

LOS BANCOS DE ITEMS EN EL ESCENARIO ACTUAL DE LA MEDICIÓN PSICOLÓGICA CON TESTS

J. GABRIEL MOLINA¹, JAIME SANMARTÍN¹
E IGNACIO PAREJA²

¹Universitat de València / ²Universitat Politècnica de València

Resumen

Los bancos de ítems representan un concepto ya arraigado en el campo de la medición psicológica con tests, si bien, su peso dentro de este campo ha crecido significativamente en estos últimos años. Este trabajo revisa el papel que han pasado a desempeñar los bancos de ítems en el actual escenario psicométrico y, para ello, se ha planteado un doble nivel de aproximación: en primer lugar, una revisión de la evolución acaecida en el campo de los tests y, a partir de esta revisión, el análisis de la relación de los bancos de ítems con algunos de los elementos más distintivos del presente de este campo (teoría de respuesta al ítem, tests informatizados y estrategias adaptativas de aplicación de los tests).

Palabras clave: Psicometría, Tests, Banco de ítems

Abstract

Item banks are no new subject in the field of testing, but their presence has become increasingly more significant in this field from a few years ago. This work has studied the role that item banks play in the present psychometric scene. In this aim, it has been carried out a two level approach: first, a review of the evolution happened in the field of testing and, from the results of this, an analysis of the relation between item banks and some of the more outstanding contents in the present of this field (item response theory, computerized tests and adaptive strategies of test administration).

Key words: Psychometrics, Testing, Item-bank.

Introducción

En el devenir del campo de la medición psicológica con tests, los años que nos preceden en esta segunda mitad de siglo han venido caracterizados por un flujo de aportaciones que han contribuido a un substancial progreso en los fundamentos formales que sustentan la construcción de los tests. También resulta revelador el hecho de que un número importante de estas contri-

buciones aparezcan, de algún modo, vinculadas a la expansión de las nuevas tecnologías de la información. En cualquier caso, lo más significativo de esta evolución es que ha arrastrado consigo el planteamiento de nuevas estrategias en la aplicación de los tests que han perseguido hacer de éstos instrumentos cada vez más eficientes en su cometido.

En esta evolución, los bancos de ítems representan un concepto ya clásico en la práctica psicométrica (ver, p. ej., Wood y Skurnik, 1969), si bien, en años recientes han adquirido un especial protagonismo -fundamentalmente, en cuanto que componente básico de los prominentes tests adaptativos informatizados (TAIs). La adaptación continuada a las respuestas de los sujetos que se da en este tipo de tests requiere de un cúmulo de ítems de propiedades conocidas - un banco de ítems- de donde seleccionar el más apropiado en cada momento.

Ahora bien, el emerger deslumbrante de los TAIs ha desvanecido un tanto la consideración de una de las piezas básicas sobre las que éstos se apoyan, los bancos de ítems. Con este trabajo se pretende ofrecer un análisis del papel que han pasado a desempeñar los bancos de ítems en el contexto de la actual medición psicológica con tests. Para ello, hemos acometido una revisión de la evolución del contexto en que los bancos de ítems se enmarcan, para después señalar la relación de éstos con algunos de los elementos clave que caracterizan el presente de la psicometría.

Evolución en la teoría y uso de los tests

En este apartado se plantea una panorámica de la evolución acaecida en el área de conocimientos que tiene en los tests su centro de atención, abarcando desde un pasado reciente hasta un presente que, indefectiblemente, en el momento de leerse, será también ya parte del pasado -más o menos lejano en función del grado de actualidad alcanzado en esta revisión, de la distancia temporal con que se contemple o también, por supuesto, de la velocidad con que la evolución de los conocimientos haya recorrido esta distancia. En cualquier caso, esta revisión pretende delinear un escenario que sirva luego de referencia para establecer la localización de los bancos de ítems en el panorama actual de la medición psicológica con tests.

Con este fin, se plantean a continuación tres líneas directrices que haciendo hincapié, respectivamente, en aspectos teóricos, instrumentales y aplicados, pretenden ofrecer una perspectiva peculiar del devenir transcurrido en este contexto.

1. Desde la Teoría Clásica de los Tests a la Teoría de Respuesta al Ítem

Este primer apartado va a incidir en el desarrollo que se ha fraguado en el campo de la Teoría de los Tests, desarrollo sobre el que realmente descansa la evolución de muchos de los aspectos de carácter aplicado finalmente tangibles para los usuarios de los tests. El enunciado de este apartado se podría decir que expresa de un modo simplificado este devenir reciente en la teoría psicométrica, pues múltiples aportaciones con entidad propia han sido vertidas al campo de la Teoría de los Tests, si bien, se observa que con frecuencia se ha tendido a englobar a todas ellas bajo los epígrafes de la Teoría Clásica de los Tests (TCT) y la Teoría de Respuesta al Ítem (TRI); especialmente, la TCT parece cobijo de todas las aportaciones que no quedan encuadradas en el contexto más definido de la TRI. No obstante, este planteamiento más simplificado nos va a resultar útil aquí como medio para destacar los rasgos que caracterizan a la TRI, así como las consecuencias aplicadas derivadas de su utilización frente a otras aportaciones más estándares.

Múltiples trabajos han tratado los aspectos que han caracterizado esta evolución y aquí, en este apartado, no vamos a seguir más que la estela de éstos con el objetivo de establecer el

marco que contextualice la ubicación de los bancos de ítems. Entre estos trabajos se señalan a continuación algunos cardinales, aunque no representen más que una fracción de los que han sido consultados y que se centran en estos aspectos: Allen y Yen, 1979; Crocker y Algina, 1986; Hambleton y van der Linden, 1982; Hulin, Drasgow y Parsons, 1983; Martínez-Arias, 1995; Meliá, 1990; Mellenbergh, 1996; Muñiz, 1990; Navas, 1994; Santisteban, 1990; Traub y Lam, 1985; Weiss, 1983; Wright y Stone, 1979.

La historia reciente de la psicometría en su vertiente más teórica ha venido marcada por una significativa evolución en los modelos formales que han dado sustento al proceso de desarrollo y aplicación de los tests, siendo el principal exponente de esta evolución la introducción y difusión de la TRI. Los orígenes de esta teoría se pueden encontrar en trabajos ya publicados hace bastante tiempo (Lawley, 1943, 1944; Lord, 1952; Richardson, 1936; Tucker, 1946), siendo el libro *Statistical theories of mental tests scores* (Lord y Novick, 1968), la publicación divulgativa más emblemática de esta teoría. En cualquier caso, bien puede decirse que no ha sido hasta estos años que nos preceden que su difusión y utilización ha empezado a hacerse más patente, empezando a eclipsar el dominio psicométrico de la aún prominente TCT.

El origen y fundamento de la TCT aparece vinculado a diversos trabajos de Spearman a principios de este siglo (Spearman, 1904, 1907, 1913), donde introdujo diversos elementos claves en esta formulación teórica, como el modelo lineal aditivo de la puntuación verdadera, la teoría del error de medida y la fiabilidad. La obra de Harold Gulliksen (Gulliksen, 1950) representa una exposición global y sistemática de este enunciado teórico, el cual ha tenido una importante repercusión entre usuarios y desarrolladores de tests durante todo este siglo y, aún hoy en día, sigue teniendo una vigencia notable en la práctica de la medición psicológica y educativa. Su simplicidad ha constituido una de las razones de peso de lo extendido de su empleo.

A la base de la TCT, al igual que de las otras Teorías de Tests planteadas, ha subyacido el objetivo de explicar de un modo formal la relación existente entre la puntuación observada en el test y la posición del sujeto en el atributo medido (Gulliksen, 1961). Una función lineal -el modelo lineal de Spearman- es la que en la TCT se ha establecido para dar cuenta de esta relación, si bien, ha sido ampliamente cuestionada la supuesta independencia que se atribuye en este modelo entre las puntuaciones verdaderas y el error. No obstante, al ser pocas las restricciones o asunciones que se plantean en la aplicación de la TCT, su utilización resulta bastante tentadora.

Entre los diversos problemas derivados de la aplicación de la TCT que han sido citados, a continuación hacemos una reseña de los que consideramos más significativos.

- Los parámetros derivados de la aplicación de este modelo que van a caracterizar a los sujetos son dependientes de las características del conjunto de ítems administrado. De forma análoga, los valores de los parámetros que proporcionan información sobre las propiedades de los ítems (dificultad, discriminación, etc.) están vinculados al grupo de sujetos a partir del que se hayan obtenido los datos de respuesta objeto de análisis. Esta dependencia conlleva importantes consecuencias: resultará difícil comparar las puntuaciones de sujetos que han sido obtenidas a partir de tests diferentes, aunque esos tests midan lo mismo; así mismo, resultará también comprometido comparar los índices psicométricos correspondientes a dos ítems cualesquiera, cuando sus valores hayan sido obtenidos a partir de grupos de sujetos diferentes.

- Tanto el error típico de medida como la fiabilidad del test, los dos principales conceptos planteados en el análisis de la precisión de las medidas dentro de la TCT, han sido objeto de importantes críticas que han cuestionado su utilidad. La fiabilidad de un test es concebida como la relación entre las puntuaciones obtenidas en formas paralelas de ese test, planteamiento que genera serios inconvenientes a la hora de hacerse operativo. A los diversos coeficientes que han sido propuestos para estimar la fiabilidad de un test también les han sido señalados

propiedades no deseables, por ejemplo, su dependencia de la muestra de sujetos a partir de la que sea obtenido este valor. Por lo que respecta al error típico de medida, plantea el problema de que se define para el test en su conjunto, no para cada ítem particular y, además, es considerado el mismo para todos los sujetos, para todos los niveles en el atributo medido, algo difícilmente concebible.

- Un reproche más global hacia la TCT proviene del hecho de estar centrada en el test, no ofreciendo recursos que den cuenta de la respuesta de los sujetos a cada ítem concreto, lo cual limita las posibilidades de análisis de la unidad elemental de medida en los tests, el ítem. Esta limitación puede explicar la surgencia de modelos explicativos a nivel de ítem, tales como los que se engloban dentro de la TRI.

La formulación teórica de la TRI ha permitido superar algunos de los problemas señalados a la TCT, algo que se ha venido poniendo de manifiesto en un número creciente de publicaciones que han incidido en aspectos formales relacionados con los modelos de respuesta al ítem. Por contra, su utilización en la práctica aparece como más diluida, probablemente debido a la mayor complejidad asociada a su comprensión y aplicación: especialistas o profesionales de la medición son los que fundamentalmente han utilizado estos modelos en la práctica psicométrica. No obstante, una difusión cada vez mayor de esta teoría en libros sobre medición, evaluación o similares (ver referencias al respecto en un apartado anterior), así como la oferta de programas informáticos que dan soporte a la aplicación de los modelos psicométricos que se engloban en la TRI (ver, por ejemplo, Muñiz, 1990) ha ido acercando su comprensión y utilización a otros usuarios no expertos a los que puede resultar útil el desarrollo y aplicación de tests basados en esta teoría.

En la TRI la atención se centra en el ítem. El ítem, unidad elemental del test, constituye un estímulo con el que se pretende elicitarse una respuesta que permita estimar la posición de los sujetos en el atributo o rasgo que se pretenda medir, el cual se supone que subyace a la respuesta dada al ítem. La TRI engloba un conjunto de modelos formales que pretenden dar cuenta de la relación entre esa respuesta observable al ítem y el rasgo que se asume que subyace a esa respuesta, estando esta relación determinada en cada modelo por una función matemática, la conocida como curva característica del ítem.

La opción de aplicar un modelo de respuesta al ítem u otro, de entre el repertorio de ellos disponible, va a depender de las condiciones que caractericen el contexto de medida: hay modelos apropiados para ítems dicotómicos, otros para ítems politómicos, unos que asumen que el rasgo es unidimensional, otros que no, etc.. Ahora bien, la selección de un modelo de respuesta al ítem, en principio apropiado a la situación de medida, no avala el éxito de la utilización de éste como medio de explicar el funcionamiento de los sujetos en los ítems. El estudio del grado de adecuación de un modelo a la realidad empírica existente debe confirmar la bondad del mismo como modelo explicativo. Y en este sentido, cabe destacar que los modelos de la TRI plantean, en general, asunciones sobre los datos bastante restrictivas, lo cual complica la viabilidad de su aplicación, si bien, en el caso de que se pueda considerar como adecuada la aplicación de uno de éstos, la información que se obtiene resulta también muy estimable.

A continuación, se comenta el interés de algunos de los rasgos derivados de la aplicación de la TRI:

- Las estimaciones de los sujetos en el rasgo medido no dependen de los ítems aplicados a los mismos, siempre y cuando éstos hayan sido seleccionados de un conjunto de ítems previamente calibrados según la TRI. En este caso, independientemente del conjunto de ítems que sea administrado a un sujeto, su puntuación estimada será la misma -salvo variaciones debidas a errores de medida. Esta propiedad provee la capacidad de poder comparar las puntuaciones de sujetos que hayan contestado a diferentes subconjuntos de ítems, lo cual ha derivado en nuevas y potentes opciones en la práctica psicométrica.

- Las estimaciones de los parámetros de la curva característica del ítem, los cuales van a caracterizar psicométricamente a los ítems, van a ser independientes también de la muestra de sujetos a partir de la que éstos sean obtenidos. Salvo errores de medida, la calibración de los parámetros del ítem no dependerá de la distribución en el rasgo medido, esto es, del grupo de sujetos a los que sean aplicados los ítems. Esta propiedad ha representado la superación de uno de los más importantes escollos planteados en el uso de la TCT, generando posibilidades interesantes a nivel aplicado.

- El concepto de función de información, como medida de la precisión con la que cada ítem -o el test en su conjunto- mide el atributo correspondiente, constituye un poderoso recurso, pues la función de información describe la precisión asociada a la estimación de los distintos valores de la escala del atributo medido, a diferencia del invariante error típico de medida de la TCT. Además, permite valorar la contribución de cada ítem concreto a la precisión del test en su conjunto, lo cual representa una importante capacidad que ha tenido importantes repercusiones en el diseño de los tests.

- Un aspecto práctico no tan positivo asociado a los modelos formales encuadrados en la TRI es que el tamaño muestral sugerido, a fin de obtener estimaciones de confianza, puede resultar de difícil acceso dado lo elevado de la magnitud muestral habitualmente aconsejada en esta tarea. A esta dificultad también habría que añadir la observación más de fondo de Blinkhorn (1997), relativa al hecho de que los modelos de la TRI parecen ajustarse mejor al desarrollo de tests de rendimiento de corte educativo que a tests de capacidades y aptitudes psicológicas, si bien, es posible encontrar un número creciente de trabajos que aproximan esta teoría de tests y la práctica psicológica.

2. Desde los tests de «lápiz y papel» a los tests informatizados

La difusión de los ordenadores ha constituido otro de los pivotes sobre los que se ha catapultado la evolución del campo de los tests. Precisamente, la aplicación de los ordenadores a la evaluación psicológica con tests, así como a los laboratorios de psicología experimental, han constituido los primeros puntos de convergencia entre la Informática y la Psicología (López, Ato, Sánchez y Velandrino, 1986; Sanmartín y Algarabel, 1986). En este apartado vamos a exponer la progresiva incorporación de los ordenadores a las distintas tareas implicadas en el desarrollo y aplicación de los tests, junto con las consecuencias derivadas de su evolución en este campo. Este desarrollo, común por otra parte a todas las áreas de conocimientos receptivas a las aportaciones de las nuevas tecnologías, está provocando un uso cada vez menor del «lápiz y papel», en favor de una creciente utilización del «ratón y monitor» o *tests informatizados*, término que será utilizado para englobar todos aquellos aspectos del desarrollo o aplicación de los tests en que los ordenadores aparecen involucrados.

Los primeros sistemas informáticos desarrollados con una directa aplicación a la medición psicológica fueron aquéllos orientados a la corrección automatizada de las respuestas a determinados tests y a la obtención de las puntuaciones correspondientes a los sujetos a los que se les había aplicado. Se trataba de sistemas a los que había que introducir los datos a través de tarjetas perforadas y que se centraban en ciertos tests como el MMPI y otros de amplia difusión (Fowler y Butcher, 1987). Este tipo de programas, tan limitados a un aspecto concreto del trabajo con los tests, es el más conocido y difundido en la actualidad entre la comunidad de potenciales usuarios de estos recursos, tal como se desprende de la observación de los catálogos de las principales empresas que comercializan la distribución de material psicométrico. Diversas razones explican la distancia entre los desarrollos presentes y la difusión y aplicación de los mismos en el mercado, pero no va a ser el análisis de estas razones el objetivo de este apartado.

El desarrollo de los tests informatizados ha venido marcado por el raudo avance del campo de la informática, siendo especialmente notoria la difusión que los ordenadores han tenido a todos los niveles, así como la creciente sofisticación y potencia de los mismos. Cada vez mejores y más asequibles ordenadores han hecho aún más fácil y efectiva la convivencia entre tests y ordenadores y, por lo tanto, las positivas resultas derivadas de este maridaje. Un indicador de la evidencia del importante desarrollo de los tests informatizados lo constituye el hecho de que la APA extendiera sus reconocidos estándares sobre tests para tener en cuenta también a aquéllos informatizados (American Psychological Association, 1986).

El modo en que la expansión de los recursos informáticos se ha ido integrando en los diversos aspectos o tareas implicadas en el desarrollo y uso de los tests no ha sido uniforme. Ha incidido inicialmente en aquellas tareas más mecánicas, como es el caso de la corrección y puntuación de las respuestas de los sujetos, para ir extendiendo su soporte a otras tareas más significativas, tales como el análisis de las respuestas o la edición de los items. Pero, complementariamente, también ha hecho posible el desarrollo de nuevos procedimientos antes impensables, sea el caso de los tests «multimedia» o de las estrategias de administración adaptativas, de modo que el advenimiento de la informática no sólo ha dado respuesta a procesos ya existentes, sino que también ha posibilitado la apertura de nuevas vías de desarrollo en el campo de la medición con tests.

A continuación se van a revisar algunas de las aportaciones en el campo de los tests informatizados, muchas de las cuales han sido recogidas de forma coincidente por diversos autores que han tratado estos aspectos y a los que remitimos para alcanzar una panorámica más profunda de toda esta evolución (Baker, 1984; Burke y Normand, 1987; Cohen y Moreland, 1992; Eyde, 1987; Lippey, 1974; López, et al., 1986; Molina, Valero y Sanmartín, 1992; Moreland, 1986; Roid, 1986; Stager y Mueller, 1991; Wise y Plake, 1990; Wolfe, 1986). Éstas y otras contribuciones constituyen la base a partir de la que se ha desarrollado la síntesis con que a continuación se describe el actual vínculo entre tests y ordenadores. Por otra parte, referencias específicas a programas informáticos que de forma genérica se nombran a continuación se pueden encontrar en Baker (1989), García-Cueto (1996) y Olea, Ponsoda y Prieto (1998).

Los aspectos relativos a los tests informatizados que seguidamente se comentan han sido diferenciados en su presentación de acuerdo a distintas tareas que en el desarrollo y aplicación de los tests se pueden distinguir. Esta diferenciación se ajusta en buena medida a la funcionalidad de las aplicaciones informáticas pioneras en este campo pues, dadas las limitaciones del momento en que fueron desarrolladas, éstas se solían limitar a dar soporte a tareas concretas. Actualmente, cada vez se tiende más a desarrollar sistemas informáticos más globales y flexibles, que integran varias de las actividades implicadas en el proceso de diseño y aplicación de los tests.

- El *desarrollo de items* se ha visto menos afectado por la tecnología informática dado el carácter más *creativo* de esta tarea. Por contra, la edición de los items es una de las facetas que de un modo más directo e inmediato se ha visto influida por el potencial inherente a la utilización de los ordenadores, pues diversos programas de propósito general -procesadores de texto, programas de diseño, aplicaciones de digitalización de imágenes, etc.- tienen su aplicación también en la edición de los items, ofreciendo enormes posibilidades en el tratamiento de textos e imágenes, e incluso en otras formas de expresión como sonido y animaciones, así como en la combinación de ellos. Pero lo que consideramos aún más importante, la edición de los items pasa a ser accesible a los propios desarrolladores de los items, reduciendo la dependencia de especialistas en edición, fotocomposición o similares, lo cual confiere una agilidad muy valiosa a esta tarea. La generación automatizada de items representa también una línea de trabajo de la que se desprenden resultados cada vez más prometedores y que parece va a constituir uno de los principales núcleos de atención en un futuro próximo.

- También en el *almacenamiento y gestión de los ítems* se ha producido un sustancial cambio que ha dejado fuera de uso los clásicos ficheros de tarjetas, cada una con la información de un ítem. Programas de propósito general como bases de datos, e incluso procesadores de texto y hojas de cálculo, han sido los que más se han utilizado inicialmente en estas lides, si bien, nuevos programas específicamente desarrollados para el trabajo con cúmulos de ítems empiezan a dar soporte de un modo más ajustado a este tipo de tareas.

- La *construcción de tests* ha experimentado una importante evolución con la incorporación de los ordenadores en este cometido. La búsqueda o selección de ítems con determinadas características dentro de un fichero o banco de ítems resulta tarea sencilla, a la vez que eficiente, si se realiza con la ayuda de alguna aplicación informática: aun en conjuntos de ítems numerosos y con criterios de búsqueda complejos, la localización de los ítems deseados no será más que cuestión de instantes. Ello ha proporcionado el recurso instrumental necesario que ha facilitado la posibilidad de construir tests a medida, esto es, un tipo de test en que los ítems que lo conforman responden a los criterios que el constructor del test haya establecido, de acuerdo a las necesidades y condiciones de la situación concreta de medición a la que se enfrente. Esta ha resultado ser una de las contribuciones más significativas en la evolución reciente del campo de los tests informatizados pues, junto a nuevas posibilidades en la administración de los tests que seguidamente se comentan, ha proporcionado las bases sobre las que se han desarrollado nuevas formas de tests, antes inviables, a los que se suele hacer referencia con el término de tests adaptativos informatizados -en adelante, TAIs.

- En la *administración de los tests* se ha visto también implicado el ordenador, incluidos los diversos periféricos de *input* y *output* asociados a éste, haciendo posible que los tests puedan ser completamente aplicados a través de un ordenador. La administración de los ítems vía ordenador ha permitido el aprovechamiento de los canales visuales y auditivos de presentación de la información con que la mayoría de los ordenadores cuentan en la actualidad, lo cual ha abierto las puertas a nuevos tipos y formatos de ítems que han redundado en un incremento en la validez de la medición pretendida. A su vez, estas nuevas posibilidades han abierto el camino a la aplicación de tests a personas con disfunciones físicas.

Otro aspecto destacado de la administración informatizada de los tests ha sido el hecho de facilitar un mejor control en la presentación de los ítems en aspectos tales como las instrucciones de ejecución del test, el tiempo de exposición de los ítems o el tiempo límite de realización de los mismos, lo cual favorece una mayor estandarización en el modo en que los tests son presentados a los sujetos. No obstante, esta resaltada cualidad de la estandarización debe ser objeto de un control riguroso, sobre todo cuando un mismo test sea administrado a través de distintos ordenadores, pues pequeñas diferencias en la configuración de los mismos -que pueden pasar fácilmente desapercibidas- podrían provocar diferencias significativas en las condiciones de administración. Realmente, la estandarización en las condiciones de presentación en los tests informatizados puede resultar más problemática que la de los tests tradicionales, a pesar de que se observa cierta tendencia a considerar lo contrario como cierto.

Por otra parte, aun cuando los tests sean administrados de modo convencional (tests de lápiz y papel), modo que sigue siendo el principalmente utilizado en la práctica, podemos encontrar diversas aplicaciones informáticas desarrolladas con el fin de dar soporte al proceso de producción de este tipo de tests. Varios programas permiten la edición e impresión de tests con una apariencia óptima, así como configurar la disposición de los ítems seleccionados en el test, permitiendo incluso generar diversas formas de un mismo test en que la posición de los ítems o el orden de las alternativas de respuesta es aleatorizado. Por otra parte, la introducción de los datos de respuesta a estos tests convencionales en un ordenador ha sido tarea que también ha sido objeto de mecanización a través del uso de lectoras ópticas y de programas de reconocimiento de texto, si bien, para ello se requiere disponer de recursos adicionales de cierto coste económico.

- La aplicación informatizada de un test plantea la posibilidad de que las respuestas sean introducidas por los propios sujetos a través de algunos de los dispositivos de *input* del ordenador (teclado, ratón, tableta gráfica, pantalla activa, etc.). La *recogida de las respuestas* por el propio ordenador va a permitir que de un modo inmediato dispongamos de los datos de respuesta a los ítems en formato digital. Ello permitirá que estos datos puedan ser después procesados informáticamente para su corrección, puntuación, análisis, etc., ahorrando el trabajo que representa su introducción a posteriori en el ordenador y evitando errores típicos asociados a ese proceso de "picado" de los datos. Otras interesantes consecuencias se derivan de la utilización del ordenador como instrumento de recogida de las respuestas de los sujetos al test: la reducción de las equivocaciones derivadas del hecho habitual de que los sujetos introduzcan sus respuestas en una hoja o espacio separado de aquél en que abordan los ítems; y, también, la capacidad para captar otras facetas de las respuestas de los sujetos, como el tiempo empleado en iniciar la respuesta a un ítem, el tiempo total de respuesta u otras formas de responder como pueda ser a través de la voz, la pulsación táctil o, tal vez en un futuro próximo, la fijación visual.

- El procesamiento informatizado de los datos de respuesta de los sujetos a los ítems ofrece un importante potencial en la *corrección* y obtención de la *puntuación* de los sujetos, potencial que ha sido aprovechado mayormente en la aplicación de la tecnología informática al campo de los tests. Gran rapidez y reducido costo son cualidades inmediatas derivadas de esta forma de trabajo, a las que hay que añadir la reducción de errores en el proceso de corrección y la posibilidad de proporcionar *feedback* inmediato a los sujetos. A estas mismas cualidades se sumaría, como factor a destacar en la obtención automatizada de las puntuaciones, la potencia de los ordenadores en el cálculo, lo cual ha facilitado la posibilidad de obtener las puntuaciones de los sujetos según diversos procedimientos de estimación de las mismas, algunos de ellos de considerable complejidad operativa.

- Junto a la obtención de las puntuaciones, algunos sistemas informáticos cubren también la *elaboración de informes* que describen esas puntuaciones. Éstos son de complejidad diversa, variando desde aquéllos en que tan sólo aparece el nombre o referencia de cada sujeto y la puntuación directa correspondiente, hasta los que obtienen múltiples puntuaciones, estadísticos e incluso elaboran exhaustivas interpretaciones a partir de esos resultados. Estos informes tan detallados han sido origen frecuente de controversia y polémica, motivada en muchas ocasiones por el mal uso o una confianza excesiva sobre los mismos.

- Para finalizar, una tarea en la que los ordenadores juegan ya desde hace tiempo un importante papel es en el *análisis psicométrico* implicado en el desarrollo y mantenimiento de los tests. Estos análisis suelen implicar cálculos estadísticos, en ocasiones complejos, en los que diversos programas informáticos, específicamente desarrollados para dar respuesta a esta tarea, brindan una poderosa ayuda y el medio de hacer viable la ejecución de los mismos. La tradición en el uso de esta clase de aplicaciones informáticas en los círculos de investigación y desarrollo de tests ha motivado la existencia de un repertorio relativamente amplio de este tipo de programas, pudiendo incluso encontrarse algunos que cubren la realización de diversas pruebas y análisis desde la perspectiva de diferentes modelos psicométricos.

3. Desde los tests convencionales a los tests adaptativos

La aplicación de estrategias adaptativas en la administración de los tests constituye el tercer aspecto que vamos a destacar como hecho especialmente revelador de la evolución reciente en el campo de los tests. A continuación se presenta una revisión de los aspectos que hemos considerado más significativos en esta faceta de la evolución de los tests y que se ha visto plasmada en una extensa línea de trabajo en la que algunas aportaciones han constituido puntos de referencia básicos a la hora de elaborar el contenido de este apartado (Bunderson, Inouye

y Olsen, 1989; Green, 1983; Hambleton, Zaal y Pieters, 1991; Lord, 1980; Lunz, Bergstrom y Gershon, 1994; Olea, Ponsoda y Prieto, 1998; Reckase, 1989; Renom, 1993; Wainer, 1990; Weiss, 1983, 1985; Weiss y Vale, 1987).

Mientras que en un test convencional siempre son aplicados los mismos ítems y en el mismo orden en las distintas administraciones que se haga del mismo, en un tests adaptativos o "a medida" se plantea una administración que se adapta a las características del sujeto(s) al que es aplicado el test y, por lo tanto, es de esperar que los tests aplicados a los distintos sujetos sean diferentes. Esta adaptabilidad en los tests puede darse respecto a distintos aspectos, tales como: el contenido de los ítems; su nivel de dificultad -o de otros parámetros que caractericen a los ítems-; el tiempo disponible para su realización; u otros que se consideren oportunos.

Dos aspectos adicionales que suelen caracterizar la aplicación de un test adaptativo son el criterio de arranque y el criterio de finalización del mismo. El primero permite establecer el punto de partida del test para cada sujeto, pudiendo ser éste un ítem al azar, uno predeterminado, uno de valor promedio en el rango del parámetro en función del que se adapte el test, u otro. Respecto al criterio de finalización, éste es el que va a determinar cuándo la aplicación de la prueba se dará por finalizada. Entre los criterios concretos que se han venido utilizando para este fin podemos señalar los siguientes: el número total de ítems administrados, el nivel de precisión alcanzado en la medida, el grado de consistencia de las respuestas, etcétera.

Hay que resaltar, en cualquier caso, que ni el concepto de test adaptativo es nuevo, ni tests convencionales y adaptativos son conceptos antagónicos, sino más bien calificativos que definen los extremos de un continuo en el que se podrían situar las múltiples diseños de administración de tests que han sido propuestos. Así, en el precursor trabajo de Alfred Binet (Binet, 1909) ya pueden encontrarse los principios de los tests adaptativos, en cuanto que consideraba la administración de una primera prueba a todos los sujetos y en función de los resultados en ésta, se administraba una segunda parte específica. También diversos tests clásicos poseen formas distintas orientadas a grupos de características diferenciadas (edad, clase social, profesión, etc.) y, en este sentido, podrían ser considerados también como adaptativos. Otros tests han sido desarrollados con una clara vocación adaptativa pero, en cualquier caso, se trata de formas de adaptabilidad más o menos rudimentaria, por lo menos en comparación con la significación que este término alcanza en algunos de los tests que actualmente se están desarrollando y aplicando. Estos tienen su mayor exponente en los genéricamente conocidos como TAIs, tests en los que la adaptación del test llega a darse a partir de la respuesta a cada ítem.

Uno de los gérmenes y punto de inflexión en esta evolución de los tests convencionales a los tests adaptativos lo podemos encontrar en el aprovechamiento del potencial derivado de la utilización de los ordenadores en la aplicación de los tests. Éste ha hecho factible un nivel de adaptabilidad en los tests a nivel de ítem, pues procesos automatizados de administración, recogida de las respuestas, codificación, corrección, puntuación, búsqueda de ítems apropiados y control de todo el proceso -incluida, por supuesto, la finalización del mismo- hacen posible que la respuesta a un ítem pueda ser tomada en cuenta, ya de forma inmediata, en la selección de los subsiguientes ítems que se presenten, así como que la presentación de los ítems no se vea afectada por dilaciones que afecten la deseable continuidad en la aplicación de un test.

También hay que destacar como elemento esencial del desarrollo de los tests adaptativos actualmente en boga, la difusión y aplicación de los modelos de la TRI, hasta tal punto que los tests adaptativos han llegado a ser contemplados como una derivación de esta teoría psicométrica. Esto no puede considerarse del todo correcto ya que también desde la TCT pueden desarrollarse tests adaptativos, sin embargo, algunas de las cualidades derivadas de la aplicación de la TRI resultan fundamentales a la hora de llevar a la práctica ciertas estrategias adaptativas, especialmente la posibilidad de situar sobre una misma escala las puntuaciones obtenidas a partir de diferentes subconjuntos de ítems. También el concepto de función de información ha

jugado un papel importante en la aplicación de los tests adaptativos, como característica en función de la que los ítems son sucesivamente seleccionados. A las aportaciones capitales de Allan Birnbaum (Birnbaum, 1968) y Frederic M. Lord (Lord, 1970) han seguido otras muchas en las que se ha resaltado y contrastado la aplicabilidad de los modelos de respuesta al ítem en el desarrollo de tests adaptativos.

La evolución en los tests adaptativos ha tenido en los TAIs su estandarte más visible. Concretamente, ha sido en Estados Unidos donde inicialmente se ha manifestado una mayor atención hacia el tema de los TAIs, pudiendo situar como epicentro de todo este movimiento a la línea de investigación encabezada por F. M. Lord en el *Educational Testing Service*, ya desde finales de la década de los 60. En nuestros días, el interés por esta forma de tests es ya generalizado, siendo abundantes y de procedencia dispersa las aportaciones que actualmente se vierten sobre esta temática. Así, por ejemplo, en nuestro país podemos destacar, entre otras: el proyecto desarrollado a partir de 1987 por el Laboratorio de Psicología Aeronáutica de la Universidad de Salamanca en colaboración con el Servicio de Psicología del Ejército del Aire (Prieto, Carro, González-Tablas, Fernández y Orgaz, 1990); la línea de trabajo dentro de la Universitat de Barcelona que ha tenido entre sus aportaciones más significativas el texto *Tests adaptativos computerizados: fundamentos y aplicaciones* y el programa informático *DEMOTAC* (Renom, 1993); la que tiene su origen en la Universidad Autónoma de Madrid y que culminó, entre otras contribuciones, en el desarrollo de un test adaptativo informatizado (Ponsoda, Olea y Revuelta, 1993); la que desde la UNED se ha centrado en la evaluación de la capacidad matemática (ver, p. ej., Barbero y Navas, 1995); la que ha tenido su ámbito de aplicación en la Comunidad de Andalucía en la evaluación del rendimiento en el área de Ciencias Sociales (ver, p. ej., García, Gil y Rodríguez, 1997); o la de los propios autores, centrada en la gestión y desarrollo informatizado de bancos de ítems (ver, p. ej., Molina, Sanmartín y Pareja, 1998). Otras aportaciones relevantes quedan en el tintero pero no es el objetivo aquí realizar una revisión de éstas.

Los TAIs han sido aplicados, fundamentalmente, en el campo de la medición de capacidades, aptitudes y rendimiento, siendo la dificultad de los ítems el aspecto que se suele considerar a la hora de adaptar el test a las características de los sujetos. También podemos encontrar aportaciones relacionadas con la medición de actitudes o de la personalidad, pero éstas son por ahora más limitadas dada la mayor dificultad asociada a la evaluación de estos aspectos. En la práctica ha sido en el ejército, la administración pública y en el campo educativo donde más profusión ha tenido la aplicación de instrumentos psicométricos de carácter adaptativo. En general, se trata de campos donde se plantean evaluaciones a gran escala y en que suele existir una mayor dotación de recursos que hace posible el desarrollo y la aplicación de procedimientos novedosos orientados a alcanzar una medición más satisfactoria.

Diversas consecuencias -algunas favorables, otras no tanto- han sido destacadas como derivadas del uso de los tests adaptativos, más concretamente, de los TAIs, aquéllos en que más genuinamente se da la adaptación del test a las contestaciones de los sujetos. A continuación, vamos a resumir algunas de las que consideramos más significativas.

- La aplicación de diseños adaptativos representa una optimización en el ajuste entre las tareas o ítems implicados en el test y el nivel en la capacidad medida de los sujetos a los que es administrado ese test. De este modo, se evita la administración de ítems que proporcionan poca información en la estimación de la capacidad de cada sujeto concreto y, por lo tanto, la típica situación de los tests convencionales en que sujetos de alta capacidad se enfrentan y "aburren" ante la abundancia de ítems, para ellos, sencillos; y lo opuesto ocurre con los de baja capacidad, en los que se puede llegar a generar cierto desaliento ante la autopercepción de incapacidad frente a una mayoría de ítems dificultosos. La convergencia en los TAIs entre el nivel de dificultad -o de aquel criterio que se considere- y el nivel del sujeto en el rasgo medido

representa una vía de solución a esta situación, constituyendo un factor motivante para todos los sujetos en la ejecución de la prueba -especialmente para los sujetos que se sitúen en los extremos de la distribución del atributo medido.

No obstante, con este nuevo esquema evaluativo también se ha puesto de manifiesto algún conflicto en su aplicación a colectivos. En un test adaptativo típico, en que el criterio de adaptación de los ítems viene determinado por la dificultad de los mismos, la mayoría de los sujetos tiende a acertar la mitad de los ítems en sus respectivos tests. Motivado por ello, la percepción de los propios sujetos de su nivel de ejecución tras finalizar el test puede ser similar -dada la adaptación del nivel de los ítems a cada sujeto-, por lo que el conocimiento posterior de los resultados o puntuaciones finales puede contrariar las expectativas de algunos de ellos y generar la convicción de que se ha producido una corrección errónea.

- Otro aspecto que, de forma paralela al anterior, se verá favorecido por la adaptación del test a los sujetos será el de la precisión de las medidas derivadas de su aplicación. Frente a los tests convencionales, de los que se ha puesto de manifiesto que proporcionan medidas más precisas para aquellas puntuaciones situadas en torno a la puntuación media en el test, los tests adaptativos ofrecen medidas de similar precisión en todo el rango de puntuaciones dado el carácter máximamente informativo de los ítems que recibe cada sujeto. Es más, se puede establecer un criterio de finalización de un TAI en función de que la precisión con que se estimen las puntuaciones de los sujetos sea la misma para todos ellos, o no inferior a un determinado valor crítico.

- Una de las características más positivamente resaltadas de estos tests ha sido el menor número de ítems necesario para obtener grados de precisión similares, en relación a otras formas de tests no adaptativas. Ello tiene importantes consecuencias en la aplicación de los tests pues supone una reducción en el tiempo de administración de los mismos y, en consecuencia, también en la fatiga que puede provocar a los sujetos y que tan negativamente puede llegar a afectar la ejecución de una prueba. Por supuesto, también tiempos de administración más reducidos afectarán positivamente los costes asociados a la aplicación del test.

La investigación ha puesto de manifiesto una reducción en torno al 35% en la longitud de los TAIs respecto a sus formas convencionales homólogas, valor apreciable y que en algunas ocasiones llega a ser mucho mayor, pero que también ha hecho emerger la cuestión de la robustez, esto es, el importante impacto que pueden tener ítems “malos” en éstos tan ajustados tests en su longitud, frente a tests convencionales en que el efecto de ítems que no funcionan muy bien puede quedar amortiguado por un numeroso remanente de ítems adecuados. En definitiva, el imperativo de desarrollar buenos ítems parece aún más esencial cuando éstos tengan que formar parte de tests adaptativos.

- La administración de estos tests requiere una serie de recursos, en principio, costosos. Tanto los programas que den soporte a ese tipo de tests, como el ordenador, monitor y demás periféricos que puedan requerirse, suponen una inversión económica inicial que implica la disponibilidad del capital necesario para ello. Aunque la actual evolución tiende a que éste sea cada vez menos cuantioso, lo cierto es que en bastantes ocasiones puede representar un obstáculo suficiente para que se acabe optando por otra alternativa.

- La seguridad de los tests también ha sido un aspecto señalado como positivamente afectado por la aplicación de los TAIs, dado que cada sujeto recibe un test concreto y, por lo tanto, puede resultar poco efectiva la transmisión de ítems memorizados a otros sujetos. Es más, es poco probable que un mismo sujeto reciba en dos ocasiones diferentes el mismo conjunto de ítems, aunque ello dependerá en gran medida de lo cuantioso del número de ítems a partir del que se desarrolle el TAI. Por otra parte, observamos como determinados recursos informáticos nos proporcionan eficientes sistemas de protección de la información contenida en el ordenador, lo cual reduce el riesgo de contemplación indiscreta de un tipo de tests que, ya de por sí, implica un menor trasiego por impresoras, fotocopiadoras o similares.

• Para finalizar, un aspecto que se hace evidente en esta concepción de test es que resulta necesaria una gran cantidad de ítems de propiedades conocidas, de donde seleccionar aquel ítem más apropiado para cada sujeto en cada momento. Esto representa un importante esfuerzo y descarga una importante responsabilidad sobre la elaboración de esos conjuntos de ítems: los *bancos de ítems*.

Los bancos de ítems en el presente de la medición psicológica con tests

Este apartado ofrece el resultado de una aproximación a los bancos de ítems desde el contexto en que se enmarca su desarrollo y aplicación. Tomando como punto de referencia las líneas directrices que en la sección anterior guiaron el análisis de la evolución acaecida en el campo de los tests, ahora nos vamos a detener en revisar el papel que juegan los bancos de ítems en este entramado. Así, se pasa a analizar a continuación la relación que guarda el concepto de banco de ítems con los tres grandes aspectos que marcaban el devenir de las tres tendencias -teórica, instrumental y aplicada- antes planteadas, esto es, la TRI, los tests informatizados y los tests adaptativos. Ello permitirá aproximarnos al concepto de banco de ítems, situándolo en relación a tres elementos que sustentan una buena parte de los desarrollos que caracterizan la psicometría actual.

1. Los bancos de ítems y la TRI

Con la TRI, los bancos de ítems han pasado a protagonizar un papel relevante en la práctica psicométrica, pues varias de las dificultades con que tropezaban las pretensiones de medición basada en bancos de ítems en el contexto de la TCT, se han visto superadas en el marco de la TRI. A continuación veremos algunos de estos aspectos en que la TRI ha constituido la base teórica en que los bancos de ítems han encontrado el medio de alcanzar algunas de las características que se ha instado posean en la práctica.

• A nivel de planteamiento, la TRI se presenta como más apropiada en el trabajo con bancos de ítems, pues sus formulaciones se centran en el ítem, más que en el test, a diferencia de lo que ocurre con la TCT. Ya que un banco de ítems es básicamente un conjunto de ítems, cualquier modelo teórico que tenga en el ítem su centro de atención, su elemento nuclear, será más adecuado, en principio, en la construcción de un banco de ítems como sistema métrico.

• A un nivel más aplicado, una demanda habitualmente requerida en la construcción de bancos de ítems es que los valores de los distintos índices psicométricos de los ítems se encuentren en la misma escala. De este modo, los valores correspondientes a cualquiera de estos índices serán comparables entre sí, con la consiguiente posibilidad de tratamiento de estos datos a posteriori. El cumplimiento de este requerimiento ha venido planteando un problema relacionado con el diseño de recogida de datos implicado en la creación de un banco de ítems donde, habida cuenta del normalmente cuantioso número de ítems, se suelen administrar distintos subconjuntos de ítems a diferentes grupos de sujetos. El análisis métrico de los ítems desde la perspectiva de la TCT supone la obtención de índices dependientes de cada uno de los grupos de sujetos a los que fueron administrados y, por lo tanto, no comparables entre sí, a no ser que se pudiera garantizar la equivalencia de los grupos de sujetos, algo bastante complicado en la práctica.

Los modelos de respuesta al ítem proporcionan una solución a este problema práctico en que diferentes conjuntos de ítems son administrados a distintos grupos de sujetos. La propiedad de invarianza en la estimación de los parámetros de los ítems, así como las posibilidades de equiparación en el marco de la TRI, hacen viable el desarrollo de bancos de ítems en que los

valores de cada uno de los parámetros que caractericen a los ítems estén en la misma escala y, por tanto, sean comparables entre sí. Esta es, tal vez, la propiedad más deseable para un banco de ítems a la que da respuesta la TRI.

• Otro problema que se planteaba en el desarrollo de bancos de ítems en el marco de la TCT es el relacionado con la medida de la fiabilidad de los ítems, dada la imposibilidad de aislar el efecto de un ítem de la precisión de la medida del test del que haya formado parte -en el caso de un banco de ítems, del test en que se hubiese integrado en la calibración inicial del banco de ítems. El que la medida de la precisión de un ítem no sea función del ítem en sí, sino que dependa de la relación entre el ítem y los demás ítems que lo hayan acompañado en un test deja sin sentido la valoración de los valores de fiabilidad de los ítems a nivel de banco de ítems.

La TRI ha aportado, a través del concepto de *información* del ítem, una medida de la precisión de cada ítem independiente del resto de los ítems, propiedad que resulta valiosa en la tesitura de los bancos de ítems a fin de superar la dificultad planteada. Además, en cuanto que el concepto de información capta la diferente precisión del ítem para los diferentes niveles del rasgo medido, constituye un importante recurso como criterio de selección de los ítems en la posterior construcción de tests a partir del banco de ítems. Por otra parte, la información proporcionada por un test puede obtenerse como la suma de la información proporcionada por cada uno de los ítems que lo integren, lo cual constituye un medio directo y sencillo de obtener un indicador de la precisión de las puntuaciones estimadas por cualquier conjunto de ítems extraído del banco de ítems.

Si bien las ventajas asociadas a la aplicación de la TRI han contribuido a mejorar sustancialmente la aplicabilidad de los bancos de ítems en el desarrollo de tests, la relación entre la TRI y los bancos de ítems no debe contemplarse como simple apoyo de éstos sobre aquélla, apoyo donde la TRI sería el sustrato teórico que hace posible un concepto aplicado, los bancos de ítems. En la conceptualización de la TRI, los bancos de ítems representan una pieza clave que da pleno sentido a la citada teoría (Kingsbury y Houser, 1993). En este mismo sentido se expresan Muñiz y Hambleton (1992, p.53):

«(los bancos de ítems...) No son una parte, ni una aplicación más de la TRI, se encuentran a la base de su formulación, lo cual no siempre se ha entendido bien. Los bancos de ítems contienen los ítems parametrizados (calibrados en la misma escala) que definen operativamente la variable medida. El rasgo es el banco, los modelos no son responsables de las extrapolaciones, reificaciones y similares.».

También hay que tener presente que la subrayada propiedad de invarianza derivada de la TRI -la relativa a la invarianza de las puntuaciones de los sujetos respecto a los tests utilizados en su obtención-, es tal cuando esas puntuaciones son estimadas a partir de tests extraídos de un determinado banco de ítems; de otro modo, la afirmación un tanto genérica y habitual de que las puntuaciones de los sujetos en un determinado atributo son independientes del test utilizado para obtenerlas, puede generar alguna interpretación equívoca.

Se puede afirmar que el potencial de la TRI tiene en los bancos de ítems su medio de eclosión. De modo rotundo se expresa en esta línea van der Linden (1986), al considerar la TRI y los bancos de ítems como dos expresiones de un mismo desarrollo -el denominado *complejo moderno*-, de forma paralela a la simbiosis teórico-práctica establecida entre la TCT y los tests estandarizados en el contexto de la psicometría clásica. Este autor destaca como diversos obstáculos pueden cercenar la utilización en la práctica de la TRI si ésta se aplica al desarrollo de tests aislados (convencionales), al modo de la práctica psicométrica que ha solido acompañar a la TCT; pero que éste no sería el caso cuando los modelos de la TRI fueran aplicados en una práctica psicométrica basada en bancos de ítems. Como principal argumento sobre el que descansa este planteamiento es esgrimido el hecho de que el número de respuestas

necesario para estimar los parámetros de los modelos de la TRI es elevado, siendo precisamente los sistemas de evaluación basados en bancos de ítems, los entornos apropiados para recoger estas cantidades de datos de respuesta, pues ellos pueden representar la acumulación y aprovechamiento efectivo de esa información.

Como contrapunto al nexo estrecho entre los bancos de ítems y la TRI, también ha sido puesto de manifiesto el uso ya remoto y no vinculado a la TRI de los bancos de ítems -como es en el caso de la evaluación de conocimientos en el contexto educativo-, así como lo innecesario de situar a todos los ítems en una escala común derivada de la TRI, cuando los bancos de ítems vayan a ser utilizados para ciertos fines (Hambleton, 1986).

2. Los bancos de ítems y los tests informatizados

Tal como quedaba indicado en un apartado precedente, con el término «tests informatizados» se hace aquí referencia a la palpable y progresiva integración de recursos informáticos a que se está asistiendo en este final de siglo en las distintas facetas implicadas en la construcción y aplicación de los tests. E inmersos en esta evolución, también los bancos de ítems aparecen vinculados a esta nueva tecnología de la información, aplicando las posibilidades que ésta ofrece a diversas tareas asociadas a su creación y utilización. Es más, van der Linden (1986) afirma que el desarrollo actual de los bancos de ítems no podría tener lugar sin el soporte del ordenador, dada la ineficiencia del trabajo manual cuando se trata de bancos de ítems de tamaño grande. En este mismo sentido se han expresado otros autores en cuyas aportaciones se ha puesto de manifiesto la significativa dependencia existente entre los bancos de ítems y los ordenadores (Baker, 1974, 1984, 1986; Hambleton, 1984; Hambleton, Anderson y Murray, 1983; Hiscox, 1983; McBride y Sympson, 1982; Mitchell, 1982; Naccarato, 1988; Prosser, 1974; Vale, 1982; Wongbundhit, 1985).

El origen del vínculo entre bancos de ítems y ordenadores puede localizarse en un referente obvio si se tiene en cuenta que uno de los rasgos definitorios que suele caracterizar a un banco de ítems es lo cuantioso del número de elementos que lo constituyen. Como consecuencia de ello, la información correspondiente a cada ítem, sumada para todos los ítems, puede suponer que el volumen de información asociado a un banco de ítems sea más que considerable. Adicionalmente, en el trabajo con bancos de ítems resulta evidente la importancia de una eficiente gestión de esa información, para así poder acceder con agilidad a los datos de cualquiera de los ítems que integren el banco de ítems y facilitar el desarrollo de tests a partir del mismo.

Todos estos condicionantes habrían generado un caldo de cultivo en el que es fácil comprender que se haya recurrido a la utilización de sistemas que den soporte al almacenamiento y gestión de la información de los bancos de ítems. Éste ha sido el caso de los clásicos ficheros, o simples cajas de cartón, donde la información de cada ítem viene en una pequeña ficha de cartulina, siendo éstas ordenadas de acuerdo al criterio que se considere más conveniente (ver, p. ej., Popyuk, 1980). Pero ha sido desde la informática, de donde han llegado los recursos que han representado una respuesta más contundente a estas necesidades, hecho que encuentra rápida explicación en las amplias posibilidades de los ordenadores en el procesamiento de información.

Los sistemas de gestión computerizada de datos -los conocidos como programas de base de datos- han constituido una poderosa solución a muchas de las necesidades planteadas en el almacenamiento y gestión de la información de un banco de ítems. Este tipo de programa lo podemos considerar como pionero en la aplicación del ordenador al trabajo específico con bancos de ítems; algo comprensible teniendo en cuenta que muchas de las funciones que éstos ofrecen pueden adaptarse a las necesidades requeridas en la gestión de la información de un banco de ítems (ver, p. ej., Bowers, 1984; Gray, 1984). Es más, han supuesto un paso adelante importante, pues se han abierto posibilidades que otrora resultaban impensables en la práctica;

sea el caso, por ejemplo, de la creación de bancos de ítems en que los ítems puedan incluir secuencias de video o sonido. Adicionalmente, a los programas de base de datos de carácter general han sucedido "software" específicamente diseñado para tratar el tipo de información normalmente asociado a los ítems, con el consiguiente mayor ajuste a las demandas requeridas en este contexto y una mayor facilidad en su uso. Un buen ejemplo de este tipo de aplicación es el programa LXR-Test (Logic eXtension Resources, 1994)

Pero la relación entre bancos de ítems e informática no se limitaría a la gestión computerizada de la información de aquéllos. Al igual que las bases de datos, otros programas genéricos se han aplicado en algunas de las tareas implicadas en la construcción de los bancos de ítems, como es el caso más frecuente de los paquetes estadísticos en el análisis psicométrico de los ítems -p. ej., utilizando los difundidos SPSS, Systat y otros de este tipo. Y una vez que ya se disponga de algún banco de ítems, también en el posterior desarrollo de tests a partir de los mismos los ordenadores están jugando un papel decisivo, tanto en las tareas de selección de los ítems que mejor se ajusten a un determinado objetivo, como en las posibilidades de administrar los tests de modo adaptativo -este es el caso, por ejemplo, del programa CONTEST (Timminga y van der Linden, 1996). Y aún es de esperar el desarrollo de más programas que se adapten al conjunto de necesidades implicadas en el trabajo con bancos de ítems -en la línea de MicroCAT (Assessment Systems Corporation, 1994)-, tal como ha sido demandado por diversos autores como medio de convertir la práctica psicométrica con bancos de ítems en una realidad más común (Baker, 1986; Hambleton y Swaminathan, 1985; Roid, 1986).

Al igual que ha ocurrido a otros niveles, el vínculo de la informática con los bancos de ítems se ha reducido básicamente a una relación unidireccional en que aquélla es aprovechada a fin de hacer más eficiente el trabajo con éstos. No obstante, esta relación no se limitaría al mero aprovechamiento instrumental orientado a computerizar lo que ya se hace, pues determinados fundamentos informáticos o el conocimiento de las posibilidades de uso de los ordenadores puede promover el planteamiento o la revisión de importantes aspectos conceptuales relacionados con los propios bancos de ítems -piénsese, por ejemplo, en bancos de ítems relacionales como derivación de las estructuras de datos relacionales que se plantean en el ámbito de las bases de datos.

Lo que sí resulta evidente en esta relación entre bancos de ítems y ordenadores es la vitalidad de la misma, pues de forma continuada se ha progresado en la aplicación de la informática en pro de la mejora en el desarrollo y uso de los bancos de ítems. Y prueba de ello es la creciente variación en el tipo de programas informáticos utilizados en este campo, donde el aprovechamiento de programas de propósito general en la satisfacción de ciertas tareas -por ejemplo, las bases de datos en el almacenamiento y gestión de la información de un banco de ítems-, va dando paso al desarrollo y aparición de programas que ya de un modo más específico abordan diversas de las tareas relacionadas con el uso de los bancos de ítems.

3. Los bancos de ítems y los tests adaptativos

Bancos de ítems y tests adaptativos son conceptos que se encuentran íntimamente vinculados (Harnisch, 1985; Linacre, 1988; Mead, 1981; Rudman, 1987). Ambos son elementos sobre los que descansa una estrategia de medición que ha gozado de una importante profusión en los años recientes y en la que los bancos de ítems constituyen el sustrato del que se alimenta lo que finalmente se hace tangible a los sujetos evaluados, un test adaptado a las características de los mismos. Los tests adaptativos demandan la existencia de un cúmulo de ítems de propiedades conocidas, para así poder ajustar el contenido o la forma en que se aplique el test a las características de los sujetos que lo reciban. Por otra parte, los atributos de un banco de ítems -conocidos a partir del proceso de construcción del mismo- van a determinar, lógicamente, la calidad de los tests adaptativos que a partir del mismo se desarrollen.

Tal como se resaltaba en un apartado anterior, los TAIs constituyen en la actualidad la expresión más sobresaliente de los tests adaptativos. En éstos es aún más patente, si cabe, el papel esencial que juegan los bancos de items en la aplicación de los mismos, pues la existencia de éstos va a permitir que en cada momento, ante cada respuesta concreta, se puedan seleccionar sucesivos y apropiados items para ser administrados. El binomio bancos de items-TAIs es ampliamente reconocido en el actual panorama de la psicometría, incluso como heredero del que formaban la TCT y los tests estandarizados (van der Linden, 1986). En cualquier caso, existen algunos matices en esta relación entre bancos de items y tests adaptativos informatizados en los que creemos conveniente incidir.

La usual referencia a los bancos de items como un componente de los TAIs -el que constituye el punto de partida de los mismos- puede derivar en una concepción un tanto restrictiva de los bancos de items, pues éstos ya formaban parte del escenario de la medición con tests bastante antes de que el concepto de TAI fuera introducido. Frente a este planteamiento, la consideración de los bancos de items como componente de una estrategia de medición en la que, en función de la estrategia o algoritmo de selección de los items del banco se obtendrá como resultado un tipo u otro de test, nos ofrece un esquema más genérico en el que los bancos de items no quedan circunscritos a la aplicación de un determinado tipo de test. Distintos tipos de tests, y no únicamente TAIs, pueden ser desarrollados a partir de un banco de items, entre ellos, otras formas de tests adaptativos y también tests no adaptativos.

Ahora bien, a pesar de esta esgrimida «independencia» de bancos de items y TAIs, lo cierto es que los TAIs han representado un decisivo impulso a la aplicación de los tests adaptativos, que en su forma no informatizada resultaban poco operativos en la práctica (Bunderson, et al., 1989); y en este impulso se han visto arrastrados también los bancos de items, como elemento básico en la aplicación de aquéllos, lo cual ha reavivado la discusión sobre los distintos aspectos vinculados a su desarrollo y utilización. En definitiva, se puede decir que los bancos de items deben buena parte del renovado interés sobre los mismos al advenimiento de la evaluación individualizada computerizada. Así, acompañando a los TAIs, han aparecido desarrollos en torno a los bancos de items que han afectado a aspectos tan variados como la calibración de los items o el almacenamiento y recuperación de la información de los mismos.

Conclusiones

Un doble nivel de aproximación a los bancos de items ha sido propuesto en este trabajo, encontrándose el segundo anidado en el primero en su desarrollo:

El primer nivel de aproximación se ha centrado en ofrecer una panorámica del contexto actual de la medición psicológica con tests, vertebrándose en función de tres dimensiones: TCT vs. TRI; tests de "lápiz y papel" vs. tests informatizados; y tests convencionales vs. tests adaptativos. Este planteamiento ha pretendido aportar una perspectiva un tanto innovadora y complementaria a las diversas revisiones ya hechas sobre el tema.

Las contribuciones derivadas del marco conceptual de la TRI, así como aquéllas consecuencia de la aplicación de la informática al campo de la medición psicológica con tests, han conferido a los bancos de items un protagonismo emergente, como pieza fundamental en el engranaje de una maquinaria orientada a desarrollar tests que pretenden medir mejor y que tiene su mayor exponente aplicado en la actualidad, en los tests adaptativos informatizados.

El segundo nivel de aproximación planteado reduce su distancia focal, centrándose ya de forma específica en los bancos de items. Tomando como punto de referencia las tres dimensiones evolutivas anteriormente trazadas, el análisis de la relación que con ellas guardan los bancos de items ha mostrado el papel importante que éstos desempeñan en el contexto actual de la medición psicológica con tests. En este sentido, resulta concluyente como los bancos de items:

constituyen un concepto básico en los modelos teóricos derivados de la TRI; representan la fuente de alimentación que sustenta el desarrollo de estrategias de evaluación máximamente adaptativas; y reflejan una vía de aprovechamiento de las aportaciones de la informática al campo de la medición en psicología.

De todas formas, la perspectiva de los bancos de ítems no es tan favorable como podría parecer, entre otras cosas, porque este campo, aunque esbozado desde hace tiempo, no se ha desarrollado técnicamente lo suficiente. Existe, como se ha podido comprobar, una dependencia especialmente importante a nivel instrumental: es necesario disponer de todo un conjunto de herramientas informáticas para poder hacer efectiva la construcción y aplicación de TAIs basados en bancos de ítems.

Para el desarrollo y evolución teórico-técnico de estas herramientas se requiere especificar con perfecto detalle todos los aspectos referentes a los posibles modelos de construcción de bancos de ítems, diseñar las estrategias de mantenimiento, actualización y revisión de éstos, así como las especificaciones de su aplicación masiva e interactiva.

Los problemas no acaban aquí. De hecho, los desarrollos futuros se enfrentan a la necesidad de tener que embarcarse en proyectos de gran envergadura, de importancia social o de justificación económica, en los que se vean implicadas las instituciones públicas o grandes empresas que respalden el importante costo que tales investigaciones requieren (Muñiz, 1997). Todo ello hace pensar que, desde la óptica aplicada, la línea de desarrollo de tests convencionales desde la TCT seguirá produciendo sus frutos, aunque la investigación en el ámbito de los TAIs deberá ir progresando, con un énfasis especial en el desarrollo de los bancos de ítems y su vertiente instrumental que haga posible que tal evolución sea efectiva, aplicable, transferible y difundible.

Referencias

- Allen, M. J. y Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.
- American Psychological Association (1986). *Guidelines for computer-based tests and interpretation*. Washington: American Psychological Association.
- Assessment Systems Corporation (1994). *MICROCAT v. 3.0*. St Paul, Minnesota 55114: 2233 University Av., Suite 200.
- Baker, F. B. (1974). Systems considerations. En G. Lippey (Eds.), *Computer-assisted test construction* New Jersey: Englewood Cliffs.
- Baker, F. B. (1984). Technology and testing: State of the art and trends for the future. *Journal Of Educational Measurement*, 21(4), 399-406.
- Baker, F. B. (1986). Item banking in computer-based instructional systems Special Issue: Test item banking. *Applied Psychological Measurement*, 10(4), 405-414.
- Baker, F. B. (1989). Computer technology in test construction and processing. En R. L. Linn (Eds.), *Educational measurement* New York: Macmillan.
- Barbero, I. y Navas, M. J. (1995). *Creación de un sistema computerizado de evaluación de la capacidad matemática*. Centro de Investigación, Documentación y Evaluación (CIDE).
- Binet, A. (1909). *Les idées modernes sur les enfants*. Paris: Ernest Flammarion.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring a examinee's ability. En F. M. Lord y M. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Blinkhorn, S. F. (1997). Past imperfect, future conditional: Fifty years of test theory. *British Journal of Mathematical and Statistical Psychology*, 50, 175-185.
- Bowers, J. J. (1984). Database management for item banking and test development: An application of dBase II for the microcomputer. En *Annual meeting of the American Educational Research Association*. New Orleans, LA:
- Bunderson, C. V., Inouye, D. K. y Olsen, J. B. (1989). The four generations of computerized educational measurement. En R. L. Linn (Eds.), *Educational measurement* New York: Macmillan.
- Burke, M. J. y Normand, J. (1987). Computerized psychological testing: Overview and critique. *Professional Psychology Research and Practice*, 18(1), 42-51.
- Cohen, R. J. y Moreland, K. L. (1992). Computer-assisted psychological assessment. En R. J. Cohen, M. E. Swerdlik y D. K. Smith (Eds.), *Psychological testing and assessment* Mountain View, CA: Mayfield Publishers Co.
- Crocker, L. y Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando: Holt, Rinehart and Winston, Inc.
- Eyde, L. D. (1987). Computerised psychological testing: An introduction. *Applied Psychology: An International Review*, 36, 223-235.

- Fowler, R. D. y Butcher, J. N. (1987). International applications of computer-based testing and interpretation. *Applied Psychology: An International Review*, 36, 419-429.
- García, E., Gil, J. y Rodríguez, G. (1997). La evaluación del rendimiento a través de TACs. En *V Congreso de Metodología de las CC. Humanas y Sociales*. Sevilla: Universidad de Sevilla.
- García-Cueto, E. (1996). Software psicométrico. En J. Muñiz (Coord.), *Psicometría*. Madrid: Universitas.
- Gray, P. J. (1984). Microcomputers: Database management software. Evaluation Guide Number 7. National Institute of Education: Washington, DC.
- Green, B. F. (1983). Adaptive testing by computer. En R. B. Ekstrom (Eds.), *Measurement, technology and individuality in education: New directions for testing and measurement*. San Francisco, CA: Jossey-Bass.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Gulliksen, H. (1961). Measurement of learning and mental abilities. *Psychometrika*, 26, 109-132.
- Hambleton, R. K. (1984). Using microcomputers to develop tests. *Educational Measurement Issues and Practice*, 3, 10-14.
- Hambleton, R. K. (1986). The changing conception of measurement: A commentary (Special Issue: Test item banking). *Applied Psychological Measurement*, 10(4), 415-421.
- Hambleton, R. K., Anderson, G. E. y Murray, L. N. (1983). Applications of microcomputers to classroom testing. *New Directions for Testing and Measurement*, 19, 65-77.
- Hambleton, R. K. y Swaminathan, H. (1985). *Item Response Theory. Principles and applications*. Boston: Kluwer-Nijhoff.
- Hambleton, R. K. y van der Linden, W. J. (1982). Advances in item response theory and applications: An introduction. *Applied Psychological Measurement*, 6, 373-378.
- Hambleton, R. K., Zaal, J. N. y Pieters, J. M. P. (1991). Computerized adaptive testing: Theory, applications and standards. En R. K. Hambleton y J. N. Zaal (Eds.), *Advances in educational and psychological testing: Theory and applications*. Boston: Kluwer.
- Harnisch, D. L. (1985). Computer application issues in certification and licensure testing. En *Annual meeting of the American Educational Research Association*. Chicago, IL:
- Hiscox, M. D. (1983). A balance sheet for educational item banking. En *Annual meeting of the National Council of Measurement in Education*. Montreal, Quebec:
- Hulin, C. L., Drasgow, F. y Parsons, C. K. (1983). *Item Response Theory. Application to Psychological Measurement*. Homewood, IL: Dow Jones-Irwin.
- Kingsbury, G. G. y Houser, R. L. (1993). Assessing the utility of item response models: Computerized adaptive testing. *Educational Measurement Issues And Practice*, 21-27.
- Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceeding of the Royal Society of Edinburgh*, 61, 273-287.
- Lawley, D. N. (1944). The factorial analysis of multiple item tests. *Proceeding of the Royal Society of Edinburgh*, 62-A, 74-82.
- Linacre, J. M. (1988). Simple and effective algorithms: Computer adaptive testing. En *Annual meeting of the American Educational Research Association*. New Orleans, LA:
- Lipsey, G. (Ed.). (1974). *Computer-assisted test construction*. New Jersey: Englewood Cliffs.
- Logic eXtension Resources (1994). *LXR-TEST v. 5.1*. Rancho Cucamonga, California 91730: 9651 Business Center Drive.
- Lopez, J. A., Ato, M., Sánchez, J. y Veladrino, A. P. (1986). Test y diagnóstico psicológico por computador. En S. Algarabel y J. Sanmartín (Eds.), *Introducción a los métodos informáticos aplicados a la psicología*. Madrid: Pirámide.
- Lord, F. M. (1952). A theory of tests scores. *Psychometric Monographs*, N° 7.
- Lord, F. M. (1970). Some test theory for tailored testing. En W. H. Holtzman (Eds.), *Computer-assisted instruction, testing and guidance*. New York: Harper & Row.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: LEA.
- Lord, F. M. y Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lunz, M. E., Bergstrom, B. A. y Gershon, R. C. (1994). Computer adaptive testing. *International Journal Of Educational Research*, 21, 623-634.
- Martínez-Arias, M. R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- McBride, J. R. y Sympson, J. B. (1982). *The computerized adaptive testing system development project*. Naval Personnel Research and Development Lab., Washington, DC.
- Mead, R. J. (1981). Basic ideas in item banking. En *Annual meeting of the National Council of Measurement in Education*. Los Angeles, CA:
- Meliá, J. L. (1990). *La construcción de la Psicometría como ciencia teórica y aplicada*. Valencia: C. Serrano.
- Mellenbergh, G. J. (1996). Measurement precision in test score and item response models. *Psychological methods*, 1(3), 293-299.
- Mitchell, A. C. (1982). Using microcomputers to help teachers to develop their assessment procedures: A development project report. *Programmed Learning and Educational Technology*, 19(3), 251-256.
- Molina, J. G., Sanmartín, J. y Pareja, I. (1998). Desarrollo de un sistema informático orientado a la construcción y gestión de bancos de ítems. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados: fundamentos, estrategias y aplicaciones*. Madrid: Pirámide.
- Molina, J. G., Valero, P. M. y Sanmartín, J. (1992). Evaluación de programas informáticos orientados al análisis de tests. *Revista de psicología Universitas Tarraconensis*, 14, 31-44.
- Moreland, K. L. (1986). Computer-Assisted Psychological Assessment in 1986: A practical guide. *Computers in Human Behavior*, 4, 221-233.
- Muñiz, J. (1990). *Teoría de respuesta a los ítems: Un nuevo enfoque en la evaluación psicológica y educativa*. Madrid: Pirámide.
- Muñiz, J. (1997). Desarrollos y perspectivas en la psicometría actual. En *V Congreso de Metodología de las CC. Humanas y Sociales*. Sevilla (España).

- Naccarato, R. W. (1988). *A guide to item banking in education* (Third Edition). Washington, DC: Office of Educational Research and Improvement.
- Navas, M. J. (1994). Teoría Clásica de los Tests versus Teoría de Respuesta al Ítem. *Psicologica*, 15(2), 175-208.
- Olea, J., Ponsoda V. y Prieto, G. (1998). *Tests informatizados: fundamentos, estrategias y aplicaciones*. Madrid: Pirámide
- Ponsoda, V., Olea, J. y Revuelta, J. (1993). ADTEST: A computer-adaptive test based on the maximum information principle. *Educational and Psychological Measurement*, 54, 680-686.
- Poppyuk, W. (1980). A model for an item banking in second language. *System*, 8, 47-52.
- Prieto, G., Carro, J., González-Tablas, M., Fernández, R. y Orgaz, B. (1990). *Construcción de una batería de tests computerizados de inteligencia espacial*. Memoria final de la investigación financiada por el CIDE (MEC). Departamento de Psicología Básica, Psicobiología y Metodología, Universidad de Salamanca.
- Prosser, F. (1974). Item banking. En G. Lippey (Eds.), *Computer-assisted test construction*. New Jersey: Englewood Cliffs.
- Reckase, M. D. (1989). Adaptive testing: The evolution of a good idea. *Educational Measurement: Issues and Practice*, 8, 11-15.
- Renom, J. (1993). *Tests adaptativos computerizados. Fundamentos y aplicaciones*. Barcelona: PPU.
- Richardson, M. W. (1936). The relationship between difficulty and the differential validity of a test. *Psychometrika*, 1, 33-49.
- Roid, G. H. (1986). Computer technology in testing. En B. S. Plake y J. C. Witt (Eds.), *The future of testing*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rudman, H. C. (1987). The future of testing is now. *Educational Measurement Issues and Practice*, 6(3), 5-11.
- Sanmartín, J. y Algarabel, S. (1986). Introducción. En S. Algarabel y J. Sanmartín (Eds.), *Introducción a los métodos informáticos aplicados a la psicología*. Madrid: Pirámide.
- Santisteban, C. (1990). *Psicometría. Teoría y práctica en la construcción de tests*. Madrid: NORMA.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlations. *American Journal of Psychology*, 18, 161-169.
- Spearman, C. (1913). Correlations of sums and differences. *British Journal of Psychology*, 5, 417-436.
- Stager, S. F. y Mueller, D. (1991). Computer use in classroom testing. En K. E. Green (Eds.), *Educational testing. Issues and applications*. New York & London: Garland Publishing, Inc.
- Timinga, E. y van der Linden, W. J. (1996). *CONTEST v. 1.0* Groningen 9700AV: The Netherlands. P. O. BOX 841.
- Traub, R. B. y Lam, Y. R. (1985). Latent structure and item sampling models for testing. *Annual Review of Psychology*, 36, 119-148.
- Tucker, L. R. (1946). Maximum validity of a test with equivalent tests. *Psychometrika*, 11, 1-13.
- Vale, C. D. (1982). *Design of a microcomputer-based adaptive testing system*. Office of Naval Research, Arlington, Vancouver.
- van der Linden, W. J. (1986). The changing conception of measurement in education and psychology Special Issue: Test item banking. *Applied Psychological Measurement*, 10, 325-332.
- Wainer, H. (1990). *Computerized Adaptive Testing: A Primer*. Hillsdale, NJ: Lawrence Erlbaum.
- Weiss, D. J. (1983). *New horizons in testing: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.
- Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 53, 774-789.
- Weiss, D. J. y Vale, C. D. (1987). Adaptive testing. *Applied Psychology: An International Review*, 36(3-4), 249-262.
- Wise, S. L. y Plake, B. S. (1990). Computer-based testing in higher education. *Measurement And Evaluation In Counseling and development*, 23, 3-9.
- Wolfe, J. H. (1986). Computerized testing technology. *Advances in Reading Language Research*, 4, 71-78.
- Wongbundhit, Y. (1985). Item banking procedure and quality control in Dade county public schools. En *Annual meeting of the American Educational Research Association*. Chicago, IL:
- Wood, R. y Skurnik, L. S. (1969). *Item banking*. London: National Foundation for Educational Research in England and Wales.
- Wright, B. D. y Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.