

Tema 4: Bancos de Ítems (BI).

Licenciatura de Psicología:
*Desarrollos actuales de la medición:
Aplicaciones en evaluación psicológica.*
José Antonio Pérez Gil
Dpto. de Psicología Experimental.
Universidad de Sevilla.

Agradecimientos: a Ana Belén Cabrera Amador y Elena Camino Montiel

Tema 4

Banco de Ítems.

4.1. Introducción.

4.2. ¿Qué es un banco de ítems?

4.3. Construcción de un Banco de Ítems.

4.3.1. Definir la estructura del banco.

4.3.2. Desarrollo y análisis cualitativo de los ítems.

4.3.3. Elección del diseño de recogida de datos.

4.3.4. Administración de los ítems.

4.3.5. Análisis de los ítems y ajuste de los datos a un Modelo.

4.3.6 Calibración de los ítems.

4.3.7 Almacenamiento de la información.

4.4. Gestión Informatizada de los Bancos de Ítems. .

2.4.1. Construcción y aplicación de test a partir de los ítems del banco.

2.4.2. Mantenimiento del Banco de ítems.

4.5. Bibliografía.

4.1. INTRODUCCIÓN

Desde que en 1905 se publicara la primera escala de Binet-Simon, desarrollada a instancias del Ministerio de Educación francés para la detección de aquellos niños que manifestaran algún retraso intelectual, el campo de la medición psicológica ha avanzado de forma extraordinaria tanto en lo que se refiere a los aspectos teóricos del problema como en el campo de la aplicación práctica.

Uno de los factores clave de esta evolución ha sido el desarrollo tecnológico y la utilización del ordenador como una herramienta imprescindible no solo para la instrucción sino también para la evaluación, puesto que ha posibilitado que la administración, puntuación e interpretación de los tests sea más objetiva y automatizada. Así mismo se ha ido desarrollando el software necesario para poder utilizar nuevos modelos psicométricos, los modelos de la Teoría de Respuesta al Ítem (TRI), cuyas ideas no habían podido ser aplicadas por falta del soporte informático y tecnológico adecuado.

La introducción de la TRI en Psicometría y la utilización a gran escala del ordenador en nuestra sociedad posibilitaron una nueva práctica en el desarrollo y construcción de tests: los bancos de ítems.

4.2. ¿QUÉ ES UN BANCO DE ÍTEMS?

La concepción de lo que es un banco de ítems ha ido cambiando a lo largo de los años, aunque la idea subyacente ha sido siempre la misma: un conjunto más o menos numeroso de ítems, que miden el mismo rasgo o habilidad y que se almacenan de tal manera que, llegado el momento, se pueda elegir de entre todos ellos los que mejor se adapten a las necesidades de uso.

Los educadores, que debían preparar distintas pruebas para evaluar el rendimiento de sus alumnos, fueron los primeros en reconocer la utilidad de ir guardando los ítems utilizados y desarrollando otros nuevos, para poder construir un nuevo test cuando llegara el momento (Wise y Plake, 1990). Podemos considerar, por lo tanto, que fueron los primeros en construir bancos de ítems.

Los primeros bancos de ítems consistían en un conjunto de tarjetas, en cada una de las cuales se escribía el enunciado de un ítem junto a sus características de contenido

y psicométricas cuando las hubiera, y a la hora de construir un test, se buscaba entre todas las tarjetas hasta encontrar aquellas que contenían los ítems más adecuados. Esto, que ya supuso un avance en su momento, podía ser válido cuando se disponía de pocos ítems, pero era poco eficiente si la colección era amplia.

Los nuevos desarrollos de la Teoría de Respuesta al Ítem (TRI) contribuyeron a sentar las bases para la construcción y utilización de los bancos de ítems, aunque en un principio no tuvieron mucha repercusión a nivel práctico por la enorme cantidad de información que debía ser gestionada y para la cual no se disponía de los medios tecnológicos necesarios.

En la década de los ochenta, con la introducción masiva del ordenador en los distintos ámbitos de la sociedad y el desarrollo de la informática, se asiste a un cambio de concepción acerca de los bancos de ítems y a una nueva forma de construcción de tests a partir de dichos bancos. Este cambio de concepción se ha recogido en las distintas definiciones aparecidas a lo largo de los años. Del análisis de las mismas se desprende que un banco de ítems tiene tres características:

1. Consiste en una colección relativamente grande y estructurada de ítems que miden un rasgo o dominio de conocimiento bien definido.
2. Los ítems se encuentran almacenados en un ordenador junto con sus características tanto de contenido como psicométricas, entre las que se incluyen las estimaciones de los parámetros obtenidas utilizando el mismo modelo de respuesta.
3. Las estimaciones de los parámetros han sido transformadas a una escala común mediante un proceso de equiparación.

Cuando se va a trabajar con bancos de ítems hay que diferenciar entre dos aspectos fundamentales:

1. La construcción del banco de ítems y,
2. La gestión informatizada de los bancos de ítems.

Ambos aspectos se desarrollarán a continuación.

4.3. CONSTRUCCIÓN DE UN BANCO DE ÍTEMS

4.3.1. DEFINIR LA ESTRUCTURA DEL BANCO

En primer lugar, hay que determinar la característica o rasgo a medir y el tipo de población a la que se destina el banco de ítems.

En segundo lugar, hay que seleccionar el tipo de ítems que se van a incluir en el banco y las distintas formas en que se van a categorizar en función de las diferentes áreas de contenido, de los procesos cognitivos que implican, etc.

Por último, es necesario definir también el marco teórico que se va a utilizar en la construcción del banco de ítems (TRI o TCT), ya que del marco elegido dependerá la información psicométrica que se obtenga y que será almacenada.

4.3.2. DESARROLLO Y ANÁLISIS CUALITATIVO DE LOS ÍTEMS

Una vez definida la estructura del banco, es necesario disponer de los ítems que van a formar parte de él; para ello, se pueden seguir dos caminos:

- a) aprovechar ítems ya existentes, lo que implica la utilización de ítems desarrollados y contrastados por otros autores, con todos los problemas que ello conlleva de derechos de autor, traducción, adaptación, etc.
- b) Construir nuevos ítems que se adapten a las especificaciones hechas en la primera parte.

Sea cual sea el camino elegido, conviene hacer un análisis cualitativo de ellos, ya que algunos de los problemas que puedan presentarse, tanto respecto al contenido como a la forma de los ítems, pueden y deben ser detectados antes de su aplicación. Una forma de llevar a cabo este tipo de análisis es utilizando el *juicio de expertos*.

4.3.3. ELECCIÓN DEL DISEÑO DE RECOGIDA DE DATOS

La elección del diseño hace referencia a la forma en que se va a llevar a cabo la aplicación de los ítems en función de su número y de las muestras disponibles, de manera que se aseguren las bases para, posteriormente, poder poner todas las estimaciones de los parámetros en una métrica común y hacer comparaciones entre las distintas estimaciones obtenidas.

Los tipos de diseños más comúnmente utilizados son:

- De un solo grupo.
- Dos grupos equivalentes.
- De anclaje.

Si el número de ítems iniciales es poco numeroso, bastará con una sola muestra y, por lo tanto el diseño elegido sería de un solo grupo.

Si, por el contrario, se dispone de un gran número de ítems, será difícil que todos sean respondidos por una sola muestra, por lo tanto será necesario la utilización del diseño de grupos equivalentes o los diseños de anclaje bien de ítems o de sujetos.

De entre todos los tipos de diseños existentes, los más utilizados son los de anclaje. En los diseños de anclaje, se aplican varios conjuntos de ítems a varias muestras de sujetos de manera que o bien todos los sujetos de las distintas muestras responden a una serie de ítems comunes que sirven de anclaje (anclaje de ítems), o bien hay un grupo de sujetos que responden a todos los ítems que se aplican (anclaje de sujetos). En el primer caso, las estimaciones de los parámetros de los ítems que han servido de anclaje serán las que se utilicen para poner en la misma escala las del resto de los ítems utilizados. En el segundo caso, serán las estimaciones del parámetro de habilidad de los sujetos que sirvieron de anclaje las que permitan llevar a cabo el proceso de equiparación.

4.3.4. ADMINISTRACIÓN DE LOS ÍTEMS

La introducción de los ordenadores y el desarrollo de la informática han multiplicado las posibilidades existentes; sin embargo, teniendo en cuenta que para la construcción de un banco es necesario disponer de muchos ítems para ser administrados, y de muchos sujetos a los que administrar dichos ítems, hoy día sigue siendo menos costoso, en el amplio sentido de la palabra, la administración de los ítems de forma convencional, utilizando tests de papel y lápiz.

4.3.5. ANÁLISIS DE LOS ÍTEMS Y AJUSTE DE LOS DATOS A UN MODELO.

Una vez administrados los ítems y procesadas las respuestas de los sujetos, se dispone de los datos que van a permitir llevar a cabo un primer análisis de los mismos. Para este primer análisis se puede recurrir al cálculo de estadísticos sencillos (proporciones de aciertos a cada ítem, proporción de respuestas en cada alternativa incorrecta, etc.), que ofrecerán una información acerca del comportamiento de los ítems

que será de gran utilidad de cara a los análisis posteriores. Una vez concluido este análisis, se procede al ajuste de los datos a algún modelo.

Aunque se puede utilizar la TCT, el marco teórico más adecuado para la construcción de los bancos de ítems es el de la TRI, ya que sus modelos proporcionan estimaciones de parámetros de los ítems independientes de la muestra de sujetos utilizada en el proceso de calibración.

Siguiendo a Hambleton y Swaminathan (1985), Hambleton (1989) y Hambleton, Swaminathan y Rogers (1991), se proponen tres caminos distintos para llevar a cabo el ajuste a un modelo TRI:

1. Evaluar los supuestos del modelo. Estos son la unidimensionalidad del espacio latente (mediante procedimientos basados en el análisis factorial, entre los que figuran el test del autovalor y el test de la línea base aleatoria); la independencia local de los ítems (comprobando que las correlaciones entre ítems son nulas); y unas ciertas propiedades métricas (discriminación constante: examinando la distribución de frecuencias de las correlaciones entre las puntuaciones ítem/test, y por otro lado, inexistencia de aciertos al azar: comprobando que la probabilidad que tienen los sujetos situados en el nivel más bajo de la escala de habilidad de responder correctamente a los ítems más difíciles de la prueba es nula).
2. Analizar la independencia entre el parámetro de habilidad de los sujetos y el conjunto de ítems administrados y entre los parámetros de los ítems y la habilidad de los sujetos. Teniendo en cuenta que una de las principales contribuciones de la TRI al proceso de medición psicológica es la posibilidad de obtener mediciones independientes tanto del instrumento utilizado como de la muestra de sujetos, es necesario comprobar hasta qué punto se verifica esta propiedad. Para ello, se analiza la invarianza de los parámetros de los sujetos cuando responden a distintos conjuntos de ítems y la invarianza de los parámetros de los ítems cuando se administran a distintas muestras de sujetos.
3. Determinar el ajuste entre las puntuaciones observadas y las predicciones realizadas a partir del modelo elegido. Para ello se comparan los patrones de respuesta observados en los sujetos y los pronosticados a partir del modelo mediante el análisis de los residuales obtenidos al aplicar el modelo, la comparación entre la curva característica obtenida para cada ítem y la

pronosticada por el modelo o la comparación entre la distribución de frecuencias de las puntuaciones observadas y la de las pronosticadas.

4.3.6. CALIBRACIÓN DE LOS ÍTEMS.

Una vez aceptado que el ajuste entre el modelo elegido y los datos es adecuado, se pasará a la fase de calibración. Esta fase incluye tres tipos de tareas:

1. La estimación de los parámetros de los ítems de acuerdo con el modelo ajustado.
2. El estudio de la precisión (o fiabilidad) de las estimaciones de los parámetros.
3. La equiparación de las estimaciones de los parámetros para situarlas sobre una escala común. El método de equiparación a utilizar dependerá del diseño de recogida de datos que se ha utilizado y del marco teórico en que se desenvuelve el desarrollo de banco de ítems. Desde la TRI, sólo será necesario hacer una transformación lineal a la métrica de alguna de ellas o a una métrica elegida arbitrariamente, de manera que todas ellas tengan el mismo origen y la misma unidad de medida, que eran en lo que diferían.

4.3.7. ALMACENAMIENTO DE LA INFORMACIÓN.

Esta última fase es la creación del banco de ítems propiamente dicho; es decir, se procede a almacenar de forma estructurada toda la información disponible, de manera que pueda ser utilizada, posteriormente, en función de las necesidades del momento.

Los bancos de ítems contienen tres tipos diferentes de información acerca de los ítems. Por una parte, el contenido del ítem tal y como aparecerá en el test que se construya: su enunciado, sus gráficos si los hubiera, etc. En segundo lugar, la información que permita el acceso al ítem. Esta información es el identificador del ítem y suele consistir en una serie de palabras clave, formando una secuencia alfanumérica, que describen a cada uno de los ítems en función de su contenido, el área curricular al que pertenecen, etc. En tercer lugar, es necesario almacenar las características psicométricas de cada ítem, por ejemplo, los parámetros estimados (a , b y c) dependiendo del modelo, el tamaño de la muestra utilizado en la calibración, el número de veces que ha sido utilizado cada ítem, etc.

4.4. GESTIÓN INFORMATIZADA DE LOS BANCOS DE ÍTEMS

Dada la enorme cantidad de información almacenada, nadie puede poner en duda la imposibilidad de llevar a cabo una gestión eficaz de ella sin la ayuda de los avances informáticos; de ahí que al abordar su estudio se haga alusión a la gestión informatizada. Esta gestión puede enfocarse hacia dos objetivos fundamentales:

- a) La construcción y aplicación de tests a partir de los ítems del banco; y
- b) El mantenimiento de banco de ítems.

4.4.1. CONSTRUCCIÓN Y APLICACIÓN DE TESTS A PARTIR DE LOS ÍTEMS DEL BANCO

Una vez que se dispone de un banco de ítems debidamente calibrado, el procedimiento general para la construcción de un test es el siguiente:

- Definir cuál va a ser su objetivo (selección, predicción, orientación, diagnóstico, etc.).
- Especificar las características del test, incluyendo entre otras, el área o áreas de contenido que debe abarcar, el tipo de formato de los ítems que van a ser seleccionados, las características psicométricas de éstos y, cuando sea posible, el proceso cognitivo que requieren los ítems para su resolución.
- Introducir todas estas especificaciones en un ordenador en el que estarán implementados los algoritmos necesarios que permitan buscar, entre todos los ítems del banco, aquel que mejor se ajuste a las características especificadas.
- Cuando un ítem reúna todas las características, es automáticamente seleccionado para formar parte del test. El proceso se repetirá hasta que se haya incluido en el test el número de ítems adecuado a la estrategia de selección utilizada.
- Una vez finalizado el proceso, el constructor del test puede revisarlo, y, en caso de que no esté de acuerdo con la inclusión de alguno de los ítems, puede reemplazarlo por otros; en este sentido, aunque la selección inicial de los ítems es automática, el test final es el resultado de un proceso iterativo que incluye tanto al constructor del test como al ordenador.

ESTRATEGIAS DE SELECCIÓN DE LOS ÍTEMS

La selección de los ítems a partir de un banco puede hacerse siguiendo distintas estrategias; la elección de una u otra dependerá del tipo de test que se quiera construir y de los fines para los que vaya a ser utilizado. Hay tres tipos básicos de tests:

1. *Tests convencionales.* Se fija de antemano el número de ítems que se quiere formen parte del test. Los tests obtenidos a partir de esta estrategia no son, por lo general, muy eficientes, ya que se administrará el mismo conjunto de ítems a todos los sujetos de la muestra, con independencia de su nivel de habilidad. Para algunos sujetos puede que haber algún ítem que resulte demasiado fácil o difícil, y, por lo tanto, éstos proporcionarán poca información acerca de su nivel de habilidad.
2. *Tests adaptativos.* Se pueden seleccionar del banco aquellos ítems que son más adecuados al nivel de habilidad de los sujetos. La idea subyacente a este tipo de estrategia es que si un ítem es adecuado para el nivel de habilidad de un sujeto, ofrecerá una mayor información del mismo. La complicación práctica de este tipo de estrategias provenía de dos fuentes: la necesidad de conocer de antemano cuál es el nivel de habilidad de los sujetos para poder elegir aquellos ítems que resultarán más adecuados y por otro lado, dado que para cada sujeto de la muestra se seleccionará del banco de ítems aquellos que mejor se adecuen a su nivel de habilidad, cada sujeto responderá a un conjunto diferente de ítems y por lo tanto, el test no podía ser puntuado mediante un simple recuento de las preguntas correctas, como ocurría con los tests convencionales. Estos dos problemas han encontrado solución dentro del marco de la TRI.
3. *Tests referidos al criterio.* El objetivo consiste en averiguar en qué medida los sujetos dominan ciertos contenidos o áreas de conocimiento a los que se conoce con el nombre de dominio o criterio. Para la evaluación del grado de conocimiento que tienen los sujetos de un determinado dominio, se puede utilizar el porcentaje de ítems contestados correctamente y considerar que mientras no se supere un determinado porcentaje no han alcanzado el nivel adecuado. Al no estar este criterio en función del nivel de habilidad de los sujetos, no se trata de una estrategia adaptativa.

APLICACIÓN DE LOS TESTS

Las formas de administración de los tests son variadas:

- Tests de papel y lápiz. Si se opta por esta forma de administración, a todos los sujetos se les presentarían los mismos ítems y en el mismo orden.
- Tests administrados mediante el ordenador. Estos pueden ser:
 - Tests convencionales informatizados. Son los mismos test que se podían presentar mediante papel y lápiz pero administrados mediante el ordenador.
 - Tests adaptativos informatizados. El procedimiento de aplicación de este tipo de tests, en términos generales, para cada sujeto, es el siguiente:

En primer lugar, es necesario obtener una estimación inicial del nivel de habilidad del sujeto; para ello, se le administra algunos ítems (pocos) que difieran en dificultad y que permitan una primera estimación inicial.

Una vez obtenida la estimación inicial del nivel de habilidad del sujeto, se le presentará el ítem del banco que tenga su máxima información en el punto correspondiente al nivel de habilidad estimado.

En función de la respuesta del sujeto a este ítem, se vuelve a estimar su nivel de habilidad, y la estimación obtenida es la que va a determinar el ítem que deberá presentarse a continuación al sujeto, siguiendo el procedimiento de selección utilizado anteriormente. El proceso se repetirá hasta que el error de estimación sea más pequeño que el nivel de tolerancia especificado previamente o hasta que se haya alcanzado el número de ítems que se deseaba presentar.

4.4.2. EL MANTENIMIENTO DEL BANCO DE ÍTEMS

Una de las peculiaridades de los bancos de ítems es su carácter dinámico. Se pueden incorporar nuevos ítems al banco a medida que se van desarrollando y, a su vez, se pueden ir eliminando aquellos que se considere que son irrelevantes.

El problema que comporta la incorporación de nuevos ítems al banco es el mismo que el planteado si en el momento de su construcción el conjunto de ítems iniciales hubiera tenido que ser aplicado a distintas muestras de sujetos. En realidad, se trata de calibrar los nuevos ítems en la misma métrica que aquellos que ya forman parte del banco; para ello será necesario llevar a cabo la equiparación de las estimaciones de los parámetros obtenidas.

4.5. BIBLIOGRAFÍA

- Barbero, M.I. (1996) Bancos de ítems. En J. Muñiz, J., (Coord). **Psicometría**. Cap. 4, p.p. 139-170. Madrid: Ed. Universitas, S.A.
- Barbero, M.I. (1999) Gestión informatizada de bancos de ítems. En J. Olea, V. Ponsoda y G. Prieto, (Eds). **Tests informatizados. Fundamentos y aplicaciones**. Cap. 3, p.p. 63-74. Madrid: Ed. Pirámide.