	,000	0			1,000				
Modelo 1	27	21,126	3	,000	7,042	,998	,984	,030	,019	,043
Modelo 2	27	29,393	3	,000	9,798	,998	,979	,036	,025	,049
Modelo 3	27	13,994	3	,003	4,665	,999	,990	,023	,012	,036
Modelo fijo	24	44,354	6	,000	7,392	,997	,984	,031	,023	,040

⁽¹⁾ Recordemos que un valor no significativo de F sería el que nos llevaría a la conclusión de que las matrices de varianzas-covarianzas son invariantes.

Para finalizar, en el caso de Matemáticas, debemos tener precaución a la hora de afirmar la igualdad en la matriz de covarianzas ya que uno de los índices de ajuste analizados (RMSEA=0,07) sobrepasa ligeramente el límite para resultar aceptable.

Discusión y conclusiones

Como hemos mencionado al comienzo de este artículo, existen varias causas que pueden explicar la variación en los patrones de correlaciones entre las variables de respuesta en estudios longitudinales como el que aborda esta investigación: cambios psicológicos, cambios organizativos del Sistema Educativo, aumento de la complejidad curricular e incluso artefactos asociados a la construcción de la escala común.

Una vez analizados los distintos modelos hipotetizados, parece evidente una mayor continuidad entre los datos provenientes de las cohortes 1ª y 2ª, es decir, entre Educación Primaria y primer ciclo de Educación Secundaria, que entre cualquiera de éstos y la cohorte 3ª.

Otro resultado relevante es la importancia como predictor de la primera aplicación, ya que es precisamente la correlación de esta variable con las de las aplicaciones sucesivas lo que se mantiene constante en los modelos que mejor ajustan.

Esta información nos lleva a pensar que la construcción de un modelo estadístico para describir y explicar el crecimiento de los centros escolares que permita seguir el progreso de unidades múltiples y anidadas puede ser el mismo para Primaria y primer ciclo de Secundaria. En cambio, para los dos ciclos de Educación Secundaria Obligatoria sería conveniente plantearse modelos diferenciados de crecimiento.

Asimismo y, enlazando con las características del contexto personal y académico de los alumnos objeto de estudio, podemos concluir que el cambio o transición de la Educación Primaria a la Educación Secundaria Obligatoria supone una variación significativa en las relaciones entre las variables dependientes medidas en ocasiones consecutivas, específicamente en el caso de la cohorte 3ª (segundo ciclo de Educación Secundaria), aunque este cambio no se manifiesta en los momentos inmediatos a la transición, sino que se hace visible cuando los alumnos encaran la parte final de la Educación Secundaria Obligatoria.

En consecuencia, cualquier estudio o evaluación que suponga una toma de datos longitudinal en los grupos de edad en los que se produce la transición entre la Educación Primaria y la Educación Secundaria debe tener muy en cuenta la posibilidad muy

sería de que la estructura correlacional entre las variables de respuesta experimente cambios de tal naturaleza que obliguen a plantearse modelos diferenciados para las cohortes en función de su proximidad al final del período obligatorio.

Esta variación asociada al tiempo se mantiene en los dos campos disciplinares que se han estudiado, Comprensión Lectora y Matemáticas, aunque cada uno de ellos tenga sus notas idiosincráticas.

Por último, y en cualquier caso, se mantiene la enorme importancia que las correlaciones de la primera aplicación con el resto tiene en cualquiera que sea el modelo y la cohorte analizados.

Referencias bibliográficas

- ARBUCKLE, J. L. (2006). *Amos 7.0 user`s guide*. Chicago: SPSS, Inc.
- BAGOZZI, R. P. & YOUJAE, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16 (spring), 74-94.
- BALDWIN, C. K. & CADWELL, L. L. (2003). Development of the Free Time Motivation Scale for adolescents. *Journal of Leisure Research*, 35 (2), 129-151.
- BATISTA, J. M. Y COENDERS, G. (2000). *Modelos de educaciones estructurales*. Madrid: La Muralla.
- BLANCO, A., GONZÁLEZ, C. Y ORDÓÑEZ, X. (2009). Complejidad del rasgo evaluado y predicción del rendimiento escolar en diseños longitudinales: un estudio de simulación desde un enfoque multinivel. *Revista de Educación*, 348.
- BLOOM, B., HASTINGS, T. Y MADAUS, G. F. (1973). *Manual de evaluación formativa y acumulativa del aprendizaje del alumno*. Santiago de Chile: Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas (CPEIP).
- BOLLEN, K. A. & LONG, J. S. (eds.) (1993). *Testing structural equation models*. Newbury Park: Sage.
- (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36, 317-346.
- (1993). *Estadística para investigadores: introducción al diseño de experimentos, análisis de datos y construcción de modelos*. Barcelona: Reverté, D.L.
- BROWNE, M. W. Y CUDECK, R. (1993). Alternative ways of assessing model fit. En BOLLEN, K. A. Y LONG, J. S. (Eds.). *Testing structural equation models* (136-162). Newbury Park: Sage.

- BYRNE, B. M. (2001). *Structural equation modeling with AMOS: basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum Associates.
- CASTRO MORERA, M., RUIZ DE MIGUEL, C. Y LÓPEZ MARTÍN, E. (2009). Forma básica de crecimiento en los modelos de valor añadido: vías para la supresión del efecto de regresión y funciones de crecimiento no-lineales. *Revista de Educación*, 348.
- ELOSUA, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepto. *Psicothema*, 17 (2), 356-362.
- GAVIRÍA, J. L., BIENCINTO, CH., LÓPEZ, E., RUIZ DE MIGUEL, C., JORNET, J., CASTRO, M., Y GONZÁLEZ, C. (2007). Cuestiones metodológicas de la evaluación de sistemas educativos. *Symposium presentado al XII Congreso Nacional de Modelos de Investigación Educativa, San Sebastián*, 105-114.
- 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, 351-387.
- JÖRESKOG, K. G. & SÖRBOM, D. (1989). *LISREL 7, a guide to the program and applications*. Chicago: SPSS Publications.
- KIPLINGER, V. (2004). *Longitudinal study of student growth in reading, writing and mathematics achievement in thirtysix Colorado public schools districts*. Colorado: Academic school district twenty.
- HAIR, J., ANDERSON, R., TATHAM, R. Y BLACK, W. (1999). *Análisis Multivariante*. Madrid: Prentice Hall.
- INSTITUTO DE EVALUACIÓN (2006). *Sistema estatal de indicadores de la educación 2006*. Madrid: MEC
- INECSE (2003). *Evaluación de la educación secundaria obligatoria 2000*. Madrid: MECD.
- (2005). *Evaluación de la educación primaria 2003*. Madrid: MEC.
- LARA, F. (2004). *La escuela como compromiso*. Madrid: Editorial Popular.
- LIZASOAIN, L. Y JOARISTI, L. (2009). Análisis de la dimensionalidad en modelos de valor añadido: estudio de pruebas de matemáticas empleando técnicas factoriales y métodos no-paramétricos basados en TRI. *Revista de Educación*, 248.
- NAVARRO, E. Y REDONDO, S. (2007). Estudio sobre el Rendimiento en Matemáticas en España a partir de los Datos del Informe PISA 2003. Un Modelo Jerárquico de Dos Niveles. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 5 (3), 118-136.
- PARDO, A. Y RUIZ, M.A. (2002). *SPSS11. Guía para el análisis de datos*. Madrid: McGraw Hill.

- SCHUMACKER, R. & LOMAX, R. (2004). *A beginner's to structural equation modelling*. Mahwah: Lawrence Erlbaum Associates.
- TABACHNICK, B. G. & FIDELL, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.
- TOPA, G., MORALES, J. F. & CAIERO, L. (2007). Social comparison and perceived breach of psychological contract: their effects on burnout in a multigroup analysis. *The Spanish Journal of Psychology*, 10 (1), 122-130.
- ZVOCH, K. & STEVENS, J.J. (2003). A multilevel, longitudinal analysis of middle school math and language achievement. *Education Policy Analysis Archives*, 11(20).

Dirección de contacto: Chantal Biencinto López. Universidad Complutense de Madrid. Facultad de Educación. Departamento de Métodos de Investigación y Diagnóstico en Educación. Avenida del Rector Royo Villanova s/n. 28040 Madrid. E-mail: alameda@edu.ucm.es

*Los datos de este artículo proceden de la investigación SEC2003-09742, proyecto de I+D financiado por el Ministerio de Ciencia y Tecnología. Este estudio no hubiera sido posible sin la desinteresada colaboración de la Subdirección General de Inspección Educativa de la Comunidad de Madrid, así como de la de cada uno de los inspectores que han participado de modo entusiasta y experto en algunas de las fases críticas de la misma