

# BORDÓN

## Revista de Pedagogía

NÚMERO MONOGRÁFICO / *SPECIAL ISSUE*

Rendimiento en matemáticas y la ciencia de la educación  
matemática: evidencia de diferentes naciones  
*Mathematics achievement and the science of mathematics  
education: evidence from different nations*

María Inés Susperreguy, Blanca Arteaga Martínez y Elida V. Laski  
(editores invitados / *guest editors*)



Volumen 70  
Número, 3  
2018

**SOCIEDAD ESPAÑOLA DE PEDAGOGÍA**

# ADAPTACIÓN DE LA PRUEBA TIMSS PARA LA EVALUACIÓN DE LA COMPETENCIA MATEMÁTICA EN ALUMNOS DE MAGISTERIO

## *Adapting the test of the Trends in International Mathematics and Science Study (TIMSS) for the assessment of the mathematical competence of students entering the undergraduate degree of Primary Education School Teacher*

BLANCA ARTEAGA MARTÍNEZ<sup>(1)</sup>, ENRIQUE NAVARRO ASENCIO<sup>(2)</sup>, ARÁNTZAZU FRAILE REY<sup>(1)</sup>  
Y PEDRO RAMOS ALONSO<sup>(1)</sup>

<sup>(1)</sup> Universidad de Alcalá

<sup>(2)</sup> Universidad Complutense de Madrid

DOI: 10.13042/Bordon.2018.63042

Fecha de recepción: 30/01/2018 • Fecha de aceptación: 12/03/2018

Autora de contacto / Corresponding author: Blanca Arteaga Martínez. E-mail: blanca.arteaga@uah.es

---

**INTRODUCCIÓN.** La base del conocimiento pedagógico que un maestro debe tener respecto a las matemáticas se sustenta de una manera especial en el conocimiento que tiene de los contenidos de la materia. El objetivo de este trabajo es la elaboración de una prueba de diagnóstico de habilidades matemáticas, a partir de la selección y adaptación de algunos de los ítems del *Trends in International Mathematics and Science Study* (TIMSS) de 2011, así como los aspectos técnicos desde la aplicación del modelo de Rasch. **MÉTODO.** Los ítems de la prueba elaborada se seleccionaron siguiendo la estructura original de bloques de contenido y dominio cognitivo establecida en el marco de TIMSS 2011 (Mullis, Martin, Ruddock, O'Sullivan y Preuschoff, 2009). El procedimiento de traducción y adaptación se ha desarrollado con el objetivo de garantizar una mayor validez lingüística. El estudio se ha realizado con una muestra de 477 estudiantes del primer curso de universidades de la Comunidad de Madrid (España). Estos estudiantes estaban matriculados en el primer año de los diferentes grados de Magisterio de Educación Primaria. **RESULTADOS.** Los 30 ítems seleccionados han sido de dos tipologías, ítems de respuesta seleccionada e ítems de respuesta construida. Dadas estas características particulares en el formato ha sido necesario estudiar la fiabilidad de la corrección realizada por los evaluadores para garantizar la concordancia en su corrección. **DISCUSIÓN.** El modelo de Rasch utilizado para la validación de la prueba nos ha permitido generar evidencias que demuestran un alto grado de confiabilidad, señalando las diferencias de comportamiento de los ítems de respuesta seleccionada y de respuesta construida (elementos que requieren que los estudiantes construyan su propia respuesta escrita). El modelo nos ha permitido una clasificación de los resultados de los ítems por nivel de dificultad, hecho que facilita una interpretación de los resultados de acuerdo a los dominios de contenido y cognitivo.

**Palabras clave:** TIMSS, Competencia matemática, Modelo de Rasch, Adaptación de pruebas, Formación docente inicial.

---

## Introducción

La evaluación diagnóstica es una herramienta crucial para planificar y diseñar programas de formación efectivos. El objetivo y diseño de este tipo de evaluaciones se centra en la recogida y estudio de información sobre el grado de desarrollo de las aptitudes del alumnado con el fin de conocer, pronosticar y tomar decisiones que favorezcan el pleno desarrollo educativo de los estudiantes. Este tipo de evaluación toma una relevancia especial en un entorno donde el objetivo es alcanzar determinadas competencias (Cano, 2008). Además, en las recomendaciones dadas por la NCTM (1991) para la evaluación en matemáticas se contempla de manera específica el diagnóstico en términos de lo que comprende el estudiante.

La utilización de evaluaciones diagnósticas nos permite adaptar los procesos de enseñanza-aprendizaje a las características particulares de cada una de las personas y, en consecuencia, potenciar la mejora del elemento evaluado (Pérez y Soto, 2011; Tourón, 2009).

Para establecer el significado de ser matemáticamente competente, vamos a referirnos a las aportaciones de Llinares (2003: 14) cuando dicta las necesidades para un estudiante de primaria y que podemos generalizar, en este caso, a sus maestros al señalar que “debe relacionarse con ser capaz de realizar determinadas tareas matemáticas y comprender por qué pueden ser utilizadas algunas nociones y procesos para resolverlas, así como la posibilidad de argumentar la conveniencia de su uso”. Estas habilidades tienen una elevada complejidad para ser evaluadas en cuanto a los niveles de desarrollo que pueden tener, por lo que resultaría complejo recogerlas en toda su magnitud.

Evaluar cada una de estas habilidades en un solo instrumento se hace difícil, pues cada una de ellas es compleja en sí misma y el nivel de desarrollo alcanzado por un sujeto está ligado al aprendizaje progresivo del individuo.

La evaluación diagnóstica bien planteada debe “evaluar habilidades cognitivas de alto nivel (como son planificación, estructuración de tareas, obtención de información, construcción de respuesta, explicación de procesos, integración de conocimientos e informaciones)” (Castro, 2011: 111), donde la tipología de ítems que se utilicen debe tener distintas características para conseguir evaluar capacidades más complejas.

Para elegir un instrumento adecuado para utilizar en un contexto de maestros en formación acudimos a las referencias especializadas, lo que nos facilitó una cantidad considerable de resultados referidos a la evaluación de creencias y actitudes hacia las matemáticas (White, Way, Perry y Southwell, 2005; Hodgen y Askew, 2007; entre otros). No obstante, son menos las investigaciones relacionadas con el nivel de logro en estudiantes de magisterio (Tatto *et al.*, 2008; Sandoval, Frit, Maldonado y Rodríguez, 2010; Nortes y Nortes Checa, 2017). Sin embargo, parece importante que el maestro domine los contenidos que enseña, para respaldar explicaciones o diagnóstico de obstáculos de aprendizaje, aportando así la base para el conocimiento pedagógico (Varas, Lacourly, López y Giacono, 2013; Ernest, 1989).

Así, el objetivo principal de este trabajo es la construcción y validación de un instrumento de medida de la competencia matemática basado en la evaluación TIMSS 2011. Para ello como objetivos específicos se plantean:

- Seleccionar y adaptar al castellano ítems de la evaluación TIMSS 2011.
- Analizar la fiabilidad de los ítems de respuesta construida cuya puntuación es determinada por un evaluador.
- Validar el instrumento construido bajo los supuestos de la Teoría Respuesta al Ítem (TRI).

El diseño de la prueba de diagnóstico de este estudio se centra en las habilidades de conocimiento y razonamiento matemático que son las

que están íntimamente ligadas a la capacidad de resolver problemas matemáticos. El marco teórico de TIMSS 2011 vincula el éxito en la resolución de problemas con la comprensión conceptual y el razonamiento (identificar, inferir y reorganizar la información), y es por ello que los ítems planteados en esta prueba se consideran adecuados para definir una primera evaluación de diagnóstico, cuyo objetivo es medir el desempeño matemático con el que los alumnos acceden al Grado de Magisterio en Educación Primaria. Se opta por este método de construcción de la prueba, ante la ausencia de pruebas específicas para este tipo de población, unido a la buena calidad psicométrica de la prueba TIMSS.

Los ítems son de dos tipologías distintas: ítems de respuesta seleccionada e ítems de respuesta construida. Estas dos tipologías facilitan la recogida de datos no únicamente referida a la demanda cognitiva, sino al rango de cognición que implica la resolución de cada uno de los ítems (Martínez, 1999).

## **Método**

El objetivo del instrumento es la medición del desempeño matemático en estudiantes que acaban de acceder al Grado de Magisterio de Educación Primaria. Se trata de una prueba diagnóstica de corte transversal.

Los ítems de la prueba pertenecen a la prueba TIMSS 2011 de rendimiento en matemáticas para 8º grado (en España, 2º de ESO). La estructura de esta prueba internacional responde a dos dimensiones, una de contenido y otra cognitiva (Mullis *et al.*, 2009). La dimensión de contenido especifica cuatro dominios: álgebra, números, datos-azar y geometría. La dimensión cognitiva describe tres procesos a evaluar: conocer, aplicar y razonar.

Se realiza una selección, con un diseño donde todos los estudiantes respondan a los mismos

ítems (Fernández-Alonso y Muñiz, 2011) y que tengan elementos con distinta dificultad pertenecientes a todos los dominios cognitivos, respetando la estructura original de la prueba en cuanto al porcentaje de ítems de cada uno de los dominios de contenido. Para la selección se tiene en cuenta la media del número de respuestas correctas a nivel internacional de cada ítem en la prueba original, seleccionando aquellos ítems que corresponden a los tres cuartiles, en el caso de que se seleccionen tres ítems de esa dimensión, o los dos cuartiles superiores, en caso de que se seleccionen dos. Se realiza de esta forma debido a la mayor capacidad informativa que tienen los ítems con una dificultad media (Martínez, Hernández y Hernández, 2006). El número de ítems de cada contenido, dominio cognitivo y dificultad se muestran en la tabla 1.

Se utiliza la versión original en inglés para realizar la traducción, intentando evitar errores como los señalados por Solano, Contreras y Backhoff (2006) al referirse a traducciones anteriores para este tipo de pruebas internacionales.

La validación de contenido sirve para admitir parte de un instrumento que fue construido para una población de segundo curso de Educación Secundaria, y que ahora queremos utilizar para estudiantes que acceden al grado de maestro. Este primer proceso tras la selección de ítems, requiere un proceso de traducción, adaptación y estandarización (Hyrkäs, Appelqvist-Schmidlechner y Oksa, 2003).

El procedimiento de traducción se realiza en dos fases para dar una mayor validez lingüística de partida; primero se realiza la traducción al castellano de la prueba y posteriormente una persona distinta que no conocía la versión original realiza la traducción de manera inversa (Beaton, Bombardier, Guillemin y Bosi, 2000). Ambas versiones son revisadas por un equipo de jueces expertos (Escobar-Pérez y Cuervo-Martínez, 2008) en el área de didáctica de la matemática para confirmar la equivalencia de los resultados.

TABLA 1. Distribución de ítems por dominios de contenido y cognitivo, por nivel de desempeño

	Nivel de desempeño		Dominio de contenido				Total
			Álgebra	Números	Datos y azar	Geometría	
Bajo	Dominio cognitivo	Conocimiento			1		1
	Total				1		1
Intermedio	Dominio cognitivo	Conocimiento	1	1		1	3
		Aplicación			1		1
		Razonamiento	1				1
	Total		2	1	1	1	5
Alto	Dominio cognitivo	Conocimiento	1	2	1		4
		Aplicación	1	3	1	1	6
		Razonamiento	1	1	1	1	4
	Total		3	6	3	2	14
Avanzado	Dominio cognitivo	Conocimiento	1			1	2
		Aplicación	1			1	2
		Razonamiento	1	2	1	1	5
	Total		3	2	1	3	9
Sobreavanzado	Dominio cognitivo	Aplicación	1				1
	Total		1				1
Total	Dominio cognitivo	Conocimiento	3	3	2	2	10
		Aplicación	3	3	2	2	10
		Razonamiento	3	3	2	2	10
	Total		9	9	6	6	30

Nota: la especificación de los dominios en cada uno de los ítems se puede consultar en el anexo 1.

La validación de la traducción se lleva a cabo a través de un juicio de expertos. Para ello, se selecciona una muestra intencional de diez jueces expertos en didáctica de las matemáticas en activo en distintas universidades españolas; para esta selección se tiene en consideración la experiencia docente en los grados de maestro, su conocimiento del sistema educativo y de evaluaciones internacionales de matemáticas. Se pide que evalúen el criterio de claridad de la redacción (se comprende desde el punto de vista sintáctico y semántico) (Quesada, Rodríguez-Gómez e Ibarra, 2013); además de un aporte cualitativo que indica fundamentalmente aspectos de redacción o términos matemáticos sinónimos a los empleados que puedan aumentar el aspecto comprensivo.

La consistencia interjueces se analiza mediante el análisis de concordancia basado en el coeficiente W de Kendall, un coeficiente de acuerdo para datos ordinales basado en el grado de varianza de la suma de rangos obtenidos de los distintos jueces (Siegel y Castellan, 1995). Los valores a partir de 0.4 se consideran aceptables (Landis y Koch, 1977).

Una vez elaborada la prueba y después de su aplicación se selecciona una muestra de diez cuadernillos que habían sido corregidos por los mismos dos correctores y se estudia la fiabilidad de la corrección con el estadístico Kappa de Cohen. Para este estadístico, los valores cercanos a 1 señalan un alto grado de acuerdo (Viera y Garrett, 2005) entre evaluadores.

Finalmente, para la validación del instrumento de medida y la estimación de puntuaciones de rendimiento se emplea la Teoría Respuesta al Ítem siguiendo los supuestos del modelo de Rasch (1960) y la adaptación de Masters (1982) para ítems de respuesta construida (Modelo de Crédito Parcial, MCP). Estos modelos asumen que la puntuación del sujeto en el instrumento de medida, entendida como suma total de las respuestas de un sujeto al conjunto de ítems, y la puntuación en el ítem (dificultad), entendida como la suma de las respuestas de los sujetos en un ítem, son estadísticos suficientes para estimar los parámetros del modelo. Al contrario de lo que ocurre en los modelos de Thurstone, que estiman el mismo valor del rasgo para los sujetos que obtienen la misma puntuación en el test.

En el modelo de Rasch la función de probabilidad de respuesta de un sujeto a un determinado reactivo depende únicamente de estos estadísticos suficientes (rasgo estimado y dificultad del ítem). En el caso de ítems dicotómicos es:

$$P_i(\theta) = \frac{\exp^D(\theta - b_i)}{1 + \exp^D(\theta - b_i)}$$

$P_i(\theta)$ : Probabilidad de acertar el ítem  $i$  para un valor determinado de habilidad. En modelos politómicos, como el MCP, es la probabilidad de superar una categoría del ítem.  $b_i$  es el parámetro de dificultad, o el nivel de rasgo necesario para responder correctamente al ítem, es decir, el valor de la habilidad en el que la probabilidad de acertar el ítem supera el 0.5. En los modelos politómicos es el nivel de rasgo necesario para superar ese paso o categoría dentro del ítem.  $\theta$  es el valor del rasgo estimado y  $D$  es una constante, igual a 1.7, que aproxima los valores a la distribución normal. La probabilidad de acertar o no un ítem depende de estos dos parámetros y toma la siguiente forma para valores de  $x$  igual a 0 y 1.

En los modelos con más de una posibilidad de puntuación, no solo acierto o error, es necesario

estimar una curva de probabilidad para cada puntuación en el ítem porque no supone el mismo nivel de dificultad obtener los distintos valores en ese ítem. En el modelo de crédito parcial la probabilidad de que el sujeto complete cada paso o categoría del ítem puede explicitarse mediante un modelo de Rasch. En nuestro caso, un ítem con tres categorías ( $R$  es el total de categorías) de respuesta (0, 1, 2) quedaría formulado de la siguiente forma:

$$P(r) = \frac{\exp(\sum_{r=0}^R(\theta - \beta_k))}{\sum_{h=0}^2 \exp(\sum_{r=0}^R(\theta - \beta_k))}$$

Siendo  $\sum_{r=0}^0(\theta - \beta_k) = 0$

La calidad de los ítems se analiza a partir de diferentes indicadores:

- La curva característica del ítem (CCI): describe los cambios en el nivel del rasgo relacionados con cambios en la probabilidad de responder a un ítem. Por tanto, sujetos con diferentes niveles de rasgo tendrán distinta probabilidad de acertar el ítem. En ítems de respuesta seleccionada con una única respuesta correcta predice la probabilidad de éxito en el ítem a partir del nivel de rasgo que posee el sujeto. En ítems el caso de los politómicos (respuesta construida) predice la probabilidad de las respuestas en cada categoría a partir del nivel de rasgo.
- La curva característica del test: relaciona las puntuaciones estimadas con el modelo de Rasch con la suma de respuestas correctas.
- Los estadísticos de ajuste de los ítems basados en los residuos como diferencia media al cuadrado (MNSQ) entre la respuesta empírica y las probabilidades esperadas que representan los gráficos CCI:
  - MNSQ (sin ponderar): no pondera los residuos, por lo que representa el

ajuste externo al ser sensible al comportamiento inesperado de ítems cuya pendiente se aleja del nivel de habilidad del sujeto, es decir, sujetos con una habilidad alta que valoran categorías bajas o al revés.

- MNSQ (ponderado): es el mismo índice pero ponderado por la cantidad de información de un ítem en el intervalo de habilidad. Se corresponde con el ajuste interno, se adecúa a ítems con pautas de respuesta irregulares implícitas en las personas y viceversa porque las personas con un nivel de rasgo cercano a la dificultad influyen más en el residual.

Un buen ajuste se obtiene cuando sus valores se aproximan a 1, pero los valores son aceptables mientras no excedan los límites del intervalo de confianza. El estadístico también se presenta de forma transformada a distribución normal como prueba de hipótesis T (valores por encima de 2 señalarán malos ajustes) (Navarro, Expósito, López y Thoilliez, 2014).

## Resultados

En primer lugar, el estudio de la claridad de los ítems por parte de los expertos obtiene, como muestra la tabla 2, un índice de concordancia aceptable ( $W=.414$ ,  $p=.000$ ). No obstante, se consideran las observaciones cualitativas aportadas para mejorar la traducción de los ítems.

TABLA 2. Resultados juicio de expertos

	Claridad
N	20
W de Kendall	.414
Chi-cuadrado	49.682
G1	6
Sig. asintót.	.000

En segundo lugar, el estudio de fiabilidad de la corrección que se lleva a cabo muestra un acuerdo perfecto entre evaluadores, con valores de Kappa iguales a uno en todos los ítems analizados.

Por último, para la validación el instrumento, en las tablas 3 y 4 se presentan los índices de ajuste para ítems de respuesta seleccionada y construida respectivamente. Como se observa, el ajuste ponderado obtiene valores aceptables en todos los ítems de respuesta seleccionada, con valores de MNSQ cercanos a 1 y valores de T por debajo de 2. Si se observan los parámetros de dificultad (B) se comprueba que hay muchos valores negativos, lo que indica que hay muchos ítems que se podrían considerar fáciles o, en otras palabras, que el nivel de rendimiento en matemáticas que se necesita para contestar correctamente a ese ítem está por debajo de la media. El ítem 1 (p66) es el más fácil. De la misma forma los ítems 7 (p72) y 9 (p74) son los más difíciles, es decir, el nivel de rendimiento en matemáticas que necesita un estudiante para responder correctamente está por encima de la media. Conviene recordar que los resultados estimados inicialmente se encuentran en una escala tipificada.

Sin embargo, no ocurre lo mismo para los de respuesta construida, el ítem 4 (p69) parece presentar problemas de ajuste como indican los valores de T por encima de 2. Tampoco parece que haya una buena estimación de los parámetros de dificultad en los ítems, 4 (p69), 8 (p73) y 10 (p75), donde las categorías de respuesta más bajas necesitan un mayor nivel de rendimiento que las superiores.

En cualquier caso, que un ítem no ajuste completamente en el modelo de Rasch no significa que esté funcionando mal en este caso particular. Este tipo de análisis necesita grandes muestras para aumentar la fiabilidad de las estimaciones y, en los ítems de respuesta construida, tener representadas las distintas categorías de valoración del ítem. Para comprobar el funcionamiento de

las puntuaciones estimadas se han calculado los promedios en matemáticas de los estudiantes que estaban en cada uno de los niveles de respuesta de

esos ítems (tablas 9 a 11) y se observa que unos mejores resultados se corresponden con una categoría más alta en el reactivo.

**TABLA 3. Índices de ajuste de los ítems de respuesta seleccionada de matemáticas**

Ítem	B	ERROR	Ajuste sin ponderar			Ajuste ponderado		
			MNSQ	CI	T	MNSQ	CI	T
1	-4.102	0.313	0.56	(0.87, 1.13)	-8.1	0.97	(0.44, 1.56)	0
2	-2.615	0.176	0.82	(0.87, 1.13)	-2.9	0.96	(0.75, 1.25)	-0.3
3	-2.673	0.175	0.83	(0.87, 1.13)	-2.8	0.95	(0.75, 1.25)	-0.3
5	-1.084	0.121	0.96	(0.87, 1.13)	-0.6	0.99	(0.90, 1.10)	-0.2
6	-1.844	0.14	0.94	(0.87, 1.13)	-0.9	1	(0.84, 1.16)	0
7	0.929	0.115	0.88	(0.87, 1.13)	-1.9	0.91	(0.91, 1.09)	-2
9	0.751	0.114	1.06	(0.87, 1.13)	1	1.07	(0.92, 1.08)	1.5
11	-2.478	0.165	0.91	(0.87, 1.13)	-1.4	0.98	(0.77, 1.23)	-0.2
13	-2.943	0.194	1.05	(0.87, 1.13)	0.8	0.99	(0.70, 1.30)	0
14	-0.269	0.111	0.93	(0.87, 1.13)	-1	0.94	(0.93, 1.07)	-1.6
15	-1.053	0.119	1.03	(0.87, 1.13)	0.4	1.02	(0.90, 1.10)	0.4
16	-3.160	0.212	1.04	(0.87, 1.13)	0.7	0.96	(0.66, 1.34)	-0.2
17	-0.911	0.117	1	(0.87, 1.13)	0	0.99	(0.91, 1.09)	-0.1
18	-2.683	0.183	0.82	(0.87, 1.13)	-2.9	0.97	(0.73, 1.27)	-0.2
19	-3.246	0.22	1.33	(0.87, 1.13)	4.6	1.01	(0.65, 1.35)	0.1
20	-0.835	0.114	1	(0.87, 1.13)	0	1.01	(0.91, 1.09)	0.3
21	-1.740	0.135	1.11	(0.87, 1.13)	1.6	1.07	(0.85, 1.15)	0.9
22	-0.133	0.109	0.95	(0.87, 1.13)	-0.8	0.96	(0.93, 1.07)	-1.1
23	-2.002	0.144	1.04	(0.87, 1.13)	0.6	0.98	(0.83, 1.17)	-0.2
24	-0.255	0.108	1.15	(0.87, 1.13)	2.2	1.11	(0.93, 1.07)	3
25	-0.738	0.117	0.9	(0.87, 1.13)	-1.5	0.94	(0.91, 1.09)	-1.4
26	-4.017	0.313	0.8	(0.87, 1.13)	-3.3	0.98	(0.44, 1.56)	0
27	-1.164	0.119	0.97	(0.87, 1.13)	-0.4	0.98	(0.90, 1.10)	-0.4
28	-2.314	0.162	0.87	(0.87, 1.13)	-2.1	0.98	(0.79, 1.21)	-0.1
29	-1.666	0.132	1.05	(0.87, 1.13)	0.7	1	(0.86, 1.14)	0.1
30	-0.768	0.115	0.92	(0.87, 1.13)	-1.3	0.95	(0.91, 1.09)	-1.2

**TABLA 4. Índices de ajuste de los ítems de respuesta construida de matemáticas**

Cat	B	ERROR	Ajuste sin ponderar			Ajuste ponderado		
			MNSQ	CI	T	MNSQ	CI	T
4	0		1.54	(0.87, 1.13)	6.8	1.28	(0.89, 1.11)	4.5
4	1	1.990	1.05	(0.87, 1.13)	0.7	1.02	(0.60, 1.40)	0.1
4	2	-1.990*	1.58	(0.87, 1.13)	7.2	1.35	(0.89, 1.11)	5.7

**TABLA 4. Índices de ajuste de los ítems de respuesta construida de matemáticas (cont.)**

Cat	B	ERROR	Ajuste sin ponderar			Ajuste ponderado			
			MNSQ	CI	T	MNSQ	CI	T	
8	0		1.22	(0.87, 1.13)	3	1.14	(0.75, 1.25)	1.1	
8	1	0.941	0.193	1.13	(0.87, 1.13)	1.9	1.01	(0.72, 1.28)	0.1
8	2	-0.941*		1.22	(0.87, 1.13)	3.1	1.09	(0.83, 1.17)	1.1
10	0			1.05	(0.85, 1.15)	0.6	0.99	(0.88, 1.12)	-0.1
10	1	0.286	0.152	1.03	(0.85, 1.15)	0.4	0.98	(0.83, 1.17)	-0.2
10	2	-0.286*		1.12	(0.85, 1.15)	1.6	1.09	(0.81, 1.19)	1
12	0			1.38	(0.87, 1.13)	5.1	0.97	(0.63, 1.37)	-0.1
12	1	-0.149	0.167	0.94	(0.87, 1.13)	-1	0.99	(0.84, 1.16)	-0.1
12	2	0.149*		0.94	(0.87, 1.13)	-1	0.97	(0.86, 1.14)	-0.4

Para observar con más detalle los ítems de respuesta construida se presentan en la figura 1 los gráficos CCI (Curva Característica del Ítem) y las proporciones de casos situados en cada categoría en las tablas 5 a 8. Estos gráficos muestran las probabilidades de situarse en cada uno de los niveles de respuesta del ítem (eje y) en función del nivel de resultados de matemáticas (eje X). Por ejemplo, en el gráfico del ítem 4 (p69), se observa que hay mayor probabilidad de obtener un valor 0 si se obtiene una puntuación baja en el resultado, sin embargo, para obtener un valor 2 (valor más alto) necesitas unos mejores resultados. No

obstante, en el valor 1 la probabilidad es similar a lo largo de todos los valores del resultado. Este aspecto puede llevar a un mal ajuste del ítem y puede ser debido por la poca variabilidad de respuestas en esa categoría, como demuestra la tabla 4, solo hay 21 casos en la categoría 1.

En el ítem 8 (p73), si ponemos atención el gráfico de CCI, se observa que la probabilidad de obtener un valor 2 aumenta a medida que aumenta la puntuación en matemáticas. Sin embargo, hay un grupo de casos que también obtienen esa puntuación y finalmente no tienen una buena

**FIGURA 1. Gráficos CCI de los ítems de respuesta construida de matemáticas**

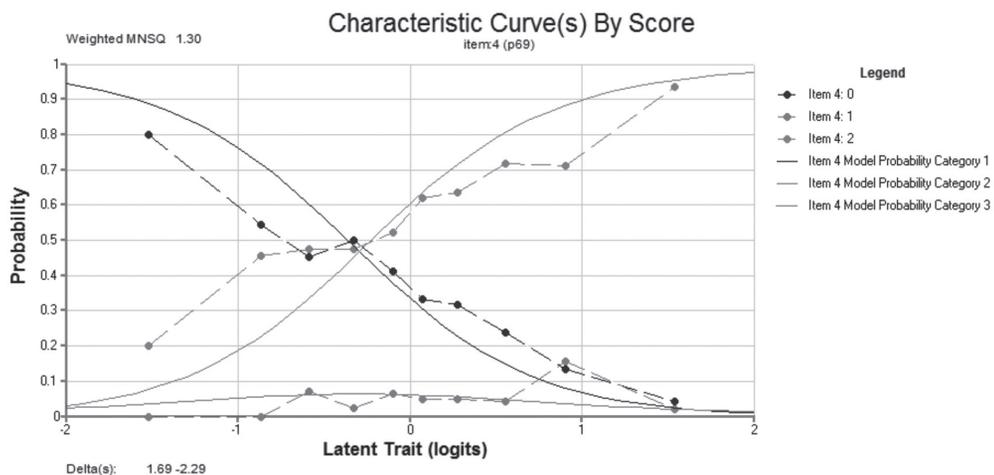
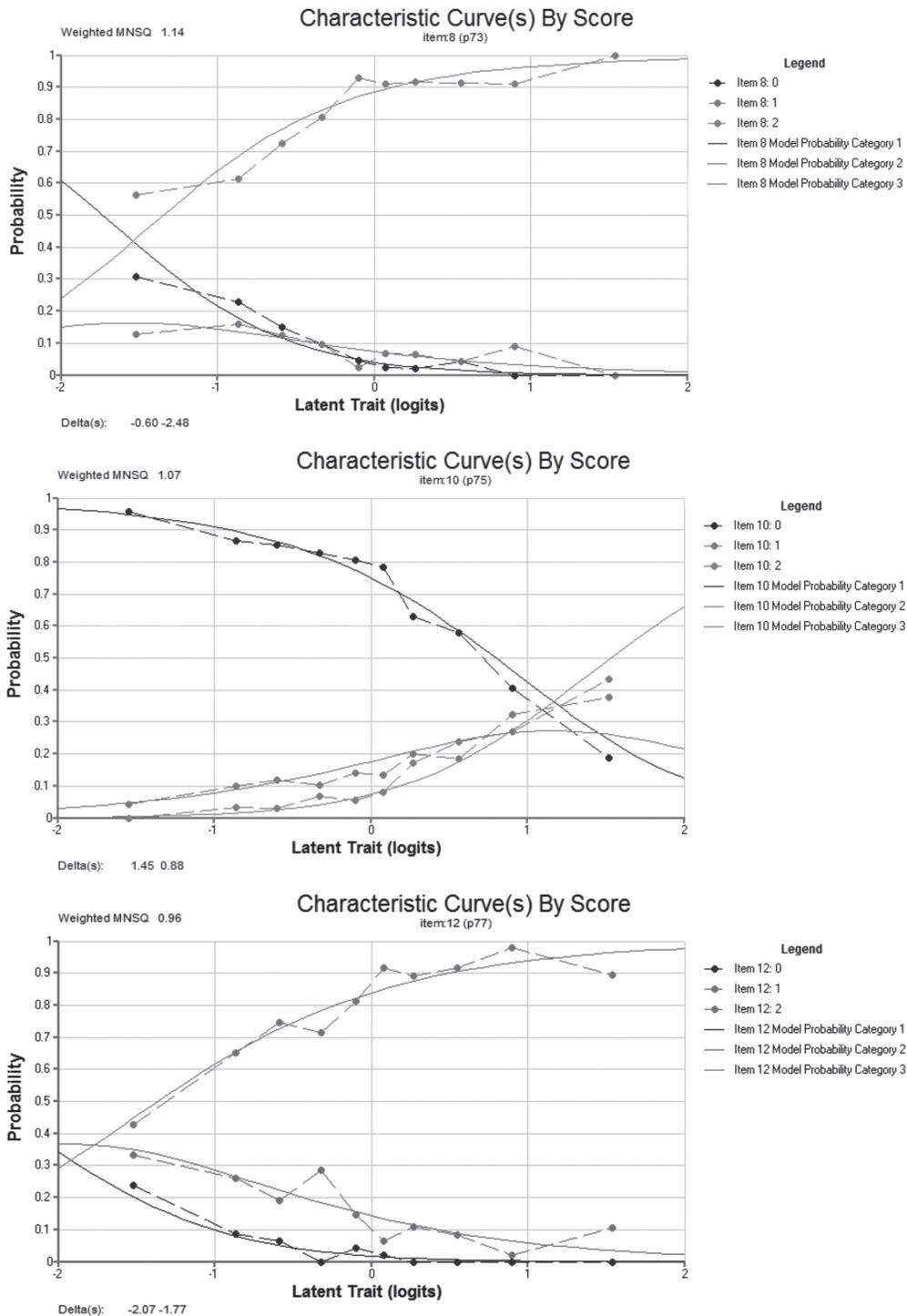


FIGURA 1. Gráficos CCI de los ítems de respuesta construida de matemáticas (cont.)



puntuación final. Se debe a que la mayor parte de los casos se sitúan en esa categoría del ítem, lo que indica una baja variabilidad en las respuestas.

En el ítem 10 (p75) ocurre lo opuesto, la mayoría obtienen un valor 0, y la tendencia es que la probabilidad disminuya a medida que se obtienen mejores resultados en matemáticas, pero incluso con puntuaciones por encima de la media puedes obtener un valor 0 en este ítem.

En las tablas 5 a 8 se presentan las proporciones de respuestas en las categorías de los ítems politómicos, para demostrar que los que tienen un ajuste inferior es por la poca representatividad de alguna de las categorías en sus niveles de respuesta. Por ejemplo, la categoría 1 del ítem 4 (p69).

**TABLA 5. Distribución de las puntuaciones en el ítem 4**

		Frecuencia	Porcentaje válido	Porcentaje acumulado
Válidos	0	155	36,3	36,3
	1	21	4,9	41,2
	2	251	58,8	100,0
	Total	427	100,0	
Perdidos	Sistema	50		
Total		477		

**TABLA 6. Distribución de las puntuaciones en el ítem 8**

		Frecuencia	Porcentaje válido	Porcentaje acumulado
Válidos	0	38	8,7	8,7
	1	33	7,6	16,3
	2	365	83,7	100,0
	Total	436	100,0	
Perdidos	Sistema	41		
Total		477		

**TABLA 7. Distribución de las puntuaciones en el ítem 10**

		Frecuencia	Porcentaje válido	Porcentaje acumulado
Válidos	0	227	67,4	67,4
	1	60	17,8	85,2
	2	50	14,8	100,0
	Total	337	100,0	
Perdidos	Sistema	140		
Total		477		

**TABLA 8. Distribución de las puntuaciones en el ítem 12**

		Frecuencia	Porcentaje válido	Porcentaje acumulado
Válidos	0	20	4,3	4,3
	1	71	15,4	19,8
	2	369	80,2	100,0
	Total	460	100,0	
Perdidos	Sistema	17		
Total		477		

Como muestran las tablas (9 a 11), el promedio en el logro en matemáticas aumenta a medida que se incrementa la categoría del ítem, aspecto que indica una buena estimación de la puntuación de rendimiento con el modelo de Rasch.

También a modo de ejemplo, se incluyen las CCI de un ítem que puede considerarse fácil y otro difícil para visualizar las diferencias (figura 2). El ítem 1 (p66) es fácil, incluso con valores bajos de resultados en matemáticas hay alta probabilidad de responder correctamente. Y el ítem 7 (p72) puede considerarse más difícil, para tener una probabilidad mayor que 0.5 de responder correctamente al ítem se necesitan unos resultados en matemáticas por encima de la media.

**TABLA 9. Puntuaciones en matemáticas de los casos con distintos niveles de respuesta en el ítem 4**

Respuesta	N	Mínimo	Máximo	Media	Desv. típ.
0	155	-2.677	1.079	-.39153	.650960
1	21	-.884	.844	.18161	.447278
2	251	-1.974	1.868	.34103	.715073

**TABLA 10. Puntuaciones en matemáticas de los casos con distintos niveles de respuesta en el ítem 8**

Respuesta	N	Mínimo	Máximo	Media	Desv. típ.
0	38	-2.677	.094	-.78928	.668668
1	33	-1.778	1.320	-.32619	.707490
2	365	-2.187	1.868	.17780	.716980

**TABLA 11. Puntuaciones en matemáticas de los casos con distintos niveles de respuesta en el ítem 10**

Respuesta	N	Mínimo	Máximo	Media	Desv. típ.
0	227	-2.677	1.320	-.14241	.662592
1	60	-.643	1.573	.57686	.596638
2	50	-.475	1.868	.81324	.523470

Respecto a la información global de la prueba de matemáticas, la fiabilidad obtenida ha sido de .776. Un valor cercano al .8 que se considera el valor óptimo para pruebas de rendimiento, sobre todo, si se van a tomar decisiones en función de los resultados. La figura 3 muestra la curva característica del test (CCT) que relaciona la suma de respuestas correctas con las puntuaciones estimadas con el modelo de Rasch.

Finalmente, en la figura 4 se incluye un mapa para representar las dificultades estimadas para los ítems y las puntuaciones. De esta forma es posible identificar qué ítems son necesarios responder correctamente para situarse en los distintos tramos de la escala. Por ejemplo, el ítem 1 es más fácil y necesita valores del rasgo por

debajo de -4; y el que más nivel de rendimiento necesita es el segundo paso del ítem 10.

En la tabla 12, se recoge la distribución de los ítems por dominio de contenido y nivel de desempeño, resultantes de la muestra.

Dado que la prueba de matemáticas se ha construido desde la prueba TIMSS, se ha considerado mantener la estructura de los resultados de la prueba de acuerdo al dominio cognitivo y al nivel de desempeño de los resultados obtenidos (tabla 13).

Podemos ver en las tablas 12 y 13 cómo en algunas variables aparecen dos niveles de logro (entre paréntesis), son aquellas de respuesta construida, correspondiendo el primer valor a la respuesta 1 y el segundo a la respuesta 2.

FIGURA 2. Gráficos CCI de los ítems 1 y 7

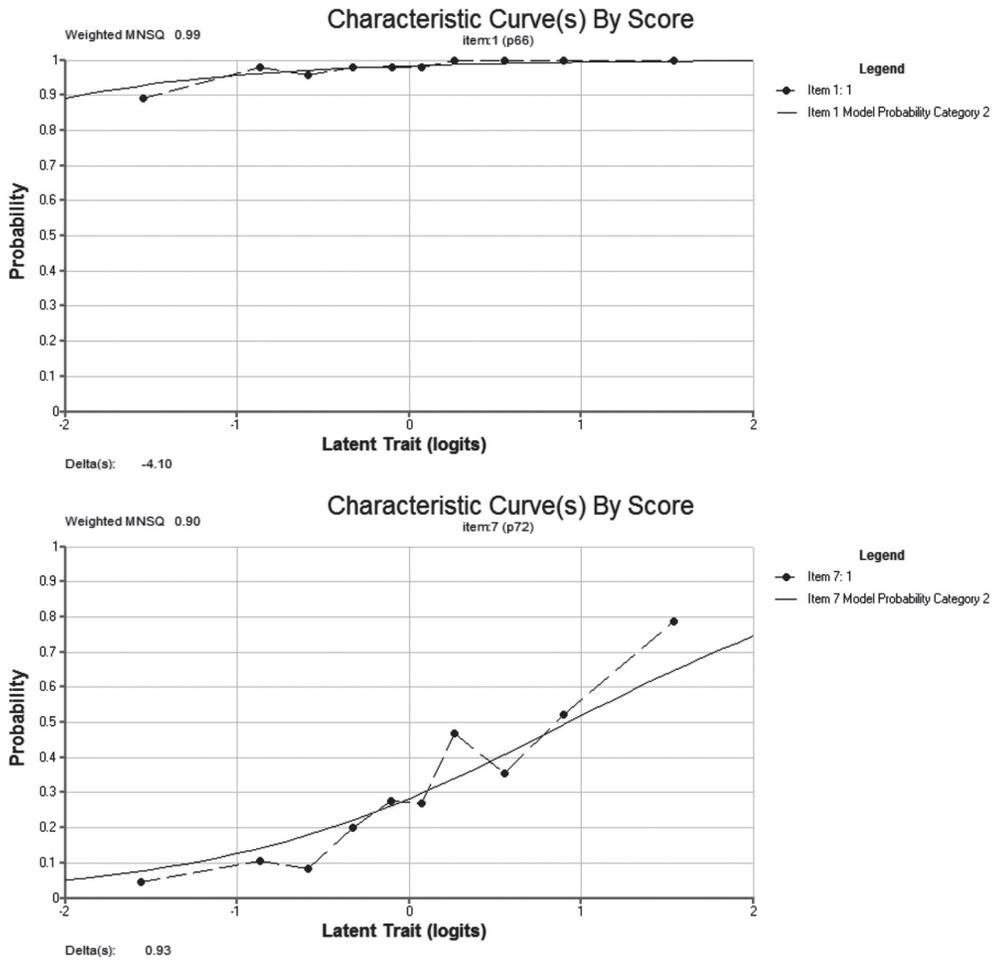


TABLA 12. Distribución de ítems por dominio contenido vs. nivel de desempeño en matemáticas

	Bajo	Intermedio	Alto	Avanzado	TOTAL
Álgebra	1(90); 2(233); 3(233); 11(252)	5(392)	8(467;518); 4(518;534)	10(558;581); 7(593)	9
Números	16(184); 13(206); 6(316); 18(232)	15(395); 17(409)	12(485;515);14(473)	9(575)	9
Datos y azar	21(326);19(175); 23(300)	20(417)	22(487); 24(475)		6
Geometría	26(98); 28(269);29(333)	27(384); 25(426); 30(423)			6
Total	14	7	6	3	30

FIGURA 3. Curva característica de la prueba de matemáticas

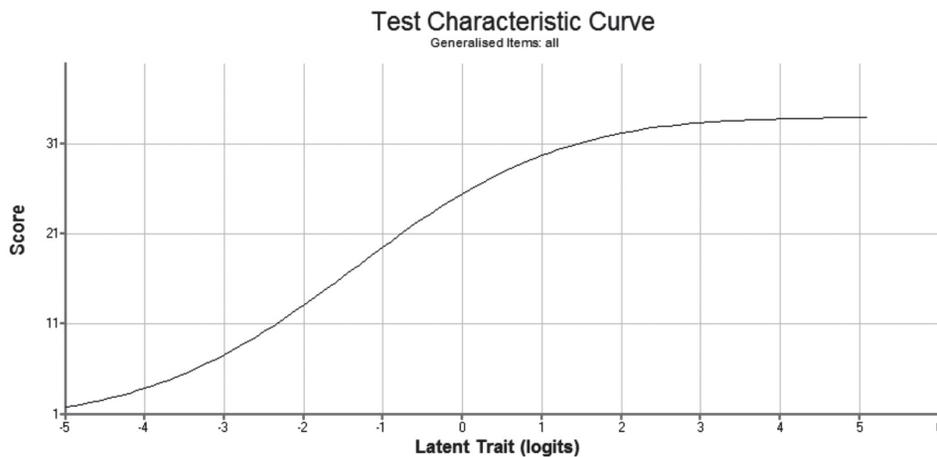
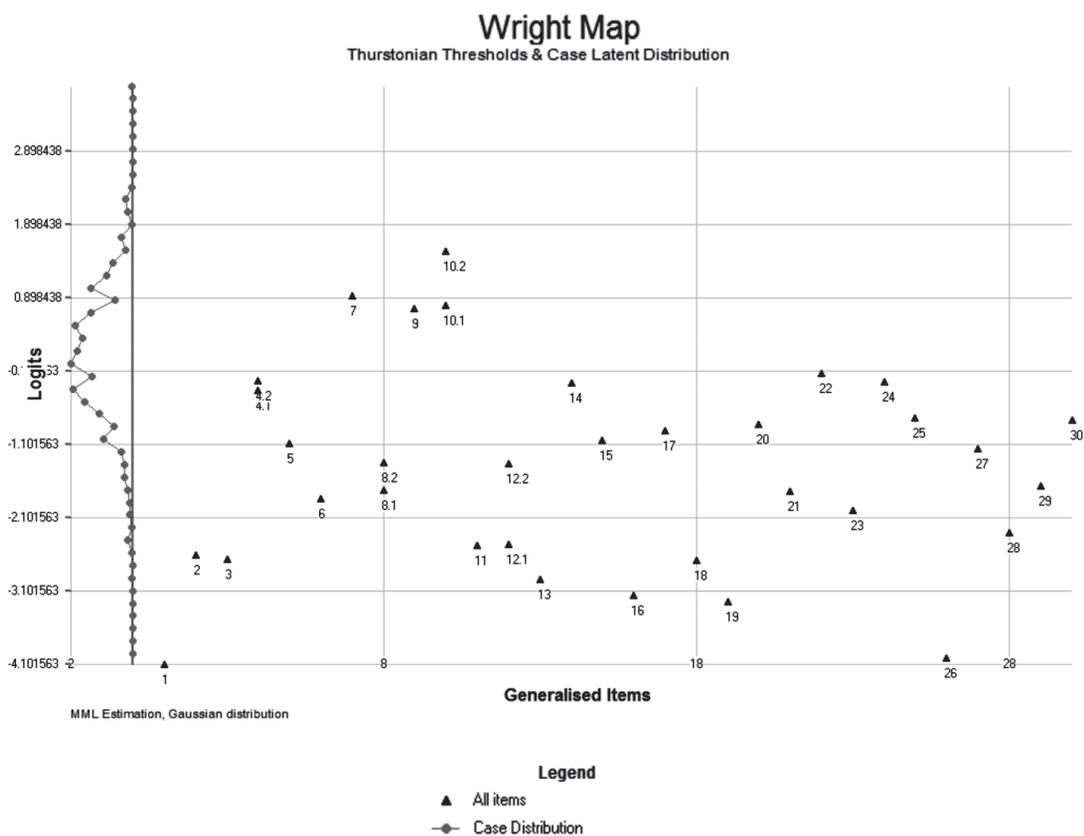


FIGURA 4. Mapa de las dificultades de los ítems y los resultados estimados en matemáticas



**TABLA 13. Dominio cognitivo/Nivel de desempeño**

	Bajo	Intermedio	Alto	Avanzado	Sobre avanzado	Total
Conocimiento	1(90); 16(184); 21(326); 26(98); 3(233); 6(316); 23(300); 29(333)		12(485;515); 4(518;534)			10
Razonamiento	18(232); 28(269)	5(392); 20(417); 25(426)	8(467;518); 14(473); 24(475)	9(575); 10(558;581)		10
Aplicación	19(175); 2(233); 13(206); 11(252)	15(395); 17(409); 27(384); 30(423)	22(487)	7(593)		10
Total	14	7	6	3		30

## Conclusiones

Hemos construido un instrumento diseñado para medir las habilidades matemáticas de los maestros en formación, que se ha utilizado como elemento de diagnóstico.

La prueba tiene una finalidad diagnóstica, de la misma forma que ocurre con TIMSS o las evaluaciones generales que realiza el MECED (<https://www.mecd.gob.es/inee/evaluaciones-nacionales.html>). El propósito es conocer el nivel de dominio de las competencias matemáticas de los futuros docentes antes de comenzar el título de grado. Con este diagnóstico, las universidades tendrán la capacidad de reforzar algún aspecto de su programa formativo o implementar actuaciones específicas sobre el alumnado si las consideran necesarias.

Partimos de la consideración de que la prueba TIMSS utilizada como elemento base para la elaboración de la prueba diagnóstica está diseñada para una población distinta; sin embargo podemos considerar el resultado en la misma como dato de logro inicial en el que los estudiantes deben acreditar un dominio suficiente.

El análisis de los ítems de la prueba mediante el modelo de Rasch de la TRI facilita un análisis detallado de cada ítem referido a su calidad

técnica y al aporte individual que hace al conjunto de la prueba. Este hecho ha resultado fundamental en este caso, dado que estamos utilizando parte de una prueba previamente diseñada para una población distinta, aportando además un error de medida en caso de existir. Este modelo genera puntuaciones en una escala de intervalo, lo que significa que diferencias entre distintos tramos de la escala suponen los mismos cambios en las probabilidades de responder de manera correcta a los ítems con la misma dificultad (Prieto y Delgado, 2003).

Los ítems, en general, tienen un buen funcionamiento y se ajustan al modelo psicométrico de Rasch. No obstante, existe una mayor presencia de ítems que pueden considerarse fáciles que está provocada porque se diseñaron para una población de menor edad.

Se observa un buen funcionamiento de la escala en su conjunto de acuerdo al modelo de Rasch, aunque cabe mencionar el comportamiento diferencial de los ítems de respuesta construida respecto a los de respuesta seleccionada. Este hecho se justifica por la escasa variabilidad de la respuesta como han evidenciado los análisis particulares de este tipo de ítems. Por ejemplo, en el ítem 4, una categoría con escasa elección (1) con el 4,9% de los valores.

La forma de mejorar este tipo de resultados implicaría el uso de una muestra mayor, aunque cabe destacar que en nuestra investigación funciona de manera adecuada la puntuación del rasgo.

La prueba funciona bien observando la curva característica, además la suma entre el nivel del rasgo y la suma de respuestas correctas están vinculadas.

Los índices de ajuste estimados con el modelo de Rasch muestran el correcto funcionamiento de los ítems incluidos en la prueba. La falta de ajuste de los ítems de respuesta construida está justificada por la baja variabilidad de las respuestas, no obstante, su funcionamiento es correcto para diferenciar entre los niveles de rendimiento. De esta forma, es posible concluir que la prueba es válida para comprobar el nivel de la competencia matemática de los estudiantes de grado.

En relación a los resultados del nivel de dificultad por dominio de contenidos, los ítems de mayor dificultad para la muestra coinciden con los que tuvieron un menor porcentaje de respuesta correcta para los estudiantes de octavo grado a nivel internacional (7 y 10), ambos pertenecientes al dominio de contenido de álgebra.

Como conclusión final y recomendación del presente trabajo, planteamos la necesidad de construir unos estándares que definan los conocimientos matemáticos que un estudiante necesita para acceder a los estudios del grado de magisterio.

## **Agradecimientos**

Este trabajo es producto de una investigación financiada por la Comunidad de Madrid, desde la Consejería de Educación e Investigación en el año 2017.

## **Referencias bibliográficas**

---

- Beaton, D. E., Bombardier, C., Guillemin, F. y Bosi, M. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25, 3186-3191, doi: 10.1097/00007632-200012150-00014
- Cano, M. E. (2011). La evaluación por competencias en la educación superior. Profesorado. *Revista de Currículum y Formación del Profesorado*, 12(3), 1-16.
- Castro, M. (2011). ¿Qué sabemos de la medida de las competencias? Características y problemas psicométricos en la evaluación de competencias. *Bordón. Revista de Pedagogía*, 63(1), 109-123.
- Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of education for teaching*, 15(1), 13-33, doi: 10.1080/0260747890150102
- Escobar-Pérez, J. y Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 6(1), 27-36.
- Fernández-Alonso, R. y Muñiz, J. (2011). Diseño de cuadernillos para la evaluación de las competencias básicas. *Aula Abierta*, 39(2), 3-34.
- Hodgen, J. y Askew, M. (2007). Emotion, identity and teacher learning: Becoming a primary mathematics teacher. *Oxford Review of Education*, 33(4), 469-487, doi: 10.1080/03054980701451090
- Hyrkäs, K., Appelqvist-Schmidlechner, K. y Oksa, L. (2003). Validating an instrument for clinical supervision using an expert panel. *International Journal of Nursing Studies*, 40(6), 619-625, doi: 10.1016/S0020-7489(03)00036-1
- Landis, J. R. y Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2), 363-374.

- Llinares, S. (2003). Matemáticas escolares y competencia matemática. En M. C. Chamorro (coord.). *Didáctica de las matemáticas para primaria* (pp. 3-30). Madrid: Prentice Educación.
- Martínez, M. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34, 207- 218.
- Martínez, M. R., Hernández, M. J. y Hernández, M. V. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47(2), 149-174, doi: 10.1007/BF02296272
- Mullis, I. V., Martín, M. O., Ruddock, G. J., O'Sullivan, C. Y. y Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Boston, M. A.: TIMSS & PIRLS International Study Center.
- Navarro, E., Expósito, E., López, E. y Thoilliez, B. (2014). EPIBI: Escala de Percepción de Indicadores de Bienestar Infantil. Validación del instrumento utilizando modelos politómicos de Rasch. *Revista de Educación*, 364, 39-65, doi:10.4438/1988-592X-RE-2014-364-254
- NCTM. National Council of Teachers of Mathematics (1991). *Professional Standards for Teaching Mathematics*. Reston, Va.: The NCTM, Inc.
- Nortes, R. y Nortes Checa, A. (2017). Competencia matemática, actitud y ansiedad hacia las Matemáticas en futuros maestros. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 20(3), 145-160, doi: 10.6018/reifop.20.3.290841
- Osterlind, S. (1998). *Constructing test items: multiple-choice, constructed-response, performance, and others formats*. Boston: Kluwer Academia Publishers.
- Pérez, A. y Soto, E. (2011). Luces y sombras de PISA. Sentido educativo de las evaluaciones externas. *Culturay Educación*, 23(2), 171-182, doi: 10.1174/113564011795944758
- Prieto, G. y Delgado, A. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen (Dinamarca): Danish Institute for Educational Research.
- Rodríguez-Gómez, G. e Ibarra, M. S. (2013). ActEval: un instrumento para el análisis y reflexión sobre la actividad evaluadora del profesorado universitario. *Revista de Educación*, 362, 69-104. doi:10-4438/1988-592X-RE-2011-362-153
- Sandoval, P., Frit, M., Maldonado, A. y Rodríguez, F. (2010). Evaluación de habilidades en matemática y comprensión lectora en estudiantes que ingresan a pedagogía en educación básica: un estudio comparativo en dos universidades del Consejo de Rectores. *Educación en Revista*, 2, 73-102.
- Solano, G., Contreras, L. A. y Backhoff, E. (2006). Traducción y adaptación de pruebas: Lecciones aprendidas y recomendaciones para países participantes en TIMSS, PISA y otras comparaciones internacionales. *Revista Electrónica de Investigación Educativa*, 8(2), 1-21.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R. y Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M)*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Tourón, J. (2009). El establecimiento de estándares de rendimiento en los sistemas educativos. *Estudios Sobre Educación*, 16, 127-146.
- Varas, L., Lacourly, N., López, A. y Giaconi, V. (2013). Evaluación del conocimiento pedagógico del contenido para enseñar matemáticas elementales. *Enseñanza de las Ciencias*, 31(1), 171-187.
- Viera, A. J. y Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*, 37(5), 360-363.
- White, A. L., Way, J., Perry, B. y Southwell, B. (2005). Mathematical attitudes, beliefs and achievement in primary pre-service mathematics teacher education. *Mathematics Teacher Education and Development*, 7, 33-52.

ANEXO I

Nº	Dominio contenido	Variable	Dominio cognitivo	Tópico principal
1	Álgebra	p66	Conocimiento	Expresiones algebraicas
2	Álgebra	p67	Aplicación	Patrones
3	Álgebra	p68	Conocimiento	Expresiones algebraicas
4	Álgebra	p69	Conocimiento	Expresiones algebraicas
5	Álgebra	p70	Razonamiento	Patrones
6	Números	p71	Conocimiento	Números naturales
7	Álgebra	p72	Aplicación	Expresiones algebraicas
8	Álgebra	p73	Razonamiento	Patrones
9	Números	p74	Razonamiento	Números naturales
10	Álgebra	p75	Razonamiento	Patrones
11	Álgebra	p76	Aplicación	Expresiones algebraicas
12	Números	p77	Conocimiento	Razón, proporción y porcentaje
13	Números	p78	Aplicación	Razón, proporción y porcentaje
14	Números	p79	Razonamiento	Fracciones y decimales
15	Números	p80	Aplicación	Números naturales
16	Números	p81	Conocimiento	Fracciones y decimales
17	Números	p82	Aplicación	Fracciones y decimales
18	Números	p83	Razonamiento	Enteros
19	Datos y azar	p84	Aplicación	Azar
20	Datos y azar	p85	Razonamiento	Azar
21	Datos y azar	p86	Conocimiento	Organización y representación de datos
22	Datos y azar	p87	Aplicación	Interpretación de los datos
23	Datos y azar	p88	Conocimiento	Azar
24	Datos y azar	p89	Razonamiento	Interpretación de los datos
25	Geometría	p90	Razonamiento	Formas geométricas
26	Geometría	p91	Conocimiento	Formas geométricas
27	Geometría	p92	Aplicación	Medida
28	Geometría	p93	Razonamiento	Formas geométricas
29	Geometría	p94	Conocimiento	Espacio y movimientos
30	Geometría	p95	Aplicación	Medida

**Abstract**

*Adapting the test of the Trends in International Mathematics and Science Study (TIMSS) for the assessment of the mathematical competence of students entering the undergraduate degree of Primary Education School Teacher*

**INTRODUCTION.** The foundation of mathematical pedagogical knowledge that a teacher must have is based in a special way on the knowledge he/she has about the contents of the subject. The objective of this paper is the elaboration of a diagnostic test of mathematical abilities, based on the selection and adaptation of some of the items of the Trends in

International Mathematics and Science Study (TIMSS) 2011, as well as the technical aspects from the application of the Rasch model. **METHOD.** The items of the elaborated test were selected following the original structure of blocks of content and cognitive domain established in the framework of the official TIMSS 2011 test (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009). The translation and adaptation procedure has been developed in order to ensure greater linguistic validity. The study was carried out with a sample of 477 students enrolled in the first year of the different Bachelordegrees of Primary Education School Teacher that are offered in universities across the Autonomous Community of Madrid (Spain). **RESULTS.** We considered two types: multiple-choice items and constructed response items. Given the characteristic of the format it has been necessary to assess the reliability of the correction made by the evaluators to guarantee the agreement in their correction. **DISCUSSION.** The Rasch model used for the validation of the test has allowed us to generate evidence that shows a high degree of reliability, pointing out the differences in behavior between the multiple-choice items and the constructed response items (elements that require students to construct their own written response). The model allowed us to classify the results of the items by level of difficulty, which facilitates an interpretation of the results according to both the content domain and the cognitive domain.

**Keywords:** TIMSS, Mathematical competence, Rasch model, Assessment adaptation, Pre-service teacher preparation.

## Résumé

---

### *Adaptation de l'évaluation TIMSS pour tester les compétences mathématiques des étudiants entrants à l'école du professorat et l'éducation*

**INTRODUCTION.** Le fondement des connaissances pédagogiques pour les mathématiques qu'un enseignant doit avoir est spécialement basé sur les connaissances que lui ou elle a de la matière. L'objectif de cet article est montrer l'élaboration d'un test diagnostique pour les aptitudes mathématiques, à partir de la sélection et de l'adaptation de certains éléments issus de l'étude des tendances mathématiques et scientifiques publiée en 2011 ('*Trends in International Mathematics and Sciences*', TIMSS.), ainsi que les aspects techniques liés à l'application du modèle de Rasch. **MÉTHODE.** Les éléments du test élaboré ont été sélectionnés en suivant la structure originelle des blocs de contenu et domaines cognitifs établis dans le cadre du test TIMSS officiel de 2011 (Mullis, Martin, Ruddock, O'sullivan & Preuschoff 2009). La procédure de traduction et d'adaptation a été développée dans le but de garantir une plus grande validité linguistique. L'étude a été réalisée auprès d'un échantillon de 477 étudiants inscrits en première année de formation pour enseignants du primaire dans les universités de la Communauté Autonome de Madrid en Espagne. **RÉSULTATS.** Nous avons considéré deux types d'éléments: les éléments à choix multiple et les éléments de réponse construite. Compte tenu de cette caractéristique de format, il a été nécessaire de tester la fiabilité de la correction faite par les évaluateurs afin de garantir l'harmonie dans leurs corrections. **DISCUSSION.** Le modèle de Rasch utilisé dans la validation du test nous a permis de générer des preuves qui montrent un haut degré de fiabilité, soulignant les différences de comportement entre les éléments à choix multiple et les éléments de réponse construite (éléments qui obligent aux étudiants à construire leur propre réponse écrite). Le modèle nous a permis de classer les résultats des éléments par

niveau de difficulté, ce qui facilite l'interprétation des résultats selon le domaine de contenu et le domaine cognitif.

**Mots clés:** *TIMSS, compétence mathématique, modèle de Rasch, adaptation des tests, formation initiale des enseignants.*

## **Perfil profesional de los autores**

---

### **Blanca Arteaga Martínez (autora de contacto)**

Profesora ayudante doctora. Doctora en Ciencias de la Educación. Licenciada en CC. Matemáticas. En la actualidad imparte docencia en la Facultad de Educación de la Universidad de Alcalá, con un perfil de Didáctica de las Matemáticas. Ha trabajado en distintas universidades, presenciales y *on-line*, donde además ha desempeñado distintos cargos de coordinación académica. Ponente en cursos de formación de profesores. Miembro del grupo de investigación “Pedagogía Adaptativa” de la Universidad Complutense de Madrid.

Correo electrónico de contacto: blanca.arteaga@uah.es

Dirección para la correspondencia: Facultad de Educación. C/ Madrid, 1, 19001 Guadalajara.

### **Enrique Navarro Asencio**

Profesor ayudante doctor. Doctor en Pedagogía por la Universidad Complutense de Madrid (UCM), obteniendo el premio extraordinario de doctorado. Actualmente es profesor ayudante doctor en la UCM y su línea de trabajo está relacionada con psicometría y evaluación del rendimiento académico y factores asociados.

Correo electrónico de contacto: enriquen@ucm.es

### **Arántzazu Fraile Rey**

Profesora visitante. Doctora en Ciencias de la Educación. Licenciada en CC. Físicas. En la actualidad imparte docencia en matemáticas y educación matemática en la Facultad de Educación de la Universidad de Alcalá en los grados de Educación Primaria y Doble Grado de Humanidades y Educación Primaria. Su línea de investigación está centrada en la resolución de problemas.

Correo electrónico de contacto: arantzazu.fraile@uah.es

### **Pedro Ramos Alonso**

Profesor titular. Doctor en Ciencias Matemáticas por la Universidad Politécnica de Madrid. Su investigación se ha centrado en geometría discreta y computacional. Desde el año 2010 trabaja en el ámbito de la educación matemática. Es profesor titular de Universidad en la Universidad de Alcalá. Imparte cursos de formación de profesorado y colabora con las editoriales Polygon Education y SM en la implantación de la metodología Singapur en matemáticas.

Correo electrónico de contacto: pedro.ramos@uah.es