# Construcción de un TAI para evaluar los conocimientos de sobre una materia universitaria

Tirado González, Sonia
Cañadas Osinski, Isabel
Bautista Ortuño, Rebeca

*Universidad Miguel Hernández*
*Área de Metodología de las Ciencias del Comportamiento*
*Departamento de Psicología de la Salud*
*Avda. de la Universidad s/n – 03202 Elche (Alicante) – España*
*Correo electrónico: sonia.tirado@umh.es*

## Resumen

Dentro de las pruebas de evaluación educativa que se utilizan en la actualidad, los tests constituyen sin lugar a dudas uno de los instrumentos más empleados para la valoración de aptitudes o de conocimientos. Se presentan en muy distintas formas, aunque el formato clásico de *lápiz y papel* sigue siendo el más utilizado. Un claro ejemplo lo conforman los exámenes universitarios, las conocidas pruebas MIR, PIR, etc., donde se evalúan los conocimientos de forma masificada. Sin embargo, a pesar de su amplio uso, no están exentos de desventajas, como son la excesiva longitud y el excesivo tiempo y, sobre todo, la desconfianza que provocan debido a la falta de precisión de las medidas obtenidas.

En los últimos años ha nacido una nueva generación de tests, los *Tests Adaptativos Informatizados* (TAI), que representan sin lugar a dudas una de las mejores alternativas como herramientas de evaluación del conocimiento adquirido. No en vano, se dice de ellos que constituyen una de las revoluciones más importantes de los últimos años en la evaluación psicológica y educativa. Dentro del marco actual de la enseñanza, en el que tanto el EEES como la UEALC abogan por un proceso de enseñanza-aprendizaje tutorizado, el proceso de elaboración de los nuevos métodos de evaluación de los conocimientos adquiridos, cobra una especial relevancia. Así, mediante nuevos componentes metodológicos, tecnológicos e informáticos, los TAI dan respuesta a las necesidades educativas específicas tanto de personas como de grupos.

Partiendo de este nuevo paradigma educativo, el trabajo que aquí se propone pretende construir un TAI para la evaluación rápida, precisa, válida y eficaz del conocimiento basado en un banco de ítems calibrado previamente.

PALABRAS CLAVE: *Teoría de Respuesta al Ítem, Test Adaptativo Informatizado, Banco de ítems, Función de información del ítem, Método de máxima información.*

KEYWORDS: *Item Response Theory, Computerized Adaptive Testing, Item Bank, Item information function, Method of maximum information.*

# Introducción

Desde sus orígenes, la Unión Europea ha contemplado la creación de un espacio educativo común para sus países miembros, regido bajo los principios de transparencia y comparabilidad de los diferentes sistemas universitarios nacionales. Este objetivo ha comenzado a ser una realidad desde la Declaración de Bolonia en 1999. La convergencia que se persigue trata de conjugar la diversidad con la posibilidad de hacer comparables y comprensibles los estudios, títulos y diplomas existentes en cada país. Se trata, en definitiva, de la creación de lo que se ha denominado Espacio Europeo de Educación Superior (EEES).

Las diferentes reuniones de los ministros de educación de la Unión Europea han permitido definir los ejes principales sobre los que desarrollar el EEES: la asunción de un modelo de titulaciones con dos niveles, la adopción de un sistema de créditos que permita su acumulación y transferencia, la promoción de la movilidad académica en Europa y el aseguramiento de niveles de calidad para el desarrollo de criterios y de metodologías comparables. En muy poco tiempo, se ha producido una gran actividad tanto desde el punto de vista legislativo como por parte de las instituciones universitarias y de los agentes sociales implicados.

Este proyecto iniciado en Europa ha llegado a extenderse a los países latinoamericanos hasta el punto de haberse creado el llamado UEALC, es decir, el Espacio Común de Educación Superior de América Latina y Caribe. Concretamente, el *Proyecto 6x4 UEALC*, partiendo de la necesidad de estrechar la cooperación y facilitar la movilidad entre los sistemas de educación superior en América Latina, responde al propósito principal de la Declaración de la Conferencia Ministerial de los países de la Unión Europea, de América Latina y del Caribe sobre la Enseñanza Superior (París, noviembre de 2000), que es la construcción de un espacio común de educación superior UEALC.

Los aspectos clave del proyecto son el desarrollo de un acercamiento a la evaluación y reconocimiento de los resultados del aprendizaje, expresados en términos de competencias, y el fortalecimiento de la pertinencia y de los vínculos de la educación superior y la investigación con la sociedad, con especial atención en el desarrollo de los mecanismos que faciliten el reconocimiento de las calificaciones y competencias de las personas, tanto para continuar con sus estudios y su vida laboral, como para incrementar la movilidad académica.

Dentro de este nuevo paradigma, uno de los retos y compromisos que se plantea es un cambio en el enfoque metodológico, transformando el sistema educativo basado en la enseñanza en otro basado en el *aprendizaje*. Para conseguir esta meta, el proceso de mejora debe ser interactivo y el profesor, como agente creador de entornos de aprendizaje que estimulen a los alumnos, debe utilizar metodologías más activas que permitan una mayor implicación y autonomía por parte del estudiante. En consecuencia, se apuesta por la enseñanza tutorizada y el trabajo personal del alumno.

Sin embargo, a pesar de lo laudable de estos objetivos, resulta un contrasentido compaginar la implantación del nuevo modelo de enseñanza-aprendizaje y la evaluación masificada de los estudiantes. Es conveniente, por tanto, disponer de instrumentos de evaluación adaptados a esta enseñanza tutorizada o, lo que es lo mismo, de herramientas valorativas del rendimiento de los estudiantes *a medida* o *personalizadas*, acordes con un método de enseñanza que pretende la autonomía del estudiante universitario. En definitiva, una enseñanza universitaria de calidad, demanda un sistema de evaluación flexible, rápido y eficaz.

Como principio general, la evaluación debe estar basada en los aprendizajes obtenidos en función de los objetivos propuestos. En este sentido, la evaluación debe adaptarse a los objetivos de contenido (por ejemplo, definir conceptos), a los que implican operaciones (por ejemplo, comprender, aplicar o reconocer) y a los relacionados con la adquisición de destrezas, actitudes y conocimientos prácticos (por ejemplo, valorar, planificar o ejecutar). Deben utilizarse, pues, procesos de evaluación comprensiva y diversificada que abarquen todos los objetivos y que faciliten la comprobación de los conocimientos adquiridos (Cruz Tomé, 1999).

Dentro del ámbito educativo, el instrumento habitual de medida de variables denominadas habilidades, aptitudes, rendimiento o conocimientos, es el *test convencional*, sistema de evaluación que en poco tiempo se ha convertido en una de las técnicas más utilizadas para valorar el rendimiento de los evaluandos. Hoy por hoy, las pruebas conocidas como *test de lápiz y papel* no son ajenas ni al profesorado ni al alumnado. Su modo de proceder es el siguiente: los estudiantes resuelven una prueba estandarizada en su forma y contenido. En otras palabras, todos los alumnos reciben los mismos ítems, o lo que es lo mismo, enunciados acompañados de dos o más opciones de respuesta, de entre las cuales debe escoger la/s alternativa/s correcta/s. El número de ítems acertados es un estimador del rendimiento tras aplicar la ya clásica fórmula de corrección del azar que, véase Muñiz (1998), es a todas luces un desatino. Además, las principales características de los ítems, a saber, discriminación, dificultad y azar, sólo pueden ser obtenidas *después* de la aplicación del test, y no *antes*. A pesar de todo ello, los estudiantes son calificados y clasificados en una escala estandarizada de rango 0-10.

En la actualidad, la Psicometría, como disciplina encargada del desarrollo de sistemas de evaluación en distintos contextos, ha abierto nuevas vías con el despliegue de diversas teorías a partir de las cuales se supervisa el proceso de construcción de tests y de escalas de medida. Una de ellas es la Teoría de Respuesta al Ítem (TRI). Entre sus actuaciones, permite el calibrado psicométrico de cada ítem (índices de dificultad, discriminación y azar) de forma independiente. En consecuencia, podemos construir progresivamente bancos de ítems psicométricamente apropiados, a partir de los cuales confeccionar uno o varios tests sobre la misma materia. La TRI está tomando mayor auge gracias a los avances informáticos, no en vano, una de sus líneas de investigación y desarrollo es la elaboración de los Tests Adaptativos Informatizados (TAI) o *Computerized Adaptive Testing* (CAT).

Elaborados a partir de bancos de ítems ya calibrados, la filosofía de los TAIs es crear *tests a medida* del sujeto evaluado. La sesión se inicia con una estrategia de arranque que establece el nivel de rasgo inicial que se asigna al evaluando y que determina cual será el primer ítem a presentar. Después de que el evaluando responde al primer ítem, mediante procedimientos estadísticos bayesianos o máximo verosímiles, se realiza una primera estimación de su nivel de rasgo. También se emplean procedimientos derivados de la TRI para seleccionar el segundo ítem, considerando que sea apropiado para el primer nivel de rasgo provisional estimado. Así pues, en cada paso del proceso, se procede a la selección y presentación sucesiva de ítems, teniendo en cuenta el patrón de respuestas (aciertos/fallos) que se dan a los ítems precedentes para la estimación del nivel de rasgo provisional (y la precisión asociada a esta estimación) en ese momento de la aplicación del TAI. Se requiere además un criterio para dar por terminada la secuencia de presentación de ítems, que normalmente tiene que ver con la consecución de cierto nivel de precisión o con el establecimiento de una determinada longitud del TAI. Aunque resulte sorprendente, mediante los algoritmos pertinentes se pueden formar los ítems más apropiados para el nivel del evaluando en cada momento, sin haberlos generado previamente y sin necesidad de calibrarlos empíricamente. Además se podría incluso corregir automáticamente mediante un sistema de experto la respuesta a preguntas abiertas, que precisan elaborar y escribir la misma (Olea, Ponsoda y Prieto, 1999).

En definitiva, dentro del campo de la evaluación, las ventajas y posibilidades de los TAIs exceden ampliamente las previstas inicialmente. En otros países como, por ejemplo, Estados Unidos, se han convertido en el nuevo modo de evaluar y su uso se ha extendido a prácticamente todos los contextos psicoeducativos, despertando gran interés tanto desde una perspectiva teórica como aplicada. Además, en estos países resulta ya casi tradicional su utilización en el campo de la clínica, en procesos de selección en las fuerzas armadas, en la evaluación de destrezas y conocimientos básicos en diferentes niveles educativos, en pruebas de certificación en las carreras de Arquitectura o Medicina, etc.

Dentro de nuestro ámbito, nuestro grupo de investigación lleva años trabajando en la construcción de pruebas de nivel adaptativas informatizadas (Cañadas, Tirado y Núñez, 2006a; Cañadas, Tirado y Núñez, 2006b; Núñez, Cañadas y Tirado, 2004; Tirado, Cañadas y Núñez, 2006). Ahora nos proponemos crear un procedimiento de medición basado en los TAIs que

permita la evaluación personalizada de los conocimientos para la Licenciatura de Psicología. Para ello, nuestro trabajo de construcción del TAI debe seguir un proceso que comprende las siguientes fases (Renom y Doval, 1999): 1) Planificación y prospección del TAI. 2) Producción de un banco de ítems. 3) Calibración del banco de ítems. 4) Implementación del banco de ítems. 5) Gestión del TAI. El cumplimiento de nuestros objetivos nos permitirá estar en consonacia con la calidad y equidad que se exigen desde el nuevo paradigma de educación y aprendizaje.

## Objetivo

A la vista de los nuevos horizontes que se plantean para la educación superior y dada nuestra formación metodológica, nos planteamos la construcción de una prueba de nivel adaptativa informatizada para evaluar los conocimientos de los estudiantes que cursan la materia de Psicometría de la Titulación de Psicología, a partir de un banco de ítems desarrollado anteriormente.

Para cubrir este objetivo, decidimos seguir un proceso que comprende las siguientes fases (Renom y Doval, 1999):

1. Planificación y prospección del TAI.
2. Producción del banco de ítems.
3. Calibración cualitativa y cuantitativa del banco de ítems.
4. Implementación del banco de ítems con el programa FastTEST Pro (Assessment Systems Corporation, 2002).
5. Gestión, mantenimiento y renovación del TAI.

A continuación expondremos los niveles alcanzados en las dos últimas fases. Las fases anteriores se exponen en un trabajo paralelo.

## Procedimiento y resultados

### *4. Implementación y ejecución del TAI*

En el mercado existen programas para la construcción y gestión de TAIs tales como *ADTEST* (Ponsoda, Olea y Revuelta, 1994), *MICROCAT* (Assessment Systems Corporation, 1996) y *FastTEST Pro v. 1.6* (Assessment Systems Corporation, 2002). En este caso, el software encargado de la gestión del banco de ítems es el *FastTEST Pro v. 1.6*, que es asimismo el encargado de gestionar el TAI una vez calibrado el banco de ítems.

Para comenzar con el proceso de evaluación se debe especificar el valor inicial o un intervalo aleatorio de valores de conocimiento o aptitud; este valor o intervalo será la referencia para que el programa escoja el primer ítem y realice la primera estimación del parámetro del estudiante, esto es, su nivel de conocimiento o aptitud. Una vez contestado este ítem, existen distintas estrategias para seleccionar el siguiente ítem del banco. En los tests convencionales se fijan de antemano el número de ítems y posteriormente se pasa a toda la muestra con independencia del nivel de aptitud de cada estudiante. En los tests adaptativos, los ítems son seleccionados según la respuesta del estudiante, adecuando la dificultad del ítem siguiente al nivel de habilidad estimada tras la respuesta del estudiante, con objeto de obtener la máxima información acerca del verdadero valor del parámetro de habilidad. En definitiva, los tests adaptativos son tests *personalizados*.

Cada vez que un ítem es administrado, la habilidad del estudiante es estimada. El próximo ítem que se administrará es escogido según esta estimación. Para la selección de los ítems, *FastTEST Pro v. 1.6* implementa el método de máxima información a partir del valor de

habilidad estimado a través de los ítems administrados. La *función de información* [ $I_j(q_i)$ ] de un ítem $j$ para un estudiante de habilidad $q_i$ depende de:

- La discriminación del ítem ($a_j$), de modo que cuanto más discriminativo sea el ítem (cuanto más acusada sea la pendiente de la CCI) más información aporta para ese nivel $q$ .
- El error típico del ítem en el nivel $q$ , cuanto menor sea el error más informativo será el ítem en el nivel $q$ .

Además hay que especificar un criterio de finalización del TAI para obtener la estimación más precisa posible de las aptitudes de los estudiantes con el mínimo número de ítems. Como la función de información del ítem es dependiente de su parámetro de discriminación, un TAI clásico seleccionará siempre los ítems del banco con mayor poder discriminativo, provocando la sobreexposición de los mismos. Para evitar este inconveniente, mejor que ser tan estricto en este criterio de selección de ítems, *FastTEST Pro v. 1.6* permite la selección aleatoria de ítems de entre un número determinado del banco que aporten mayor información en el nivel de habilidad estimado. En nuestro trabajo, para la estimación del parámetro de conocimiento o aptitud se empleó el método de máxima verosimilitud al igual que se hizo con los ítems.

Ahora bien, ¿cuándo finaliza la aplicación del TAI? Se pueden definir distintas estrategias:

- Fijar el número máximo de ítems que se administrarán.
- Establecer el error de estimación de aptitud máximo que no se desea alcanzar.
- Especificar un valor de corte de habilidad y el rango del error típico que se quiere en torno al parámetro de aptitud.

En función de las necesidades del evaluador, se optará por una estrategia u otra o incluso por una combinación de todas ellas. En general, en los tests que evalúan conocimiento, la estrategia que se utiliza es la primera, es decir, se determina a priori el número de ítems a aplicar al sujeto. Dado que los contenidos no son necesariamente los mismos al ir adaptándose cada ítem al nivel del sujeto, el ahorro de ítems y de tiempo se convierten en una gran ventaja cuando se trata de evaluaciones masivas, siendo la información que se obtiene de los sujetos muy precisa. Sin embargo, dado que una de las ventajas que presentan los TAIs frente a los tests convencionales es la precisión de medida, para la finalización de nuestro test se ha optado por fijar el error máximo permitido en la estimación de la habilidad de Psicometría, un valor igual a 0,20.

El resultado del estudiante en el test queda reflejado en un informe en el que aparecen detallados los siguientes datos:

- El nivel de conocimiento estimado del estudiante, tras el acierto o el fallo al ítem presentado, así como el error típico de estimación.
- Los ítems del banco que se le han mostrado al estudiante.
- El valor final del parámetro de aptitud y el error típico asociado a él, datos sobre los cuales se considerará a un estudiante como apto o no en el rasgo evaluado según criterio del evaluador.

La Figura 1 muestra los resultados de un sujeto tras aplicarle el TAI.

Figura 1. Informe de los resultados del sujeto 1

Theta was estimated by maximum likelihood.

The standard error band plotted as ---- is plus or minus 2.00 standard errors.
X = Initial theta value C = Correct answer I = Incorrect answer

```
Item  Theta  SE   -3.......-2.........-1.........0.........+1.........+2.........+3
 0  -0.30* 1.00*        --------------------X--------------------
 1  -4.00* 1.00*  <--------------------          .
 2  -1.19  1.00*  ------------------C----------------------
 3  -1.96  1.00*  ---------I--------------------
 4  -1.39  0.91   ---------------C------------------
 5  -1.08  0.73     ---------------C--------------
 6  -0.85  0.64       -------------C-------------
 7  -0.65  0.59         ------------C------------
 8  -0.89  0.55         -----------I-----------
 9  -0.77  0.51          ----------C----------
10  -0.65  0.48          ----------C----------
11  -0.82  0.46          ---------I----------
12  -0.75  0.44          ---------C---------
13  -0.65  0.42           --------C--------
14  -0.75  0.41           --------I--------
15  -0.68  0.39           --------C-------
```

*Arbitrarily assigned value.

The final theta estimate based on 15 items was -0.68 with a standard error of 0.39, resulting in a 2.00 standard error band of -1.47 to 0.10

| Item | Correct Answer | Answer | Correct? | Theta | SE | SE Band -2.0 SE | SE Band +2.0 SE | Item Identifier |
|---|---|---|---|---|---|---|---|---|
| 0 | | | | -0.30* | 1.00* | -2.30* | 1.70* | |
| 1 | 1 | 2 | N | -4.00* | 1.00* | -6.00* | -2.00* | PSICOMETRÍA\|BAREMOS\| \|66 |
| 2 | 2 | 2 | Y | -1.19 | 1.00* | -3.19 | 0.81* | PSICOMETRÍA \|BAREMOS\| \|70 |
| 3 | 1 | 2 | N | -1.96 | 1.00* | -3.96 | 0.04* | PSICOMETRÍA \|FIABILIDAD\| \|33 |
| 4 | 2 | 2 | Y | -1.39 | 0.91 | -3.20 | 0.42 | PSICOMETRÍA \|FIABILIDAD\| \|47 |
| 5 | 2 | 2 | Y | -1.08 | 0.73 | -2.54 | 0.39 | PSICOMETRÍA \|ITEMS\| \|17 |
| 6 | 2 | 2 | Y | -0.85 | 0.64 | -2.14 | 0.43 | PSICOMETRÍA \|FIABILIDAD\| \|43 |
| 7 | 2 | 2 | Y | -0.65 | 0.59 | -1.83 | 0.54 | PSICOMETRÍA \|ITEMS\| \|62 |
| 8 | 1 | 2 | N | -0.89 | 0.55 | -2.00 | 0.21 | PSICOMETRÍA \|FIABILIDAD\| \|38 |
| 9 | 2 | 2 | Y | -0.77 | 0.51 | -1.79 | 0.25 | PSICOMETRÍA \|CONCEPTOS\| \|09 |
| 10 | 2 | 2 | Y | -0.65 | 0.48 | -1.62 | 0.31 | PSICOMETRÍA \|BAREMOS\| \|76 |
| 11 | 1 | 2 | N | -0.82 | 0.46 | -1.74 | 0.10 | PSICOMETRÍA \|FIABILIDAD\| \|44 |
| 12 | 2 | 2 | Y | -0.75 | 0.44 | -1.62 | 0.13 | PSICOMETRÍA \|CONCEPTOS\| \|15 |
| 13 | 2 | 2 | Y | -0.65 | 0.42 | -1.50 | 0.20 | PSICOMETRÍA \|FIABILIDAD\| \|57 |
| 14 | 1 | 2 | N | -0.75 | 0.41 | -1.57 | 0.06 | PSICOMETRÍA \|FIABILIDAD\| \|39 |
| 15 | 2 | 2 | Y | -0.68 | 0.39 | -1.47 | 0.10 | PSICOMETRÍA \|BAREMOS\| \|78 |

Dado que, para comenzar el TAI, le pedimos al programa que hiciera una estimación inicial en un rango de valores de -1 a +1, el primer ítem que administra tiene una dificultad cercana a la neutralidad ($b_j$=-0.30). Al mostrar una respuesta que muestra un desacierto, el siguiente ítem es de dificultad inferior. Al contestar este ítem mostrando un acierto, el siguiente es más difícil. Este proceso continúa hasta que se consigue estimar la habilidad (conocimientos en psicometría) con un error inferior o igual al fijado por el evaluador. Con un error de estimación de 0.39, la habilidad del sujeto es -0.68, para lo cual ha contestado a 15 ítems. Teniendo en cuenta que el rango de valores de la habilidad es [-3,+3], podríamos decir que el nivel de conocimientos de este sujeto se encuentra en torno a la media.

## *5. Mantenimiento y renovación del TAI*

En definitiva, la creación de un TAI lleva asociadas una serie de ventajas, que no sólo se restringen al ahorro de ítems y tiempo. En efecto, los TAIs nos permiten conocer las estrategias y mecanismos de respuesta del examinado con mayor flexibilidad que los cuestionarios convencionales, permitiéndonos incluso el diagnóstico de problemas de aprendizaje. Como

señalan Olea y cols. (1999), un TAI nos permite un nivel de estandarización en las medidas obtenidas muy superior al que se obtendría con esfuerzo equivalente en un TC.

Sands y Waters (1997) describen el potencial de los más de 1.000 *testing centres* fruto de la combinación de la metodología TAI y las redes de comunicación, así como la imagen pública de alta tecnología y sofisticación asociada a las instituciones que aplican una evaluación informatizada como los TAIs.

En nuestra investigación hemos conseguido un banco de 111 ítems, seleccionados tras un riguroso escrutinio de una muestra inicial de 224, que abarcan todo el rango de valores de dificultad y con óptimas estimaciones en el parámetro de pseudo-azar (Núñez, Cañadas y Tirado, 2004; Núñez, Cañadas y Tirado, 2005). Sin embargo, antes de administrar el test como prueba adaptativa se necesita ponerlo a prueba para comprobar si el software gestiona adecuadamente el TAI y la consecución de una serie de objetivos para los que se solicita la ayuda:

- Evaluar el TAI construido con el banco de 111 ítems del que se dispone actualmente.
- Ampliar el banco, incorporando ítems que amplíen el rango de valores los parámetros de dificultad, discriminación y pseudo-azar en las cuatro áreas de conocimiento.
- Observar la sobreexposición e infrautilización de los ítems del banco.
- Aumentar la información del banco añadiendo ítems más discriminativos.
- Probar la validez de criterio.
- Comprobar cuál sería la mejor estrategia para finalizar el TAI.

Para ampliar el banco de ítems se recurrirá a muestras de estudiantes de Educación Secundaria y Grado Superior de Formación Profesional Específica, quienes contestarán a los cuestionarios en formato de lápiz y papel. Para el resto de objetivos, la muestra de estudio serán los estudiantes de 1º de Psicología en los años académicos 2005/06 y 2006/07.

**Referencias Bibliográficas**


Abad, F. J., Olea, J., Real, E. y Ponsoda, V. (2002). Estimación de habilidad y precisión en tests adaptativos informatizados y tests óptimos: Un caso práctico. *Revista Electrónica de Metodología Aplicada*, *7*, 1-20.

Ackerman, T.A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement, 29*, 67-91.

Ansley, T.N. y Forsyth, R.A. (1985). An examination of the characteristic of unidimensional IRT parameters estimates derived from two-dimensional data. *Applied Psychological Measurement, 9*, 37-48.

Assesment Systems Corporation (1995). *XCALIBRE: Marginal maximum-likelihood estimation for the 2- and 3- parameter logistic IRT model* (version 1.10). Saint Paul, MN: Autor.

Assesment Systems Corporation (1996). *User's manual for MicroCAT. The MicroCAT™ testing system*. Saint Paul, MN: Autor.

Assesment Systems Corporation (2002). *User's manual for the FastTEST Professional Testing System* (version 1.6). Saint Paul, MN: Autor.

Barbero, M.I. (1999). Gestión informatizada de bancos de ítems. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados. Fundamentos y aplicaciones* (pp. 63-83). Madrid: Pirámide.

Bock, R.D. y Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of AM algorithm. *Psychometrika, 46*, 443-459.

Bock, R.D. y Lieberman, M. (1970). Fitting a response model for n dicotomously scored items. *Psychometrika, 35*, 179-197.

Green, B. F. Jr. (1983). Notes on the efficacy of tailored tests. En H. Wainer & S. Messick (Eds.), *Principals of Modern Psychological Measurement*, Hillsdale, N.J.: Lawrence Erlbaum Associates.

Hambleton, R.K. y Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.

Kok, F.G. (1988). Item bias and test multidimensionality. En R. Langeheine y J. Rost (Eds.), *Latent trait and latent class models* (pp. 263-275). New York: Plenum.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Mislevy, R.J. y Bock R.D. (1990). *PC-BILOG 3.04: Item analysis and test scoring with binary logistic models*. Mooresville, IN: Scientific Software.

Núñez, R. M., Cañadas, I. y Tirado, S. (2004). *Estudio piloto sobre la construcción de una prueba adaptativa de acceso a la titulación de Psicología*. Comunicación presentada en el III Congreso de la Sociedad Internacional de Profesionales de Metodología de Encuestas. Granada, septiembre 2004.

Núñez, R. M., Cañadas, I. y Tirado, S. (2005). *Effect of the number of response alternatives on item parameters in IRT*. Póster presentado en el 9th European Congress of Psychology. Granada, julio 2005.

Olea, J., Abad, F.J., Ponsoda, V. y Ximénez, M.C. (2004). Un test adaptativo informatizado para evaluar el conocimiento del inglés escrito: Diseño y comprobaciones psicométricas. *Psicothema, 16*, 519-525.

Ponsoda, V., Olea, J. y Revuelta, J. (1994). ADTEST: a computer-adaptive test based on the maximum information principle. *Educational and Psychological Measurement, 54*, 680-686.

Ponsoda, V., Wise, S., Olea, J. y Revuelta, J. (1997). An Investigation of Self Adapted Testing in a Spanish High School Population. *Educational and Psychological Measurement. 57*, 2, 210-221.

Reckase, M.D., Carlson, J.E., Ackerman, T.A. y Spray, J.A. (1986). *The interpretation of unidimensional IRT parameters when estimated from multidimensional data*. Paper presented at the annual meeting of the Psychological Society, Toronto, Canada.

Renom, J. y Doval, E. (1999). Tests adaptativos informatizados. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados. Fundamentos y aplicaciones* (pp. 127-162). Madrid: Pirámide.

Revuelta, J. y Ponsoda, V. (1998). Un test adaptativo informatizado de análisis lógico basado en la generación automática de items. *Psicothema, 10*, 3, 753-760.

Shealy, R.T. y Stout, W.F. (1993). An item response theory model for test bias and differential test functioning. En P.W. Holland y H. Wainer (Eds.), *Differential item functioning* (pp. 197-239). Hillsdale, NJ: LEA.

Wainer, H. (2000). CATs: Whither and whence. *Psicológica, 21*, 121-133.

Yen, W.M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5*, 245-262.

Yen, W.M. (1984). Effect of item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*, 125-145.