



ESTUDIO DE LA DIMENSIONALIDAD EMPLEANDO ANÁLISIS FACTORIAL CLÁSICO Y ANÁLISIS FACTORIAL DE INFORMACIÓN TOTAL: ANÁLISIS DE PRUEBAS DE MATEMÁTICAS DE PRIMARIA (5º Y 6º CURSOS) Y SECUNDARIA OBLIGATORIA

[Study of dimensionality using classic factor analysis and full information factor analysis: analysis of primary (5th and 6th degrees) and compulsory secondary mathematics tests]

por

[Article record](#)

[About authors](#)

[HTML format](#)

Joaristi, Luis (luis.joaristi@ehu.es)

Lizasoain, Luis (luis.lizasoain@ehu.es)

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

Abstract

The aim of this paper is to analyze the dimensional structure of a set of mathematical achievement tests used to assess the academic achievement in the Community of Madrid during 2005-06 and 2006-07 years and in three cohorts corresponding to the following academic degrees: 5th-6th of primary education, 1st-2nd and 3-4th of compulsory secondary education (ESO). The results coming from classical factor analysis and full information factor analysis are compared. Therefore a longitudinal study is performed in order to test whether such structure remains constant along time. The results confirm that most of the tests are mostly unidimensional structures. Meanwhile, it was proven that as the complexity of the contents increase, the complexity of the dimensional structure tests increase as well.

Keywords

Dimensionality, primary education, compulsory secondary education, mathematics achievement tests, factor analysis, full information factor analysis.

Resumen

El objetivo de este trabajo es analizar la estructura dimensional de las pruebas empleadas para evaluar el rendimiento académico en la asignatura de matemáticas en el contexto de la evaluación realizada en la Comunidad de Madrid en los cursos académicos 2005-06 y 2006-07 en tres cohortes: 5º y 6º de Educación Primaria, 1º-2º y 3º-4º de ESO. Para ello, se triangulan los resultados obtenidos de la aplicación de las técnicas factoriales clásicas con los del Análisis Factorial de Información total. Asimismo, se realiza un estudio longitudinal con objeto de ver si dicha estructura se mantiene constante a lo largo del tiempo. Los resultados confirman estructuras esencialmente unidimensionales. A su vez, se comprueba que el aumento de la complejidad de los contenidos implica un aumento en la complejidad de la estructura dimensional de las pruebas.

Descriptores

Dimensionalidad, educación primaria, ESO, pruebas de rendimiento de matemáticas, análisis factorial, Análisis Factorial de Información Total.

INTRODUCCIÓN

Cuando en un proceso evaluativo se emplean diseños longitudinales en los que cada

individuo es medido en varias ocasiones, esto implica que las puntuaciones obtenidas por los sujetos en cada uno de esos momentos deben ser equiparadas de forma que se

pueda disponer de una escala común a todas ellas. Las pruebas que se empleen deben tener validez curricular de forma que su contenido debe reflejar en cada momento los propios del curso que se evalúa. Por otra parte, las puntuaciones obtenidas deben tener una métrica común para todas las aplicaciones. La violación de estos supuestos puede tener consecuencias importantes para la validez de los resultados (Gaviria y Ruiz de Miguel, 2007).

Cuando, como es nuestro caso, se trata de una evaluación longitudinal, las pruebas tienen que adaptarse al currículo de cada año. Y aquí nos encontramos con dos tipos de competencias evaluables, a saber: de una parte aquellas habilidades de naturaleza transversal cuya estructura es básicamente lineal, como puede ser el caso de la «Comprensión Lectora», en la que la evolución en la habilidad no está asociada a un cambio en la complejidad de la estructura disciplinar. Pero por otro lado, hay materias, como las Matemáticas, en las que aunque existe una secuencia lógica en la organización de los contenidos, la incorporación de nuevos contenidos curso por curso lleva aparejado un aumento en la complejidad estructural de los mismos.

Para la realización de comparaciones a lo largo del tiempo, entre e intra centros, se emplean fundamentalmente diseños de valor añadido. Lo que garantiza la «comparabilidad» de las puntuaciones obtenidas en el contexto de un diseño longitudinal es la existencia de un modelo de medida que suponga una métrica común de las escalas de medida. Una muestra de este tipo de métrica común se puede encontrar en el trabajo de Antillón, Larrazolo y Backhoff (2006) en el que equiparan las puntuaciones de una prueba de acceso a la universidad mediante el procedimiento equipercantil. El que se pueda estimar en una escala común requiere a su vez satisfacer el supuesto de independencia local. La forma más fuerte de este supuesto es el requisito de unidimensionalidad que implica una única y común dimensión para todas las

escalas. La forma débil supone una estructura de dimensionalidad común a las mismas.

Cuando el proceso evaluativo se desarrolla a lo largo del tiempo y comprende cohortes de distintas edades (y en nuestro caso de diferentes niveles, primaria y secundaria), debemos verificar que en cada una de las cohortes las pruebas empleadas tienen la misma estructura dimensional; para así posteriormente poder emplear un modelo común de valor añadido en las mismas.

Stout et al. (1996) afirman que la evaluación de la dimensionalidad de una prueba implica dos tipos de problema: en primer lugar, la verificación o refutación de la unidimensionalidad y segundo, la descripción de la estructura multidimensional de la prueba, si ésta fuese necesaria. La verificación de la unidimensionalidad es imprescindible en aquellos procedimientos psicométricos que presuponen que los datos se ajustan a un modelo latente unidimensional o, cuando menos, que el desajuste respecto de esta condición no invalida la utilización de un procedimiento específico.

Para afrontar esta tarea, Martínez Arias et al. (2006) apuntan que las estrategias más habituales comienzan llevando a cabo un análisis factorial lineal sobre la matriz de correlaciones entre ítems, habitualmente la matriz de correlaciones tetracóricas. A partir de las soluciones aportadas por estas técnicas, se evalúa la dimensionalidad en función de distintos criterios: valores propios superiores a la unidad, diagrama de sedimentación, cociente entre los dos primeros valores propios, etc. Por ejemplo, Bayot, Hernández y de Julián (2005) emplean el análisis factorial exploratorio para estudiar las propiedades psicométricas de una escala de competencia parental. Sin embargo, está extendida la idea de que este grupo de técnicas no resuelve el problema adecuadamente. Así, se han propuesto distintas alternativas, como el análisis factorial no lineal. Esta técnica está enmarcada en la TRI multidimensional y es

realizable por medio del programa NO-HARM (Fraser y McDonald, 1988). Otro enfoque lo constituye el Análisis Factorial de Información Total de Bock et al. (2003) implementado en el programa TESTFACT. Además esta cuestión se puede abordar también mediante métodos no paramétricos, en los que no se parte de ningún supuesto sobre las funciones de respuesta. Los procedimientos HCA/CCPROX, DETECT y DIMTEST (disponibles en el programa DIMPACK) están basados en el concepto de unidimensionalidad esencial de Stout y en la estimación de las covarianzas condicionales de cada par de ítems. En otro trabajo (Lizasoain y Joaristi, en prensa) hemos analizado la dimensionalidad de estas pruebas empleando estos métodos no paramétricos.

En este estudio se pretende comprobar si es plausible la invarianza de la estructura dimensional en cohortes paralelas. Desde esta perspectiva, el objetivo de este trabajo es analizar la estructura dimensional de las pruebas empleadas para evaluar el rendimiento académico en la asignatura de matemáticas en el contexto de la evaluación realizada en la Comunidad de Madrid en los cursos académicos 2005-06 y 2006-07 en tres cohortes: 5º y 6º de Educación Primaria, 1º-2º y 3º-4º de ESO. Los datos utilizados para la realización de este trabajo proceden del Proyecto de I+D con referencia SEC2003-09742, ya finalizado, titulado: *El valor añadido en educación y la función de producción educativa: un estudio longitudinal*. Se ha evaluado la posible unidimensionalidad y la simplicidad o complejidad de la estructura de las pruebas. Asimismo se ha realizado un

estudio longitudinal con objeto de ver si dicha estructura se mantiene constante a lo largo del tiempo. Para ello, se han triangulado los resultados de las técnicas factoriales clásicas y del Análisis Factorial de Información Total.

PROCESO E INSTRUMENTOS DE EVALUACIÓN

Aquí nos vamos a limitar a apuntar a grandes rasgos las características generales del marco en que se encuadra el presente trabajo. La población de referencia considerada fue el conjunto de centros educativos de la Comunidad de Madrid con alumnos escolarizados en el último ciclo de Primaria y en la Secundaria Obligatoria en el curso académico 2005-06. Se midió a los alumnos al inicio y al final del curso académico, durante dos cursos consecutivos, 2005-06 y 2006-07. Se definieron tres segmentos educativos de interés: el último ciclo de Educación Primaria y los dos ciclos de la Secundaria Obligatoria.

Como se puede ver en la Tabla I, cada uno de estos segmentos educativos constituye una cohorte. Para los alumnos que en 2005-06 entraron en 5º de Primaria (cohorte 1ª) se han recogido datos en octubre de 2005 (primera medición), y junio de 2006 (segunda medición). En el curso 2006-07, cuando estos alumnos estaban en 6º de Primaria, se recogió información en noviembre de 2006 (tercera medición) y junio de 2007 (cuarta medición). Con las cohortes 2ª (correspondiente al primer ciclo de Secundaria) y 3ª (correspondiente al segundo ciclo de Secundaria) se ha procedido de manera análoga.

Tabla I. Tamaño de la muestra por cohortes y aplicaciones

Cohorte	Curso	Octubre 2005	Junio 2006	Curso	Noviembre 2006	Junio 2007
1ª	5º Primaria	4211	4173	6º Primaria	3627	3757
2ª	1º ESO	5106	4882	2º ESO	3327	3403
3ª	3º ESO	4736	4272	4º ESO	2761	2772

Las variables dependientes utilizadas han sido las puntuaciones obtenidas en las pruebas de Matemáticas elaboradas *ad hoc* apli-

cadadas a los alumnos de los centros de la muestra. Como ya se ha señalado, las mediciones se realizaron al principio y al final del

curso escolar, obteniendo un total de cuatro mediciones para cada alumno en la materia de Matemáticas.

Cabe aclarar los códigos que se pueden observar en distintas partes del trabajo. Por un lado, cada prueba tiene un código compuesto por la letra «m», que indica que se trata de pruebas de Matemáticas; le sigue un número que indica el curso: 5º Primaria, 6º Primaria, 1º ESO, 2º ESO, 3º ESO y 4º ESO son codificados por los números 5, 6, 1, 2, 3 y 4 respectivamente. Le sigue una letra que indica si la prueba corresponde a si la aplicación se ha realizado en octubre (o) o en junio (j); a continuación sigue otro número que hace referencia al año en que se ha realizado la prueba, de manera que 5 indica 2005, 6 el

2006 y 7 el 2007. El último elemento es una letra que especifica la forma de la prueba, a y b. Tal codificación se presenta en la Tabla IV. En cuanto a la posterior identificación de las dimensiones resultantes, hay que precisar que previamente se ha realizado una clasificación sustantiva de los distintos ítems desde el punto de vista de los «Contenidos» y de los «Dominios».

En la Tabla II se representa el número de ítems que, según un criterio sustantivo, está asociado a cada tipo de contenido – Estadística, Geometría, Medida, Números, Funciones y Tablas y Gráficos– según las distintas cohortes, fechas de aplicación y formas de la aplicación.

Tabla II. Número de ítems por cohorte, aplicación y forma en los distintos Contenidos

		APLICACIÓN							
		octubre 05		junio 06		noviembre 06		junio 07	
COHORTE	CONTENIDO	A*	B	A	B	A	B	A	B
5º + 6º Primaria	Estadística (ES)**			3	3	3	3	4	5
	Geometría (GE)	8	8	11	11	10	10	8	8
	Medida (ME)	11	12	11	11	10	10	11	10
	Números (NU)	21	20	15	15	15	16	15	15
	Total	40	40	40	40	38	39	38	38
1º + 2º ESO	Geometría	8	9	9	7	9	9	11	13
	Medida	6	6	3	2	2	2	2	3
	Números	19	20	23	24	25	26	23	20
	Tablas y gráficos (TA)	4	4	3	3	3	2	4	4
	Total	37	39	38	36	39	39	40	40
3º + 4º ESO	Estadística	6	5	7	7	4	5	12	12
	Funciones (FU)	5	6	8	9	10	9	10	9
	Geometría	13	12	6	6	6	5	8	8
	Números	16	17	17	18	18	18	8	9
	Total	40	40	38	40	38	37	38	38

*Indica la forma de la aplicación (hay dos formas, A y B).

**Entre paréntesis se representa el código de los distintos Contenidos.

En la Tabla III se representa el número de ítems que, sustantivamente, está asociado a cada tipo de dominio –Aplicación, Conoci-

mientos y Resolución de problemas– según las distintas cohortes, fechas de aplicación y formas de la aplicación.

Tabla III. Número de ítems por cohorte, aplicación y forma en los distintos Dominios

		APLICACIÓN							
		octubre 05		junio 06		noviembre 06		junio 07	
COHORTE	DOMINIO	A	B	A	B	A	B	A	B
5º + 6º Primaria	Aplicación (AP)	10	10	12	13	13	13	9	9
	Conocimientos (CO)	24	25	23	22	19	22	22	22
	Resolución de problemas (RP)	6	5	5	5	6	4	7	7
	Total	40	40	40	40	38	39	38	38
1º + 2º ESO	Aplicación	8	13	15	13	11	11	11	11
	Conocimientos	17	17	15	15	18	16	20	22
	Resolución de problemas	12	9	8	8	10	12	9	7
	Total	37	39	38	36	39	39	40	40
3º + 4º ESO	Aplicación	23	19	17	22	18	20	18	15
	Conocimientos	13	16	12	11	14	11	7	9
	Resolución de problemas	4	5	9	7	6	6	13	14
	Total	40	40	38	40	38	37	38	38

**Entre paréntesis se representa el código de los distintos Dominios.

En ambas tablas se puede observar que el número de ítems no es el mismo en todas las pruebas. Si inicialmente eran 40, ha habido unos pocos ítems que no han resultado operativos. Se debe aclarar el tipo de codificación que se utilizará, por ejemplo, en la Tabla VI. En ella se presentan los ítems que constituyen la prueba correspondiente a 6º de Primaria de noviembre de 2006 bajo la forma B de la aplicación (código m6o6b). Las dos primeras letras se refieren al tipo de «Contenido» (Tabla II) y las dos siguientes de forma abreviada a los distintos tipos de «Dominio» (Tabla III); les sigue un número que identifica al ítem dentro de la prueba y, por fin, puede aparecer la letra A (forma a de la prueba), B (forma B) o ninguna (incluida en ambas formas de la prueba).

ANÁLISIS MEDIANTE TÉCNICAS FACTORIALES

En el presente apartado se describen los distintos análisis de datos realizados. Con el fin de establecer el número de factores, se comenzó con el análisis más clásico, el de Componentes Principales. Para facilitar el análisis de la unidimensionalidad se completaron los resultados por medio de una técnica más específica, tal como el Análisis Factorial de Información Total. A continuación se interpretan los

resultados en cuanto a identificación de los factores y unidimensionalidad. Los resultados han sido obtenidos utilizando los programas SPSS, Excel y TESTFACT.

Análisis Factorial clásico

Debido a que se trata de ítems dicotómicos, la matriz que se ha de analizar es la de las correlaciones tetracóricas, pues el análisis basado en la matriz de correlaciones de Pearson suele dar origen a la aparición de factores de dificultad formados por ítems muy fáciles o difíciles, es decir, los ítems aparecen agrupados por algunas de sus características estadísticas. Esto hace que la solución presente alguna dimensión no asociada a ningún aspecto sustantivo y que la unidimensionalidad quede enmascarada (Gaviria, 1990; Nunnally y Bernstein, 1994). Al respecto también se puede encontrar un estudio relativo a la conveniencia de la utilización de las correlaciones tetracóricas de ítems dicotómicos para su análisis factorial en Muthén (1988).

La matriz de correlaciones tetracóricas pueden no ser definida positiva, condición para que los valores propios (es decir las varianzas) sean positivos. Podemos hacernos una idea de la complejidad del estudio si se tiene en cuenta que se sometieron a análisis las matrices de datos correspondientes a 24 pruebas compues-

tas por ítems dicotómicos; se trataba de tres cohortes, cuatro aplicaciones y dos formas. Las matrices de correlaciones tetracóricas se obtuvieron por medio del programa TESTFACT. Para llevar a cabo los análisis a los que hace referencia este epígrafe, las matrices así obtenidas fueron sometidas a análisis mediante el programa SPSS. Como es habitual, se aplicó la medida de KMO (sobre la proporción de la varianza común) y en la prueba de esfericidad de Bartlett (sobre la hipótesis nula de que la matriz de correlaciones es la matriz identidad).

En la mayor parte de las pruebas la proporción de varianza común fue superior a 0,80; las excepciones correspondieron a m5o5b, m6o6a y m3j6a, que presentan valores próximos a 0,70. Los datos de las pruebas m6o6b y m3j6b produjeron matrices de correlaciones tetracóricas que no son definidas positivas. En cuanto a la esfericidad en todas las pruebas (excepto las mencionadas m6o6b y m3j6b, que no dieron matrices definidas positivas) se rechazó la hipótesis nula.

Con el gráfico de sedimentación, los resultados mostraron un patrón muy claro siendo adecuada para la mayoría la solución de dos dimensiones; excepto las pruebas m5o5a, m5o5b, m1o5a, m2o6a, m2j7b y m3o5a que precisaron de tres o cuatro dimensiones.

En la Tabla IV se pueden ver los valores propios y sus proporciones sobre la inercia total correspondientes a un análisis en el que se han retenido los factores de valor propio mayor o igual a uno. Junto a estos, la última columna muestra un cociente; que es el resultado de dividir el primer valor propio entre el segundo. Los resultados están clasificados por cohorte, aplicación y forma de la aplicación. Según el criterio del cociente entre los dos primeros valores propios (Martínez Arias, 2006), se puede observar que son las cohortes de 5º y 6º de Primaria y la de 1º y 2º de ESO aquellas en las que dicho cociente es en general mayor que 5. Por fecha de aplicación, es la de junio de 2006 la que en más ocasiones presenta unidimensionalidad.

Tabla IV. Valores propios, tasas de inercia, y cocientes

Prueba	Detalle de la Prueba	λ_1	λ_2	Proporción de la inercia de F_1	Proporción de la inercia de F_2	λ_1 / λ_2
m5o5a	5º Prim octubre 2005 A	10,21	2,08	,2618	,0533	4,92
m5o5b	5º Prim octubre 2005 B	9,58	2,07	,2456	,0531	4,63
m5j6a	5º Prim junio 2006 A	10,80	1,93	,2769	,0495	5,59
m5j6b	5º Prim junio 2006 B	11,13	1,94	,2854	,0497	5,73
m6o6a	6º Prim noviembre 2006 A	10,07	1,91	,2582	,0490	5,26
m6o6b	6º Prim noviembre 2006 B	9,91	1,95	,2541	,0500	5,08
m6j7a	6º Prim junio 2007 A	9,94	1,67	,2549	,0428	5,97
m6j7b	6º Prim junio 2007 B	9,60	1,62	,2462	,0415	5,93
m1o5a	1º ESO octubre 2005 A	8,11	1,62	,2079	,0415	5,01
m1o5b	1º ESO octubre 2005 B	7,24	1,85	,1856	,0474	3,91
m1j6a	1º ESO junio 2006 A	8,32	1,53	,2133	,0392	5,44
m1j6b	1º ESO junio 2006 B	8,55	1,59	,2192	,0408	5,37
m2o6a	2º ESO noviembre 2006 A	8,97	1,91	,2300	,0490	4,70
m2o6b	2º ESO noviembre 2006 B	9,01	1,79	,2310	,0459	5,03
m2j7a	2º ESO junio 2007 A	8,37	1,81	,2146	,0464	4,63
m2j7b	2º ESO junio 2007 B	8,19	1,87	,2100	,0479	4,38
m3o5a	3º ESO octubre 2005 A	7,41	2,30	,1900	,0590	3,22
m3o5b	3º ESO octubre 2005 B	7,06	1,70	,1810	,0436	4,15
m3j6a	3º ESO junio 2006 A	5,06	1,85	,1297	,0474	2,74
m3j6b	3º ESO junio 2006 B	8,59	2,09	,2203	,0536	4,12
m4o6a	4º ESO noviembre 2006 A	7,13	1,78	,1828	,0456	3,99
m4o6b	4º ESO noviembre 2006 B	7,44	1,66	,1908	,0426	4,47
m4j7a	4º ESO junio 2007 A	6,53	1,58	,1674	,0405	4,14
m4j7b	4º ESO junio 2007 B	6,37	1,71	,1633	,0438	3,72

Obsérvese que los resultados para m6o6b y m3j6a se muestran en cursiva y recuérdese que son éstas las pruebas cuyas matrices de correlaciones tetracóricas no son definidas positivas. Los resultados para estas dos prue-

bas han sido obtenidos utilizando TESTFACT. Como resumen de la Tabla IV se presenta la Tabla V, en que se remarcan sobre fondo gris las pruebas en que sí se cumple el criterio.

Tabla V. Cociente entre primer y segundo valor propio

APLICACIÓN	octubre 05		junio 06		noviembre 06		junio 07	
	A	B	A	B	A	B	A	B
COHORTE								
5º + 6º Primaria								
1º + 2º ESO								
3º + 4º ESO								

Como se ve, de las 24 pruebas sólo en diez se da unidimensionalidad según el criterio del cociente entre los valores propios de los dos primeros factores. Sin embargo, en ámbitos como éste y dado el elevado número de pruebas, es casi imprescindible la utilización de otras herramientas, algunas de ellas *ad hoc* para el análisis de la unidimensionalidad.

Análisis Factorial de Información Total

Como antes hemos afirmado, las técnicas factoriales *clásicas* proceden mediante la diagonalización de la matriz de correlaciones entre todas las variables objeto de estudio. En el caso que nos ocupa, se trata de una matriz de correlaciones tetracóricas dado que las respuestas a los ítems se consideran como una variable dicotomizada con distribución normal subyacente. Debido a que las matrices de dos de las pruebas no eran definidas positivas y para las cuales SPSS no da solución, empleamos el programa TESTFACT, que supera la dificultad reemplazando la matriz original por otra «suavizada». La lógica del proceso consiste en que, partiendo de la descomposición espectral de las matrices simétricas, la matriz de correlaciones suavizada se obtiene reemplazando los elementos de la matriz diagonal de los valores propios que han resultado negativos por ceros o valores positivos pequeños.

TESTFACT aplica el método de factorización MINRES o Residuos Mínimos o Mínimos Cuadrados no ponderados, consistente en que no hay estimación previa de las communalidades. Su objetivo es maximizar las correlaciones entre las variables objeto de análisis basándose en los mínimos cuadrados para minimizar los residuos. Es sabido que para interpretar o identificar los factores es muy conveniente realizar algún tipo de rotación con el fin, desde el punto de vista estadístico, de obtener una estructura simple. Debido a que la discusión se centra en los métodos VARIMAX (ortogonal) y PROMAX (oblicuo), Finch (2006) aprecia que el método oblicuo es más efectivo en la identificación de una estructura simple.

Sólo queda obtener la solución factorial a través del Análisis Factorial de Información Total para después proceder a las rotaciones convenientes. A diferencia de los métodos factoriales clásicos que, como antes hemos afirmado, operan diagonalizando la matriz de correlaciones; este tipo de análisis no utiliza las correlaciones entre las variables del modelo sino que se basa, por una parte, en el empleo de la matriz de los patrones de respuesta de los sujetos a los ítems, y por otro, en incorporar la información proveniente de los parámetros de la TRI para estimar las cargas factoriales (por ejemplo la dificultad de los ítems, la adivinación por azar, etc.).

Dicho de otra manera, se utilizan como datos de entrada, no las correlaciones entre variables, sino las frecuencias de todos los patrones de respuesta o vectores de respuesta al ítem. En definitiva, en vez de operar con un modelo factorial de uso *general*, nos encontramos aquí con una técnica factorial diseñada específicamente para analizar desde la perspectiva de este conjunto de modelos, las respuestas de los sujetos a los ítems de una prueba, empleando -aprovechando podríamos decir- toda la información disponible al respecto. Veamos brevemente las características básicas del mismo. Para una descripción más rigurosa de la fundamentación matemática de esta técnica puede

recurrirse a los trabajos de Bock y Aitkin (1981), de Muraki y Engelhard (1985) y de Bock, Gibbons y Muraki (1988).

Los ya citados Bock y Aitkin (1981) aplican el modelo del factor múltiple de Thurstone a datos de respuesta (y_{ij}) al ítem (j) de m factores suponiendo que no describe una variable manifiesta sino un proceso de respuesta inobservable en que cada ítem binario j tiene asociada una respuesta cuantitativa continua subyacente y una γ_j . (Es decir, cada ítem se estudia bajo m factores comunes θ , cuyas cargas factoriales α hay que estimar.)

$$y_{ij} = \alpha_{j1} \theta_{1i} + \dots + \alpha_{jk} \theta_{ki} \dots + \alpha_{jm} \theta_{mi} + \varepsilon_{ij} = \sum_{k=1}^m \alpha_{jk} \theta_{ki} + \varepsilon_{ij}$$

con:

- $\varepsilon_{ij} \sim N(0, \sigma_j^2)$ con $\sigma_j^2 = 1 - \sum_k \alpha_{jk}^2$
- $m =$ número de factores, siendo:
- $\alpha_j = (\alpha_{j1} \dots \alpha_{jk} \dots \alpha_{jm})$, es decir, $k = 1, \dots, m$ para las cargas factoriales;
- $\theta_i = (\theta_{1i} \dots \theta_{ki} \dots \theta_{mi})$, es decir, $k = 1, \dots, m$ para las variables latentes no medibles (vector de aptitudes m -dimensional del sujeto i).

Para un sujeto i seleccionado al azar:

- $\theta_i = (\theta_{1i} \dots \theta_{ki} \dots \theta_{mi}) \sim N(\mathbf{0}_m, \mathbf{I}_m)$ ¹
- $\mathbf{y}_i = (y_{i1} \dots y_{ij} \dots y_{in}) \sim N(\mathbf{0}_n, \mathbf{I}_n)$
- $\varepsilon_{ij} \sim N[\mathbf{0}_n, \text{diag}\{\sigma_j^2\}]$

En el análisis factorial de ítems dicotomizados, se genera una respuesta correcta del sujeto $i \in \{1, \dots, N\}$ en el ítem $j \in \{1, \dots, n\}$ y el factor $k \in \{1, \dots, m\}$ ante un umbral γ_j (dificultad del ítem j), si $y_{ij} \geq \gamma_j$. El proceso inobservable y_{ij} es latente y realizado sólo como un vector de variables-respuesta dicotómicas $\mathbf{x}_i = (x_{i1} \dots x_{ij} \dots x_{in})$ con:

$$x_{ij} = \begin{cases} 1 & \text{si } y_{ij} \geq \gamma_j \\ 0 & \text{si } y_{ij} < \gamma_j \end{cases} \quad \forall j = 1 \dots n$$

NOTA: el modelo de análisis factorial para variables categóricas se basa en que y_{ij} (variable del proceso de respuesta) es latente y realizada en un vector de respuestas de ítems dicotómicos (0, 1) para n ítems x_{ij} que cumplen que $x_{ij} = 1$ si $y_{ij} \geq \gamma_j$.

La probabilidad de respuesta correcta o positiva ($x_{ij} = 1$) del sujeto i al ítem j , dada la "aptitud latente m -dimensional" θ_i del sujeto i es:

$$P(x_{ij} = 1 | \theta_i) = \Phi_{N(0,1)} \left[\frac{\sum_{k=1}^m \alpha_{jk} \theta_{ki} - \gamma_j}{\sigma_j} \right],$$

a lo que denotamos $\Phi_j(\theta_i)$, en que Φ es la función de distribución de la ley normal estándar.

En la estimación de los cargas factoriales se parte del supuesto de que los datos proceden de una población en que la aptitud sigue una

¹ En negrita se representan vectores y matrices.

distribución multivariante $\Theta \sim N(0, I)$ en el caso en que se trate de factores ortogonales. Además, como en todo análisis factorial, se supone que $y_j \sim N(0,1)$.

Si el número de ítems es n y un modelo de respuesta binaria se representa por el vector $\mathbf{x}_t = (x_{t_1}, x_{t_2}, \dots, x_{t_j}, \dots, x_{t_n})$, en que

x_{t_j} toma los valores 1 y 0 según sea contestado correcta o incorrectamente. La probabilidad condicional es:

$$P(\mathbf{x} = \mathbf{x}_t | \boldsymbol{\theta}_i) = \prod_{j=1}^n (\Phi_j(\boldsymbol{\theta}_i))^{x_{t_j}} (1 - \Phi_j(\boldsymbol{\theta}_i))^{(1-x_{t_j})}$$

Su probabilidad marginal es la condicional $\forall \boldsymbol{\theta}_i$:

$$P_t = P(\mathbf{x} = \mathbf{x}_t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{j=1}^n (\Phi_j(\boldsymbol{\theta}))^{x_{t_j}} (1 - \Phi_j(\boldsymbol{\theta}))^{(1-x_{t_j})} g(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} L_t(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

en que $g(\boldsymbol{\theta})$ es, para N sujetos aleatoriamente extraídos de la población, la función de densidad de la distribución normal multivariante del rasgo latente $\boldsymbol{\theta}$. Esta integral se calcula por medio de la cuadratura de Gauss-Hermite.

El procedimiento de estimación de máxima verosimilitud marginal se aplica a la tabla de contingencia de los patrones de respuesta (Abad, Ponsada y Revuelta, 2006), a partir de la cual se construye la función a optimizar:

$$L_M = P(X) = \frac{N!}{\prod_{h=1}^s n_h!} \prod_{h=1}^s P_h^{n_h}$$

en que L_M es la probabilidad de una muestra con:

- $n_1 \dots n_h \dots n_s$ son las frecuencias de los patrones de respuesta $x_1 \dots x_h \dots x_s$,
- N es el tamaño de la muestra,
- s es el número de patrones diferentes ($s \leq \min\{2^n, N\}$).

NOTA: Obsérvese que la denominación del procedimiento incluye el término “total” o “completa” (full) por esta razón, porque se estudian todos los patrones de respuesta. Es decir, se utilizan como datos las frecuencias de todos los patrones de respuesta o vectores

de respuesta al ítem, \mathbf{x}_t con $t = \{1, \dots, h, \dots, s\}$.

Las estimaciones MV marginal de las γ_j y cargas factoriales α_{jk} maximizan la ecuación de L_M , obteniéndose éstas de las estimaciones MV marginal de los parámetros de intersección (c_j) y pendiente (a_{j1}, \dots, a_{jm}) de la función de respuesta utilizando el algoritmo EM, que se utiliza para la obtención de estimaciones de Máxima Verosimilitud de parámetros en modelos que dependen de variables no observadas. (Bock y Aitkin introducen el método de estimación MV Marginal para estimar los parámetros de los modelos TRI de ojiva normal de 1 y 2 parámetros utilizando EM.).

En el proceso de estimación hay que tener presente que el algoritmo EM converge lentamente; por ello para el proceso iterativo conviene que los valores de partida sean adecuados. Estos se obtienen por Análisis de Factores Principales MINRES sobre la matriz de correlaciones tetracóricas. Así, se obtienen factores ortogonales y sus cargas factoriales son las adecuadas para el comienzo del proceso iterativo del Análisis Factorial de Información Total después de convertir esas cargas factoriales en pendientes e intersección de los ítems, pues además conviene expresar el modelo en términos de

a_j y c_j pues se obtiene una expresión simplificada de L_M .

Una vez obtenida la convergencia, es decir la solución final de las pendientes y la intersección, éstas se transforman (Takane y De Leeuw, 1987) en la solución factorial definitiva a través de:

$$\text{cargas factoriales: } \alpha_j = \frac{a_j}{d_j} \text{ dificultad}$$

$$\text{estandarizada } \gamma_j = -\frac{c_j}{d_j} \text{ con } d_j^2 = 1 + \sum_k a_{jk}^2$$

Posteriormente, con el fin de buscar la estructura simple, se podrá rotar esta solución de forma ortogonal (VARIMAX) u oblicua (PROMAX).

En resumen, tal y como al inicio apuntamos, el Análisis Factorial de Información Total parte de los patrones de respuesta a los ítems y no de la matriz de correlaciones ni de covarianzas. Se utiliza el algoritmo EM en las ecuaciones del método de estimación de Máxima Verosimilitud Marginal.

PROCESO Y RESULTADOS

Las soluciones factoriales de las 24 matrices de correlaciones tetracóricas han sido obtenidas con TESTFACT. Para ello se ha

utilizado el método de factorización MINRES o Mínimos Cuadrados no ponderados. Asimismo se ha utilizado este programa para obtener las puntuaciones totales en la aplicación correspondiente. Recordemos que en el método de factorización MINRES (residuos mínimos) no hay estimación previa de las communalidades y que su objetivo es maximizar las correlaciones entre las variables objeto de análisis basándose en los mínimos cuadrados para minimizar los residuos.

Como se ha visto, dos de las matrices de correlaciones no son definidas positivas, por lo que SPSS no ha proporcionado ningún resultado. Por ello se han comprobado y completado los resultados por medio del comando FACTOR de TESTFACT. Utilizando este programa se ha obtenido la solución sin rotar por medio del Análisis Factorial de Información Total. El programa facilita también la solución factorial bajo rotación ortogonal VARIMAX.

Como última fase del análisis, se han utilizado las saturaciones obtenidas tras rotación oblicua PROMAX de parámetro cuatro, que se presentan en la Tabla VI. Ahora disponemos de una solución más fácilmente interpretable en la medida en que en el segundo factor disponemos de ítems que, sólo o mayoritariamente, cargan en el mismo. La correlación entre ambos factores es de 0,73.

Tabla VI. Cargas factoriales en los dos primeros factores con rotación PROMAX

Ítem	F1	F2	Ítem	F1	F2	ítem	F1	F2
ESCO1	0.002	0.545	NUCO14	0.476	0.090	MEAP46B	0.509	0.040
MEAP2	0.415	0.137	NUCO15	0.457	0.089	MECO47B	0.449	0.059
ESAP3	-0.094	0.545	NUCO16	0.613	-0.089	MEAP48B	0.545	-0.107
ESAP4	0.546	0.046	NUAP17	0.568	-0.028	NUCO49B	0.570	-0.038
GECO5	-0.122	0.722	NUPR18	0.516	-0.160	NUCO50B	0.359	0.101
GECO6	-0.087	0.740	MEAP19	0.604	-0.046	NUCO51B	0.196	0.215
GECO7	0.064	0.526	NUCO39B	0.273	0.157	NUAP52B	0.461	-0.007
GECO8	0.047	0.484	NUCO40B	0.366	0.023	NUAP53B	0.534	-0.077
GEAP9	0.129	0.479	NUCO41B	0.447	0.024	MECO54B	0.379	0.129
GEAP10	-0.137	0.739	NUCO42B	0.599	0.069	GEPR55B	0.245	0.124
GEAP11	0.070	0.649	NUCO43B	0.563	0.037	MECO56B	0.576	-0.201
GECO12	0.203	0.267	NUPR44B	0.666	-0.044	MECO57B	0.661	-0.061
GEAP13	0.115	0.183	MECO45B	0.305	0.144	MEPR58B	0.589	-0.122

Así pues, se obtendrán las cargas factoriales de todas las pruebas tras aplicar a las 24 matrices de correlaciones tetracóricas, suavizadas si fuese preciso, el Análisis Factorial de Información Total seguido de una rotación PROMAX.

INTERPRETACIÓN DE LOS FACTORES

Asociación con facilidad y dificultad estandarizada

El objetivo de los resultados de la Tabla VII es identificar aquellas pruebas en que puede sospecharse de la presencia de «artefactos» provocados por la semejanza de las respuestas (Martínez Arias, 2006). Cuanto

mayor sea la correlación entra facilidad y F1 (o F2), más riesgo de que la agrupación de ítems se haya producido estadísticamente y no substancialmente. La forma de interpretar los resultados adjuntos consiste en que si, por ejemplo, en una determinada prueba, la correlación entre la Facilidad de los ítems correspondientes y de sus saturaciones en F1 es de 0.795 (6º de Primaria octubre de 2006 bajo la forma A), implica que cuanto más fácil es un ítem, mayor saturación tiene en F1. Recordar que las saturaciones se han obtenido por medio del método MINRES y rotación PROMAX. Los valores de Facilidad y Dificultad estandarizada se han obtenido con TESTFACT.

Tabla VII. Correlaciones de Facilidad y Dificultad estandarizada con F1 y F2

Prueba	F1 y Facilidad	F2 y Facilidad	F1 y Dificultad	F2 y Dificultad	Cohorte	Medición
5º Prim octubre 2005 A	-,070	,654 **	-,354	-,463 *	1ª	1
5º Prim octubre 2005 B	-,283	,819 **	,448 **	-,956 **	1ª	1
5º Prim junio 2006 A	,283	,115	,369 *	-,894 **	1ª	2
5º Prim junio 2006 B	,492 **	-,300	-,138	-,270	1ª	2
6º Prim noviembre 2006 A	,795 **	-,721 **	-,958 **	,898 **	1ª	3
6º Prim noviembre 2006 B	-,050	,194	,739 **	-,907 **	1ª	3
6º Prim junio 2007 A	,810 **	-,732 **	-,987 **	,871 **	1ª	4
6º Prim junio 2007 B	,351 *	-,037	-,904 **	,594 **	1ª	4
1º ESO octubre 2005 A	-,632 **	,831 **	,834 **	-,985 **	2ª	1
1º ESO octubre 2005 B	,849 **	-,681 **	-,956 **	,855 **	2ª	1
1º ESO junio 2006 A	,903 **	-,768 **	-,945 **	,962 **	2ª	2
1º ESO junio 2006 B	,832 **	-,705 **	-,984 **	,907 **	2ª	2
2º ESO noviembre 2006 A	,792 **	-,379 *	-,995 **	,704 **	2ª	3
2º ESO noviembre 2006 B	,871 **	-,648 **	-,989 **	,815 **	2ª	3
2º ESO junio 2007 A	,915 **	-,475 **	-,978 **	,534 **	2ª	4
2º ESO junio 2007 B	,912 **	-,500 **	-,997 **	,691 **	2ª	4
3º ESO octubre 2005 A	,857 **	-,671 **	-,917 **	,750 **	3ª	1
3º ESO octubre 2005 B	,844 **	-,323 *	-,954 **	,386 *	3ª	1
3º ESO junio 2006 A	-,591 **	-,663 **	,191	-,021	3ª	2
3º ESO junio 2006 B	-,467 **	,818 **	,918 **	-,914 **	3ª	2
4º ESO noviembre 2006 A	-,309	,676 **	-,181	-,066	3ª	3
4º ESO noviembre 2006 B	,150	,023	-,315	,787 **	3ª	3
4º ESO junio 2007 A	,662 **	-,517 **	-,426 **	,933 **	3ª	4
4º ESO junio 2007 B	,563 **	-,320 *	-,633 **	,979 **	3ª	4

**p < 0,01 (bilateral)

*p < 0,05 (bilateral)

En la Tabla VII se presentan las 24 correlaciones de Pearson entre la facilidad de los ítems de cada aplicación y la saturación en

cada uno de los dos primeros factores, F1 y F2. En ella se somborean las casillas con correlaciones más altas; éstas podrían ser debi-

das a que en esas pruebas hay demasiados ítems fáciles, pues las correlaciones son positivas. Se observa que es en la segunda cohorte en que mayores correlaciones con el primer factor que hay. En cuanto al segundo factor, no se observan correlaciones positivas altas más que en tres pruebas; por el contrario hay varias en que la correlación es alta y negativa, pudiendo ser debido a que en tales pruebas hay pocos ítems fáciles. En lo que se refiere a la «Dificultad estandarizada», se puede ver que la estructura de las correlaciones entre ella y F1 es básicamente semejante a la de la Facilidad. En el segundo factor hay ya bastantes discrepancias con respecto a la Facilidad.

Interpretación de las dimensiones según Contenido y Dominio

El criterio utilizado en la identificación y caracterización de los factores se ha elaborado partiendo en primer lugar de las saturaciones de los ítems en los dos primeros factores, obtenidas con TESTFACT según antes hemos visto. A continuación, siguiendo el

criterio de Stevens (2002), se han retenido sólo los ítems cuyas saturaciones o cargas factoriales son superiores en valor absoluto a 0,4. Es decir, ahora la condición de cada ítem es si ha sido retenido o no para la identificación de cada factor.

Para cada una de las 24 pruebas, se han obtenido cuatro tablas de contingencia cruzando las dos nuevas variables mencionadas (ítem retenido o no para cada factor) con la variable relativa al contenido de los ítems bajo las seis modalidades que ya hemos visto (estadística, geometría, medida, números, funciones y tablas) y la variable sobre el dominio cognitivo (tres modalidades: conocimientos, aplicaciones y resolución de problemas). Posteriormente, para cada una de las 24 pruebas se ha obtenido, de las cuatro tablas de contingencia, el residuo tipificado corregido de Haberman. Por fin, si se tienen en cuenta sólo los residuos corregidos más notables, resultan las Tablas VIII y IX.

Tabla VIII. Resumen de la asociación entre los dos primeros factores y Contenido

COHORTE	Contenido			
	F1		F2	
1ª cohorte	Estadística	2	Estadística	4
	Geometría	3	Geometría	4
	Medida	6	Medida	1
	Números	4	Números	2
2ª cohorte	Geometría	1	Geometría	5
	Medida	6	Medida	1
	Números	4	Números	4
	Tablas y gráficos	4	Tablas y gráficos	0
3ª cohorte	Estadística	3	Estadística	5
	Funciones	1	Funciones	4
	Geometría	1	Geometría	1
	Números	6	Números	3

En la Tabla VIII se especifican los contenidos específicos para cada cohorte y el número de pruebas, resultando un total de 8 pruebas (4 aplicaciones por 2 formas). Por ejemplo, para la primera cohorte, en Estadística hay dos aplicaciones en que F1 está asociado a tal contenido. Así, en cuanto al primer factor, en la primera cohorte dominan la

Medida y los Números; en la segunda sigue siendo Medida el contenido dominante, seguido de Números y Tablas y gráficos; por fin, en la tercera cohorte son los Números seguidos de Estadística los dominantes. En lo que se refiere al segundo factor, en la primera cohorte está asociado a Estadística y Geometría; en la segunda cohorte aparecen

Geometría y Números y en la tercera, Estadística y Funciones. En resumen, podemos concluir que el primer factor está fundamentalmente asociado a Medida y Números;

mientras que el segundo factor es algo más disperso, pero aparecen sobre todo Geometría y Estadística, siendo la primera el contenido más claro.

Tabla IX. Resumen de la asociación entre los dos primeros factores y Dominio

COHORTE	Dominio			
	F1		F2	
1ª cohorte	Conocimientos	4	Conocimientos	0
	Aplicaciones	1	Aplicaciones	4
	Resolución de problemas	3	Resolución de problemas	3
2ª cohorte	Conocimientos	3	Conocimientos	2
	Aplicaciones	4	Aplicaciones	3
	Resolución de problemas	1	Resolución de problemas	5
3ª cohorte	Conocimientos	2	Conocimientos	4
	Aplicaciones	3	Aplicaciones	1
	Resolución de problemas	4	Resolución de problemas	3

En lo que se refiere a la caracterización en función de los dominios cognitivos, observando la Tabla IX, vemos que –para las dos primeras cohortes– en el primer factor aparecen los tres dominios, mientras que el segundo se asocia básicamente a aplicaciones y resolución de problemas.

Estudio longitudinal

Hasta ahora en este trabajo hemos estado abordando el examen conjunto de las 24 pruebas pero en la mayoría de los casos el enfoque adoptado ha sido el de considerarlas de forma independiente. Sólo muy ocasionalmente hemos realizado algún comentario comparando las tres cohortes objeto de estudio. Pero además de contar con tres cohortes, el diseño de la evaluación es longitudinal con cuatro medidas, dos por curso. Y ante esto, la cuestión es: ¿podemos realizar algún enfoque o estudio de carácter longitudinal relativo a la estructura dimensional de estas pruebas? Acabamos de ver cómo –dentro de cada cohorte– la estructura dimensional no es exactamente la misma para las ocho pruebas de cada una, pero ciertamente tampoco es muy distinta. Ante esta situación, ¿se puede, por ejemplo, afirmar que la estructura de las pruebas –sea ésta la que sea– se mantiene a lo largo del tiempo? La

pregunta es importante porque si esta estructura es muy cambiante, esto va a dificultar la elaboración, validación e interpretación adecuada de modelos de valor añadido que nos puedan informar adecuadamente del cambio experimentado por los estudiantes.

El enfoque tradicionalmente adoptado para este tipo de estudio es la realización de análisis factoriales confirmatorios que permitan evaluar el grado de ajuste de un modelo dimensional previamente formulado. Pero aquí nos encontramos con una dificultad a la hora de plantearnos un enfoque así y es que las pruebas no son las mismas por lo que, cómo máximo, tal estudio podría hacerse para cada par consecutivo de aplicaciones con los ítems de anclaje. Pero obviamente sólo ellos no representan la estructura dimensional del conjunto de una prueba por lo que esta vía ha de ser desechada.

Ante esta situación, pensamos que sí es viable plantearse el estudio de indicadores indirectos de la consistencia temporal de la estructura dimensional de las pruebas empleadas en el proceso de evaluación. Y la propuesta que formulamos es la siguiente: si esta estructura se mantiene a lo largo del tiempo estas pruebas ordenarán a los sujetos de manera similar en

las cuatro aplicaciones. O dicho de otra manera, el rango de cada estudiante con respecto al grupo de referencia ha de ser muy similar en las cuatro aplicaciones si la estructura de las cuatro pruebas es similar por más que lógicamente éstas varíen en función del currículo de cada momento.

Y para ello hemos procedido de la siguiente manera: para cada cohorte hemos generado un archivo de datos en SPSS que contiene el identificador del estudiante y de su centro, y para cada una de las cuatro aplicaciones, la puntuación directa de la prueba (número de ítems correctamente respondidos), y las puntuaciones en los dos factores extraídos de TEST-FACT después de la rotación PROMAX. A continuación se han transformado las puntua-

ciones obteniendo los rangos para cada una de las mismas. A su vez, esta ordenación por rangos se ha efectuado de dos maneras: una globalmente tomando a toda la muestra de cada cohorte como grupo de referencia; y una segunda segmentando por centro escolar con objeto de controlar posibles efectos de regresión a la media. Y para examinar la concordancia de estas ordenaciones se ha calculado para cada par de variables el coeficiente de Spearman.

A continuación, en la Tabla X se presentan las correlaciones para los tres conjuntos de variables (*punt* para puntuaciones directas, *fact1* y *fact2* para las dos puntuaciones factoriales) correspondientes a la cohorte de 1º-2º de la ESO. Todos son significativos al 0,0001.

Tabla X. Coeficientes Rho de Spearman

Rho de Spearman Puntuaciones	m1j6	m2o6	m2j7	Rho de Spearman Factor 1	m1j6	m2o6	m2j7	Rho de Spearman Factor 2	m1j6	m2o6	m2j7
m1o5	,717 4448	,690 2600	,637 2647	m1o5	,605 4448	,592 2600	,525 2647	m1o5	,457 4448	,651 2600	,540 2647
m1j6		,740 2556	,700 2601	m1j6		,689 2556	,651 2601	m1j6		,621 2556	,663 2601
m2o6			,699 2888	m2o6			,658 2888	m2o6			,658 2888

Como vemos, las correlaciones más altas se alcanzan entre las puntuaciones directas y, lógicamente, son mayores cuanto más próximas en el tiempo sean las aplicaciones. Para no alargar excesivamente este trabajo, digamos que este patrón de asociación se da tanto en la cohorte de 5º y 6º de primaria como en ésta que acabamos de ver. Por el contrario, en la de

3º y 4º de la ESO las correlaciones son notablemente más bajas. En la Tabla XI mostramos para las tres cohortes los valores más altos de los coeficientes, obtenidos entre las puntuaciones directas de los tres pares de aplicaciones consecutivas, así como los promedios por cohorte.

Tabla XI. Valores máximos de Rho

Rho de Spearman Aplicaciones consecutivas (puntuaciones)	5º-6º	1º-2º	3º-4º
Oct-05-Jun-06	,621	,717	,120
Jun-06-Oct-06	,760	,740	,362
Oct-06-Jun-07	,745	,699	,597
Promedios	,708	,718	,359

Con respecto a los rangos por centros, lógicamente no vamos a poner aquí ninguna tabla. Baste decir al respecto que los valores son muy similares y que se mantiene la diferencia de la tercera cohorte con respecto a las otras dos. Obviamente también hay diferencias notables entre algunos centros sobre todo en aquellos con pocos estudiantes por grupo. El promedio global de los coeficientes para las tres cohortes es de 0,66 con una desviación típica de 0,14. Como valores extremos señalamos que el máximo alcanzado es de 0,849 y el mínimo de 0,021.

En conclusión, creemos que se puede afirmar que en las dos primeras cohortes la estructura dimensional se mantiene a lo largo del tiempo, mientras que en la tercera es necesario un estudio más en profundidad de los cambios que se dan entre cada aplicación.

Por último, nos planteamos también si era posible realizar algún análisis que nos evaluase el grado de mantenimiento de esta estructura de orden en las cuatro aplicaciones y no sólo par a par. La técnica empleada ha sido el índice de concordancia *W* de Kendall. Un ejemplo de su uso para examinar la concordancia de rangos que varían en el tiempo lo encontramos en el trabajo de Baumgartner et al. (1999) que emplearon este procedimiento para evaluar conglomerados de imágenes de resonancia magnética.

En nuestro caso, y dadas las limitaciones computacionales de SPSS para las pruebas no paramétricas (un máximo de 500 variables), hemos procedido de la siguiente manera: en primer lugar hemos seleccionado para cada una de las tres cohortes una muestra aleatoria de 450 sujetos. Una vez transpuesta la matriz, disponemos de un fichero con 450 variables (los sujetos) y 12 casos (los jueces), en este caso considerando como tales los rangos globales de las tres puntuaciones (directa, factor uno y factor dos) de cada una de las cuatro aplicaciones. A continuación hemos realizado el contraste de la *W* para evaluar la concordancia entre los 12 jueces conjuntamente (las 12 puntuaciones) y los tres grupos de cuatro, una por aplicación (las cuatro puntuaciones directas, los cuatro primeros factores y los cuatro segundos). Dado el número de variables, los grados de libertad eran muy elevados por lo que los contrastes resultaban todos significativos (se contrasta mediante *chi* cuadrado). Para poder contrastar con grados de libertad menores, repetimos el proceso con muestras aleatorias de 30 y 50 sujetos. En la Tabla XII aparecen los valores de la *W* para cada cohorte, tamaño muestral y número de jueces considerados simultáneamente. Todos resultaron significativos a 0,0001.

Tabla XII. Valores de la *W* de Kendall

	Cohorte 5º-6º (N=4551)			Cohorte 1º-2º (N=5592)			Cohorte 3º-4º (N=5175)		
N muestral:	30	50	450	30	50	450	30	50	450
Todo (12)	,492	,544	,567	,519	,590	,591	,379	,295	,400
Punt. (4)	,627	,714	,696	,597	,677	,654	,496	,371	,469
Fact. 1 (4)	,496	,499	,559	,546	,659	,600	,415	,368	,444
Fact. 2 (4)	,510	,544	,565	,513	,523	,612	,373	,342	,443

Como anteriormente, los valores superiores de la *W* se dan cuando contrastamos la concordancia simultánea de las puntuaciones directas de las cuatro aplicaciones. Para las dos primeras cohortes tenemos valores comprendi-

dos entre 0,65 y 0,70 lo que refuerza la aseveración anterior de que la estructura se mantiene a lo largo del tiempo. También al igual que antes, para la tercera cohorte el índice de concordancia es menor.

CONCLUSIONES FINALES

A modo de resumen, podemos concluir lo siguiente:

- La mayoría de las pruebas tienen una estructura de una cierta complejidad y las que resultan ser de estructura más simple tienden a mayor multidimensionalidad.
- Esto nos reafirma en la idea de que, mayoritariamente, nos encontramos con una dimensión latente principal de la que participan la gran mayoría de los ítems y una secundaria.
- En las pruebas de estructura simple, podría hablarse probablemente de dos dimensiones más marcadas. En la primera seguirían cargando muchos ítems, y la segunda, más claramente definida, se aglutinaría alrededor de contenidos como la geometría, la medida o la estadística o por dominios cognitivos de aplicación o resolución de problemas.
- Para las tres cohortes y desde la perspectiva de la dimensionalidad, las formas A y B no resultan equivalentes.
- Desde una perspectiva longitudinal, en las dos primeras cohortes la estructura dimensional de las pruebas se mantiene básicamente a lo largo de las cuatro aplicaciones. No ocurre lo mismo con la tercera.
- La triangulación realizada mediante los procedimientos factoriales «clásicos» y el Análisis Factorial de Información Total ha resultado una estrategia metodológica válida que nos ha permitido acumular evidencia empírica que consideramos suficiente para apoyar las conclusiones anteriores.

Todos estos resultados confirman que el aumento de la complejidad de los contenidos implica un aumento en la complejidad de la estructura dimensional de las pruebas. En el futuro, siempre que la evaluación de valor añadido se refiera a un abanico de cursos tan amplio como el que aquí se ha estudiado, se recomienda tener muy en cuenta la posibilidad de contar con estructuras de dimensionalidad de complejidad creciente en paralelo con el desarrollo del currículo.

REFERENCIAS BIBLIOGRÁFICAS

- Abad, F. J., Ponsoda, V. y Revuelta, J. (2006). *Modelos politómicos de respuesta al ítem*. Madrid: La Muralla.
- Antillón, Luz Elena, Larrazolo, Norma y Backhoff, Eduardo (2006). Igualación equipercantil del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *RELIEVE*, v. 12, n. 2. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_2.htm. Consultado en 10 de Octubre de 2008.
- Baumgartner, R., Somorjai, R., Summers, R. & Richter, R. (1999). Assessment of Cluster Homogeneity in Fmri Data Using Kendall's Coefficient of Concordance. *Magnetic Resonance Imaging*, 17, 10, 1525–1532.
- Bayot, Agustín, Hernández Viadel, José Vicente y De Julián, Luis Felipe (2005). Análisis factorial exploratorio y propiedades psicométricas de la escala de competencia parental percibida. Versión para padres/madres (ECP-p). *RELIEVE*, v. 11, n. 2. http://www.uv.es/RELIEVE/v11n2/RELIEVEv11n2_2.htm. Consultado en 10 de Octubre de 2008.
- Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459.
- Bock, R. D., Gibbons, R. D., & Muraki, E. (1988). Full information item factor analysis. *Applied Psychological Measurement*, 12, 261-280.
- Bock, R. D., Gibbons, R., Schilling, S. G., Muraki, E., Wilson, D. T. & Wood, R. (2003). *TESTFACT 4.0, Computer software and manual*. Lincolnwood. Scientific Software International.
- Finch, H. (2006). Comparison of the Performance of Varimax and Promax Rotations: Factor Structure Recovery for Dichotomous Ítems. *Journal of Educational Measurement*, 43, 39–52.
- Fraser, C. & McDonald, R. P. (1988). NOHARM II. Least-Squares item factor analysis. *Multivariate Behavioral Research*, 23, 267-269.

Gaviria, J.L. (1990). Factores de dificultad en el análisis de ítems. Qué son, por qué aparecen y posibles soluciones. *Revista Complutense de Educación*, 1, 1, 95-98.

Gaviria, J.L. y Ruiz de Miguel, C. (2007). Importancia de algunos supuestos psicométricos en la evaluación de los sistemas educativos. Calibración y equiparación en las pruebas de Estándares Nacionales de México. *Revista de Educación*, 343, 223-248.

Lizasoain, L. y Joaristi, L. (en prensa). Análisis de la dimensionalidad en modelos de valor añadido: estudio de las pruebas de matemáticas empleando métodos no paramétricos basados en TRI. Artículo aceptado para su publicación en el número especial de la *Revista de Educación* sobre valor añadido. Fecha prevista de publicación: Enero 2009.

Martínez, M. R., Hernández, M. J. y Hernández, M. V. (2006). *Psicometría*. Madrid: Alianza.

Muraki, E., & Engelhard, G. (1985). Full information item factor analysis: applications of EAP scores. *Applied Psychological Measurement*, 9, 417-430.

Muthén, B. & Hofacker, C. (1988). Testing The Assumptions Underlying Tetrachoric Correlations. *Psychometrika*, 53, 563-578.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory*. New Cork: McGraw Hill.

Stevens, J. (2002). *Applied Multivariate statistics for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Stout, W., Nandakumar, R. & Habing, R. (1996) Analysis Of Latent Dimensionality Of Dichotomously And Polytomously Scored Test Data. *Behaviormetrika*, 23, 1, 37-65.

Takane, Y., De Leeuw, J. (1987) On The Relationship Between Item Response Theory And Factor Analysis Of Discretized Variables. *Psychometrika*, 52, 3, 393-408

Walker, C.M., Azen, R. & Schmitt, T. (2006). Statistical Versus Substantive Dimensionality. The Effect of Distributional Differences on Dimensionality Assessment Using DIMTEST. *Educational and Psychological Measurement*. 66, 5, 721-738.


Waller, N. G. (2002). *MicroFACT for Windows- Factor Analysis for Dichotomous and Ordered Polytomous Response Data*. St. Paul, MN: Assessment Systems Corporation.


Wilson, D. T., Word, R. & Gibbons, R. (1998). *TESTFACT. Test scoring, Item Statistics and Item Factor Analysis*. Chicago: Scientific Software Internacional, Inc.

NOTAS

[1] En negrita se representan vectores y matrices.

ABOUT THE AUTHORS / SOBRE LOS AUTORES

Joaristi, Luis (luis.joaristi@ehu.es). Ingeniero y Profesor Titular del Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universidad del País Vasco-Euskal Herriko Unibertsitatea. Avenida de Tolosa, 70. 20018 Donostia - San Sebastián (España). Entre sus principales líneas de investigación destacan el análisis de datos aplicado a la investigación educativa y la evaluación de sistemas educativos. [Buscar otros artículos de este autor en Scholar Google](#) 

Lizasoain, Luis (luis.lizasoain@ehu.es). Profesor Titular del Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universidad del País Vasco-Euskal Herriko Unibertsitatea. Avenida de Tolosa, 70. 20018 Donostia - San Sebastián (España). Entre sus principales líneas de investigación destacan el análisis de datos aplicado a la investigación educativa y la evaluación de sistemas educativos. [Buscar otros artículos de este autor en Scholar Google](#) 

Joaristi, L. y Lizasoain, L. (2008). Estudio de la dimensionalidad empleando análisis factorial clásico y análisis factorial de información total: Análisis de pruebas matemáticas de primaria (5º y 6º cursos) y secundaria obligatoria. *RELIEVE*, v. 14, n. 2, p. 1-18. http://www.uv.es/RELIEVE/v14n1/RELIEVEv14n2_2.htm

ARTICLE RECORD / FICHA DEL ARTÍCULO

Reference / Referencia	Joaristi, Luis y Lizasoain, Luis (2008). Estudio de la dimensionalidad empleando análisis factorial clásico y análisis factorial de información total: análisis de pruebas de matemáticas de primaria (5º y 6º cursos) y secundaria obligatoria. <i>RELIEVE</i> , v. 14, n. 2. http://www.uv.es/RELIEVE/v14n2/RELIEVEv14n2_2.htm . Consultado en (<i>poner fecha</i>).
Title / Título	Estudio de la dimensionalidad empleando análisis factorial clásico y análisis factorial de información total: análisis de pruebas de matemáticas de primaria (5º y 6º cursos) y secundaria obligatoria. [<i>Study of dimensionality using classic factor analysis and full information factor analysis: analysis of primary (5th and 6th degrees) and compulsory secondary mathematics tests</i>]
Authors/Autores	Joaristi, Luis y Lizasoain, Luis
Review / Revista	Revista ELectrónica de Investigación y EValuación Educativa (RELIEVE), v. 14, n. 2
ISSN	1134-4032
Publication date / Fecha de publicación	2008 (Reception Date : 2008 October 20; Approval Date : 2008 October25; Publication Date : 2008 October 30).
Abstract / Resumen	<p><i>The aim of this paper is to analyze the dimensional structure of a set of mathematical achievement tests used to assess the academic achievement in the Community of Madrid during 2005-06 and 2006-07 years and in three cohorts corresponding to the following academic degrees: 5th-6th of primary education, 1st-2nd and 3-4th of compulsory secondary education (ESO). The results coming from classical factor analysis and full information factor analysis are compared. Therefore a longitudinal study is performed in order to test whether such structure remains constant along time. The results confirm that most of the tests are mostly unidimensional structures. Meanwhile, it was proven that as the complexity of the contents increase, the complexity of the dimensional structure tests increase as well.</i></p> <p>El objetivo de este trabajo es analizar la estructura dimensional de las pruebas empleadas para evaluar el rendimiento académico en la asignatura de matemáticas en el contexto de la evaluación realizada en la Comunidad de Madrid en los cursos académicos 2005-06 y 2006-07 en tres cohortes: 5º y 6º de Educación Primaria, 1º-2º y 3º-4º de ESO. Para ello, se triangulan los resultados obtenidos de la aplicación de las técnicas factoriales clásicas con los del Análisis Factorial de Información total. Asimismo, se realiza un estudio longitudinal con objeto de ver si dicha estructura se mantiene constante a lo largo del tiempo. Los resultados confirman estructuras esencialmente unidimensionales. A su vez, se comprueba que el aumento de la complejidad de los contenidos implica un aumento en la complejidad de la estructura dimensional de las pruebas.</p>
Keywords / Descriptores	<p><i>Dimensionality, primary education, compulsory secondary education, mathematics achievement tests, factor analysis, full information factor analysis.</i></p> <p>Dimensionalidad, educación primaria, ESO, pruebas de rendimiento de matemáticas, análisis factorial, Análisis Factorial de Información Total.</p>
Institution / Institución	Universidad del País Vasco-Euskal Herriko Unibertsitatea (España).
Publication site / Dirección	http://www.uv.es/RELIEVE
Language / Idioma	Español (Title, abstract and keywords in English)

Revista ELectrónica de Investigación y EValuación Educativa (RELIEVE)

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).