

CONSTRUCCIÓN DE UN BANCO DE ITEMS INFORMATIZADO PARA LA EVALUACIÓN DE CONOCIMIENTOS SOBRE UNA MATERIA UNIVERSITARIA¹

Rafael Jesús Martínez Cervantes
Rafael Moreno Rodríguez²
Universidad de Sevilla

RESUMEN

Se expone el proceso de construcción de un banco de items informatizado como herramienta de aprovechamiento y mejora de la evaluación de conocimientos sobre una materia universitaria. Se tomó como material de partida las pruebas utilizadas en las convocatorias oficiales de la asignatura de *Fundamentos Metodológicos en Psicología* de la Universidad de Sevilla, desde el curso 1994/95 hasta el 2000/2001 con un total de 570 items. Se describe el diseño sistemático de dichos exámenes, así como el archivo y corrección automática de las respuestas de los alumnos. Con esa información se procedió a un análisis de las propiedades psicométricas tanto de las pruebas como de cada ítem, y al diseño de la base de datos en la que archivar esos contenidos y propiedades. Finalmente se valoran las principales ventajas e implicaciones de los bancos de items informatizados.

ABSTRACT

The construction of a computerized item-bank is described as a tool for improving the evaluation of knowledge about university subjects. The raw material consisted of the tests used in the official exams of the course entitled *Fundamentals of Methodology in Psychology* held at the University of Seville from 1994 to 2001, with a total of 570 items. Both the systematic design of such exams and the automatic scoring of students' answers are described. An analysis of the psychometric properties of every test and item was made using that information, with a database being designed to store item contents and properties. Finally, different advantages and implications of such computerized item-banks are discussed.

En el ámbito de la educación universitaria, la *evaluación* del rendimiento es parte consustancial de la enseñanza y por tanto de las actividades de su profesorado. Puesto que esta evaluación tiene repercusiones de tipo legal al determinar aspectos tales como la adquisición de títulos académicos y capacitación profesional, es imprescindible que dicha tarea se llegue a realizar con las garantías adecuadas de *igualdad, méritos y capacidad* de las personas evaluadas. Estos mismos planteamientos son aplicables también a pruebas de acceso a la universidad y de destrezas de postgraduados como en los exámenes M.I.R. o P.I.R.

Para dar respuestas a estas necesidades, desde las perspectivas educativas y de la evaluación psicológica actuales se considera que el disponer de *bancos de items* amplios y de calidad es una herramienta importante. Un banco de items no es simplemente una

colección de preguntas. Supone un conjunto de preguntas relativas a un determinado dominio de conocimientos o de destrezas, almacenadas y clasificadas en función de los componentes de ese contenido que cada una evalúe, y de sus propiedades psicométricas -tales como la dificultad y la capacidad de cada ítem para discriminar entre los examinandos con suficientes conocimientos y sin ellos (Muñiz, 1998)-. Aunque no es frecuente hacerlo, las preguntas también pueden clasificarse con arreglo a diversos aspectos de su formato -por ejemplo por tipo de pregunta abierta o cerrada, de elección dicotómica o múltiple, y por otras características como su longitud y claridad de redacción- (Martínez, Moreno y Carmona, 2001).

Las principales ventajas de un banco de ítems son las siguientes: 1) Su construcción favorece la especificación del contenido a evaluar y aspectos del formato con que se evalúa, 2) aporta además conocimiento de propiedades psicométricas de cada ítem. Una vez construido, y sobre la base de las asociaciones que contiene entre cada ítem y sus propiedades, 3) el banco permite seleccionar a aquellos de sus ítems que vayan a componer una determinada prueba en función de las propiedades psicométricas, contenido y formato que interese aplicar en cada momento, y 4) por último sirve también de guía para construir nuevos ítems que amplíen dicho banco y que posean el contenido y las propiedades psicométricas que en cada caso convengan.

En la actualidad, los bancos de ítems llegan a cobrar pleno sentido en el contexto de un sistema de evaluación *informatizado*, posibilitado sobre todo por la difusión de la infraestructura tecnológica necesaria (Barbero, 1996). La tecnología informática aporta grandes ventajas de eficiencia, rapidez y facilidad a la evaluación de conocimientos y aptitudes, tanto en la construcción y archivo de los ítems con determinadas características y propiedades como en la recuperación o muestreo selectivo de estos ítems para una prueba determinada.

En el ámbito universitario español es muy poco frecuente la construcción y uso de estos bancos de ítems informatizados y susceptibles de ser renovados de modo permanente. Es cierto que gran parte del profesorado probablemente acumula y conserva las pruebas que realiza en los sucesivos cursos académicos, y también que en general mejora con ellas su habilidad para confeccionarlas. Sin embargo, la mayoría de las veces ello se hace de manera intuitiva y no estructurada. Menos frecuente aún es la evaluación sistemática de la calidad psicométrica -por ejemplo dimensionalidad, dificultad y capacidad discriminativa- de las pruebas realizadas, cercenándose con ello la posibilidad de mejorar de modo sistemático la calidad de nuevas pruebas por construir.

A pesar de esta situación consideramos que hoy día resulta factible introducir en la práctica de la evaluación universitaria la construcción de bancos de ítems informatizados a) que permitan aprovechar de manera sistemática la información contenida en las pruebas ya realizadas de muchas asignaturas universitarias, y b) que puedan convertirse en la base para la construcción de nuevos ítems y pruebas con estimación de las propiedades psicométricas deseadas.

El objetivo del presente trabajo es mostrar la factibilidad de esas posibilidades en la actividad del profesorado universitario de nuestro entorno económico y cultural actual. En concreto, pretendemos describir el proceso de construcción de un banco de ítems informatizado a partir de pruebas ya aplicadas en cursos anteriores en una materia universitaria, como en nuestro caso son las preguntas de elección múltiple utilizadas desde

el curso 1994-95 en la asignatura *Fundamentos Metodológicos en Psicología*, perteneciente a la licenciatura de Psicología de la Universidad de Sevilla.

MÉTODO

Sujetos

Alumnos matriculados en la asignatura de *Fundamentos Metodológicos en Psicología* (FMP) y que se presentaron a las convocatorias oficiales desde el curso 1994/95 hasta el curso 2000/2001.

Instrumentos

Pruebas utilizadas en las convocatorias oficiales de FMP desde el curso 1994/95 hasta el curso 2000/2001. Se trata de 19 pruebas con 30 preguntas de alternativas múltiples cada una, que suponen un total de 570 ítems.

Infraestructura informática compuesta por diversos programas comerciales en sucesivas versiones actualizadas y distintos equipos de ordenadores personales. Para la edición de los textos y preguntas de las pruebas de examen se utilizaron como procesadores de texto los programas *WordPerfect* y *Word*. Para la corrección y análisis estadístico de las respuestas se utilizó el programa SPSS, así como el programa *MicroCAT™ Testing System* para el análisis psicométrico de las pruebas. Las capacidades de procesamiento y memoria de los ordenadores utilizados han ido evolucionando de forma paralela a los programas, desde unos primeros equipos compuestos por ordenadores personales PC 486 y HD 80 Mb, hasta ordenadores Pentium III con HD de 20 Gb.

Procedimiento

El proceso de construcción del banco de ítems ha supuesto la sucesión de las etapas que a continuación se describen.

Construcción sistemática de pruebas. Antes de considerar siquiera la elaboración de un banco de ítems, se disponía de las pruebas aplicadas en cursos anteriores. Esas pruebas habían sido construidas cada año por los profesores de la asignatura³ de una manera sistemática. Sin embargo, esa sistematización no es imprescindible para llegar a tener un banco de ítems. Puede servir igualmente toda colección de pruebas ya aplicadas y corregidas. En todo caso el procedimiento seguido por nosotros es el siguiente (ver diagrama 1). En los tres primeros meses de cada curso académico construimos simultáneamente todas las pruebas requeridas para las tres convocatorias -Febrero, Septiembre y Diciembre- existentes en nuestra asignatura, que es de primer cuatrimestre. La construcción de las pruebas comienza con la delimitación del universo de contenidos a evaluar en términos de una *Tabla de especificación* (López y otros, 1995). Sigue el muestreo de los contenidos que van a componer cada prueba, decidiéndose sobre los que se van a preguntar en cada examen. Cada año se decide por consenso de los profesores de la asignatura el procedimiento de muestreo a seguir de entre los disponibles (Moreno, Martínez y Chacón, 2000). A veces se hace un muestreo de contenidos a partir de cada subapartado, y otras se hace sólo a partir de algunos de los subapartados elegidos previamente. Esas elecciones se realizan a veces intencionalmente, tratando de que estén

incluidos los subpartados considerados más relevantes, y otras veces se hacen por procedimientos aleatorios.

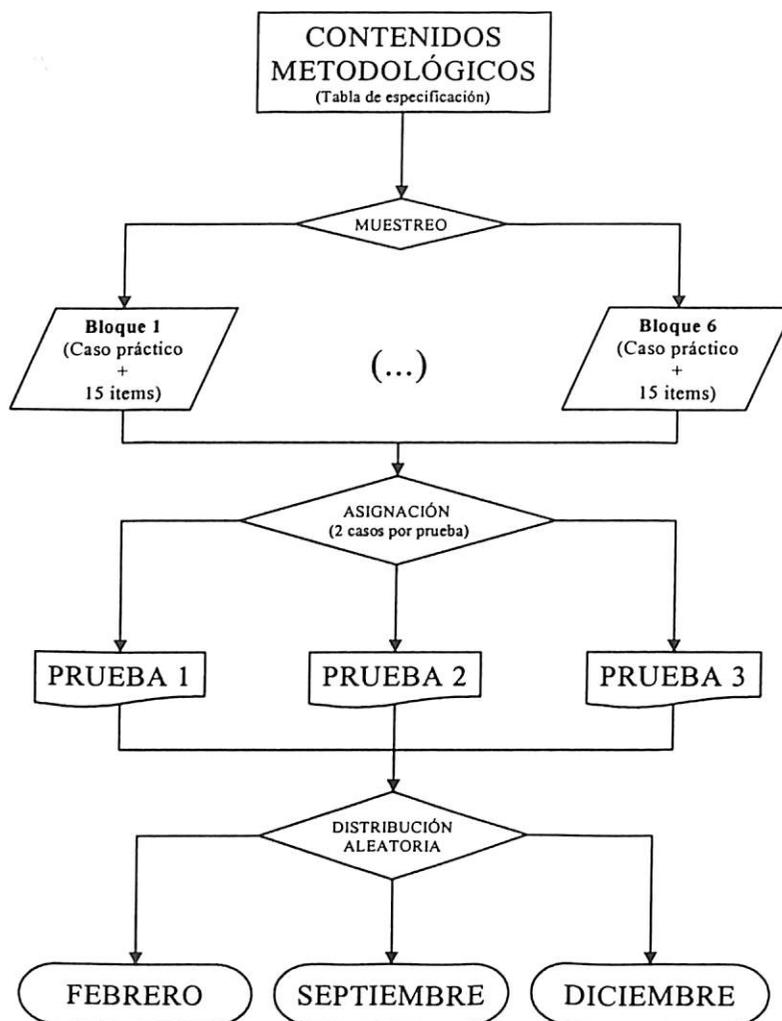


Diagrama 1.- Proceso de construcción sistemática de las pruebas de evaluación y su distribución por convocatorias.

Para esas elecciones se tiene en cuenta que el objetivo es construir seis bloques distintos, dos por prueba. Cada bloque consta de un caso práctico, consistente en la descripción de una actuación profesional o de una investigación formal psicológica en forma de resumen utilizado por la revistas de Psicología. Respecto a cada caso práctico se

construyen 15 preguntas acerca de la materia constituyente de la asignatura. Con el uso de los casos prácticos se pretende limitar el aprendizaje meramente memorístico, y fomentar la adquisición de habilidades metodológicas representativas de las que el estudiante evaluado va a necesitar en su futura práctica profesional. Más que evaluar si los examinandos saben exponer o identificar una serie de nociones de la materia, nos interesa si son capaces de aplicar ese conocimiento a situaciones reales y analizarlas metodológicamente.

Una vez definidos los contenidos a incluir en cada bloque, la construcción de éstos se realiza de la siguiente manera. Se decide al azar qué dos bloques –cada uno con un caso práctico y 15 preguntas– debe construir cada profesor. Cada caso práctico es construido bien *ex novo*, bien extraído tal cual de la literatura o bien adaptado a partir de alguno existente en ella, pero siempre con el criterio de que contenga los contenidos que han sido muestreados. A continuación se construyen los 15 ítems que deben evaluar los contenidos incluidos en cada caso práctico. Todos los ítems tienen el mismo formato de elección múltiple cerrada con seis alternativas de respuesta, cuatro específicas de cada ítem y dos más –*Todas la anteriores*, y *Ninguna de las anteriores*– que aparecen tan sólo en las instrucciones y en un recordatorio en la parte superior de cada página de la prueba. A pesar de que el uso de estas dos últimas alternativas está en discusión en la literatura de construcción de ítems, optamos por seguir usándolas dado que dicha literatura no aporta datos claros sobre ellas y que sin embargo sí estamos obteniendo a través de nuestros exámenes. En la construcción de cada pregunta tratamos de aplicar las diferentes reglas que ofrece la bibliografía (ver por ejemplo, Bartolomé, 1993, Roid y Haladyna, 1982; Haladyna, 1994; Osterlind, 1989). En general esas reglas intentan que el contenido que se desea preguntar quede bien especificado en los ítems, y que no aparezcan claves adicionales de formato que orienten indebidamente sobre la respuesta correcta al sujeto sin los conocimientos suficientes, o confundan o desorienten a los sujetos que sí los tengan (Martínez, Moreno y Carmona, 2001). Todo ese material es elaborado con un procesador informático de textos que facilita su edición y el proceso de revisión que se inicia aproximadamente un mes antes de la celebración de cada examen.

En esta última etapa, se decide al azar o intencionalmente qué dos casos prácticos y sus correspondientes preguntas van a ser emparejados para constituir cada una de las tres pruebas del curso, y aleatoriamente qué prueba va a aplicarse en cada una de las tres convocatorias. Tras esto, los profesores revisan individualmente el material construido para el examen más próximo. Una vez que identifican posibles dificultades de contenido y formato que puedan tener los casos prácticos y preguntas, ponen en común sus opiniones y deciden las modificaciones a introducir. Ese proceso se realiza hasta decidir que la prueba está lista para ser usada, lo que ocurre normalmente después de dos o tres revisiones. Tras éstas, cada prueba es confeccionada en ocho páginas, que para mejor manejo de los alumnos y economía de papel son convertidas en copistería en dos hojas grapadas e impresas por ambas caras. En la última aparece la Hoja de Respuestas con la petición de los datos identificadores de cada examinando y una tabla con treinta celdas a rellenar con los dígitos correspondientes a la alternativa elegida como correcta en cada pregunta.

Archivo informatizado y corrección automática de respuestas de los sujetos. El archivo informático de las respuestas de los examinandos se ha venido realizando hasta ahora de forma manual por parte de los distintos profesores que impartían la asignatura.

Este procedimiento consiste en la edición y grabación mediante un procesador de texto de un archivo en formato ASCII de los datos de identificación y las alternativas de respuestas elegidas por los alumnos. Este archivo queda configurado como *texto fijo* en el cual las columnas están asignadas de forma predeterminada a cada campo, desde los datos identificadores de apellidos y nombre (columnas 1 a 33) hasta las respuestas a cada ítem de la prueba (columnas 34 a 64), correspondiendo cada fila a los datos de cada alumno. En las columnas de respuestas se identifica con dígitos de 1 a 6 la alternativa elegida, y con un 9 la pregunta que no es contestada. Este tipo de formato facilita las funciones de importación de datos desde los distintos programas de análisis utilizables (en nuestro caso MicroCAT y SPSS). En el SPSS las funciones de importación se despliegan de manera interactiva desde el menú *Archivo*. En cambio en el MicroCAT la importación es directa, siendo necesario añadir a los ficheros líneas de comandos que indiquen el número de preguntas de la prueba, las preguntas no contestadas, la clave de corrección para cada pregunta, el número de alternativas y qué preguntas se van a incluir en el análisis.

Una vez importados los datos y para restar los posibles aciertos al azar de la puntuación total, elaboramos un programa de corrección automática mediante comandos de sintaxis para SPSS (Anexo I). Con este mismo programa y paquete estadístico se obtiene la calificación cualitativa y numérica de cada alumno y se facilita una serie de análisis descriptivos de las pruebas (tales como distribuciones de frecuencias, medidas de dispersión, de tendencia central, y estudios de la dimensionalidad). Además se realiza un análisis psicométrico de las pruebas para evaluar su consistencia interna, en nuestro caso a través del índice alfa de Cronbach y el de Spearman-Brown de dos mitades, y mostrando para cada ítem sus índices de dificultad (en proporciones de aciertos) y de discriminación entre sujetos hábiles y no (mediante sus correlaciones biserials ítem-test corregidas). Un diagnóstico más pormenorizado de cada ítem es realizado con el programa MicroCAT, obteniendo para cada uno un índice de discriminación basado en la diferencia de los porcentajes de aciertos en los grupos extremos de baja y alta habilidad, así como la distribución de las respuestas por alternativas en ambos grupos.

Construcción de la base de datos inicial, previa al banco de ítems. Para llegar a tener el banco de ítems es preciso confeccionar previamente una base de datos, en nuestro caso a partir de los exámenes realizados en años anteriores. A tal fin utilizamos como *software* el programa *Microsoft® Access* versión 2000 por su versatilidad y facilidad de uso, aunque cabe cualquier otro adecuado incluyendo los específicamente diseñados para bancos de ítems (como por ejemplo el programa *Bankit*). La construcción de nuestra base comenzó con la definición de dos tablas diferentes. Una referida al contenido y a diferentes características de los ítems, y otra a sus propiedades psicométricas. Aunque se podía haber confeccionado una sola tabla con toda la información, preferimos la división mencionada por permitir una mayor modularidad y distribución de tareas.

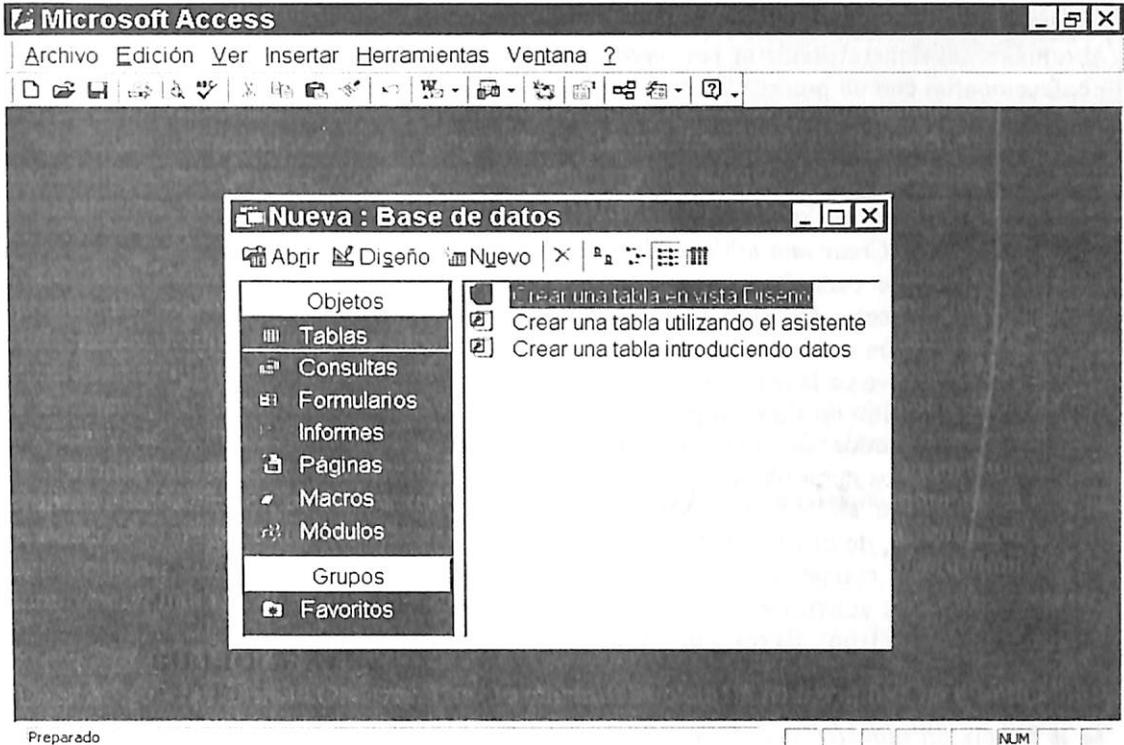


Figura 1.- Ventana con los diferentes menús para la creación de tablas en una base de datos.

La primera tabla fue definida del siguiente modo. Siguiendo el programa utilizado, debe comenzarse por crear una nueva base de datos pulsando en el icono de *Nueva base* o a través del menú *Archivo, abrir*. Tras dar nombre a dicha base, y pulsar en *Enter*, en el menú de la izquierda de la ventana que aparece debe pulsarse en la entrada *Tabla* y después en *Crear una tabla en vista diseño* (ver figura 1), aunque también puede elegirse la entrada *Crear una tabla utilizando el asistente* siguiendo en tal caso las instrucciones que éste va mostrando. En la primera forma, para cada campo a crear aparecen las columnas *Nombre del campo*, *Tipo de datos* y *descripción*. En nuestro caso creamos los siguientes: Número identificador de cada ítem (dígitos desde el 1 en adelante), Curso académico en el que se utilizó (con cuatro dígitos), Convocatoria de la prueba (Febrero, Septiembre o Diciembre), Caso práctico dentro de la prueba (1 o 2), Texto del caso práctico, Enunciado del ítem y Alternativas de respuesta. El programa pide además definir el tipo y tamaño de cada campo. En nuestro caso el identificador de cada ítem, y el de cada Caso son *numéricos*, con tamaño definido -de entre los que ofrece el programa- como *entero* para el primero y como *byte* para el segundo. Los campos de Curso, Convocatoria y Enunciado de cada pregunta son de tipo *texto*, con tamaño respectivamente de 10, 50 y 255 caracteres, mientras que el de Alternativas de respuesta tuvo que ser definido de tipo *memo* puesto que sobrepasaba los 255 caracteres que el programa admite como máximo para el tipo *texto*. Por último, el

campo para el Texto del caso práctico fue de tipo *hipervínculo* con el fin de introducir sólo el nombre del fichero donde se encuentra cada caso y no todo su texto. Estos ficheros se confeccionaron con un procesador de textos a partir de las pruebas aplicadas, denominando a cada fichero con el año, convocatoria y número de caso al que corresponde; por ejemplo el fichero 1998F1 contiene el primer caso de la convocatoria de Febrero del año académico indicado.

A continuación se elaboró la tabla de propiedades estadísticas, siguiendo en este caso la opción de *Crear una tabla utilizando el asistente*. El primer campo de esta tabla es el identificador de cada Item, con nombre y tipo idéntico al de la tabla anterior. A continuación una serie de campos recogen datos sobre las condiciones de aplicación, en concreto la Posición o número de orden del ítem en la prueba (de tipo *numérico* y tamaño de un *byte*), la Clave de la respuesta correcta (del mismo tipo y tamaño que el anterior) y el tamaño de la Muestra de alumnos presentados (de tipo *numérico* y tamaño *entero*). El resto de los campos, todos de tipo *numérico* y tamaño *doble*, recogen las características psicométricas de los ítems obtenidas mediante el programa MicroCAT. De éstos, los cuatro primeros describen al ítem en referencia a la alternativa correcta y son el índice de discriminación (*d*), de dificultad (*ID*), la proporción de aciertos en los grupos extremos de sujetos hábiles o competentes (*Pc*) e inhábiles (*Pi*), y la correlación biserial-puntual corregida entre los aciertos en el ítem y la puntuación total obtenida en la prueba por la muestra de sujetos (*rbp*). El resto de campos de la tabla repetían la información contenida en estos cinco últimos campos pero en referencia a cada una de las alternativas de respuesta. Para denominar a dichos campos se utilizó las mismas raíces anteriores, a las que se le añadía un número final de 1 a 6 para señalar a qué alternativa correspondían (por ejemplo *ID6*, *Pc3* o *rbp4*).

Para establecer el necesario vínculo entre las dos tablas previamente elaboradas se utilizó la opción *Relaciones* del menú *Herramientas*. A continuación se seleccionaron dichas tablas, que para ser vinculadas deben contener ambas un mismo campo de tipo *Clave*, con igual denominación y características y que en nuestro caso era el campo *Ítem*. El programa *Access* utilizado identifica de forma automática estos campos comunes relacionando así a ambas tablas.

Creadas y vinculadas las tablas el siguiente paso es introducir los datos de cada ítem. A tal fin el programa utilizado permite crear un formulario. Para ello debe pulsarse en *Formulario* en la misma ventana donde se hizo en *Tabla* (ver figura 1), y después en *Crear un formulario en vista diseño*. En este caso sin embargo para mayor comodidad preferimos optar por la opción *Crear un formulario utilizando el asistente*. Siguiendo las sucesivas ventanas de esta ayuda, debe ir señalándose los campos ya creados que se desean incluir en el formulario, su distribución espacial, estilo de presentación gráfica y título que se le desea dar. Una vez que se ejecuta esa definición, pueden introducirse las modificaciones que se estimen oportunas bien desde la tabla –redefiniendo por ejemplo el tamaño de un campo–, volviendo a usar el *Asistente*, o directamente en el formulario usando el ratón al modo usual de *Windows*®. De ese modo configuramos un formulario como el que aparece en la figura 2, poniendo especial atención a que cada ítem cupiera en una sola pantalla, para así facilitar una visualización más cómoda de todos sus campos.

Figura 2.- Ejemplo con datos ya introducidos del formulario *Enunciado, alternativas y propiedades.*

Disponiendo ya del formulario, la siguiente tarea consiste en pasar a él los datos de las tablas. En un principio intentamos importar los datos de modo automático a partir de archivos de datos en formato ASCII. En el caso de la primera tabla, y ante los errores aparecidos en los textos importados, optamos por hacerlo de modo más directo, usando el siguiente procedimiento. Empezamos creando los registros para los sucesivos ítems de una prueba, tecleando sus números identificadores. A continuación, se escribía para el primer ítem el Curso académico, Convocatoria, número del Caso práctico y título con hipervínculo del fichero en donde se encontraba el Texto del caso, copiándolo en el resto de ítems con las funciones *Copiar* y *Pegar* de *Windows*. Con esos mismos comandos se importaban los textos de los Enunciados y Alternativas de respuestas desde los documentos escritos con procesador de textos en los que se tenían las pruebas desde su confección. Dado lo repetitivo del procedimiento, esta tarea de importación de los 570 ítems no ocupó más de cuatro horas.

La importación de los datos a la tabla de propiedades estadísticas sí pudo ser automática, utilizando ficheros obtenidos con el programa *MicroCAT*. Estos ficheros contienen en formato ASCII los resultados del análisis psicométrico ordenados en columnas correspondientes a los mismos campos de la tabla de propiedades estadísticas anteriormente descrita, con una fila para cada uno de los ítems analizados. Posteriormente, desde el programa *Microsoft Access*, y teniendo activada la tabla a la que se van a importar los datos, se seleccionó del menú *Archivo* la opción *Obtener datos externos* y dentro de este menú a la opción *Importar*. Aparece a continuación un cuadro de diálogo para seleccionar el nombre y tipo del archivo del que se van a obtener los datos. En este cuadro, similar al de cualquier programa, es posible explorar los directorios y carpetas del ordenador para seleccionar los archivos pertinentes. En nuestro caso, y una vez seleccionado el archivo, se indicaba que era de tipo texto tras lo cual aparecía una ventana con el *Asistente para importación de texto*. Este asistente guía con comodidad el proceso de importación

solicitando una serie de datos para detectar la configuración de la información almacenada en el archivo: si está delimitado o es de anchura fija en columnas. Al ser nuestros archivos de texto fijo se debían señalar los cambios de un campo al siguiente mediante líneas con la guía de una regla en pantalla. Finalizaba el proceso al indicar la tabla a la que han de importarse los datos (pudiendo ser una tabla de nueva creación o una ya existente como fue en nuestro caso). Este proceso no presentó problemas de importación en la medida en que se había cuidado que en el archivo de texto no aparecieran problemas de formato -lo cual se puede revisar mediante un procesador de texto, cuidando de archivar los cambios en formato ASCII- y a que se diseñaron previamente los campos de la tabla en correspondencia con los datos exportados por el programa de análisis psicométrico.

Una vez creada la base de datos, incluyendo tanto sus tablas como formularios correspondientes, puede ser considerada un banco de items desde el momento en que incluya sus elementos característicos, es decir tanto descripción y categorización de los contenidos de cada ítem como indicación de sus propiedades psicométricas.

RESULTADOS

Tras el procedimiento expuesto obtuvimos un banco de items sobre la asignatura de *Fundamentos Metodológicos en Psicología* con un total de 570 items, 50 campos y un tamaño de 5,74 Mbytes. Ese banco de items acumula una amplia información sobre los exámenes realizados y cada una de sus preguntas, de la que puede ser extraída o seleccionada la que más interese en cada momento. Para ello *Microsoft®-Access* ofrece la modalidad de *Consultas* que puede obtenerse pulsando en la misma ventana de *Tabla* y *Formulario* que ya antes hemos mencionado.

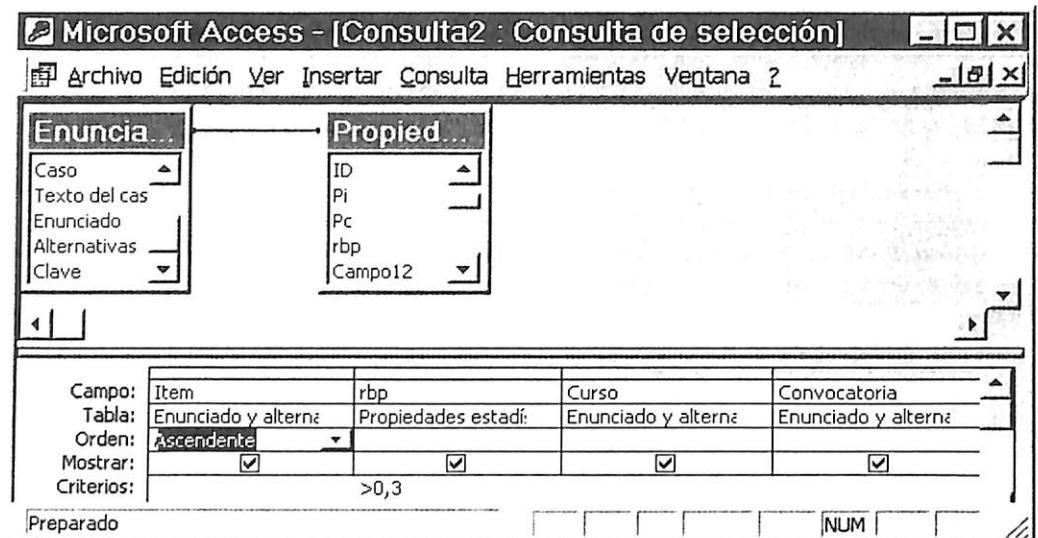


Figura 3.- Ventana de Microsoft Access con los menús para realizar consultas a una base de datos.

El cuadro de diálogo para las consultas del programa Microsoft Access se divide en dos ventanas. En la superior se deben incluir las tablas de la base de datos de las que se pretende obtener información mediante una consulta. En el caso de tablas vinculadas, como es el que nos ocupa, éstas aparecen unidas por una flecha bidireccional. Una vez insertadas las tablas deben indicarse en la ventana inferior cada uno de los campos que se quieren *mostrar* en la consulta, indicando a su vez la tabla de la que proceden los datos, si se quiere un *orden* ascendente o descendente de los registros a partir de un campo, así como cualquier *criterio de selección* de los registros. En el ejemplo mostrado en la figura 3 se muestra una selección de cuatro campos, Item, Curso y Convocatoria de la tabla de enunciados y alternativas y la correlación biserial-puntual (rbp) de la tabla de propiedades estadísticas. En ese mismo ejemplo el campo Item se utiliza además para ordenar los registros de forma ascendente y en el campo rbp se ha establecido un criterio de selección de aquellos items cuyo índice sea superior a 0,30. El resultado de esta consulta ofrece el conjunto de los items más discriminativos de los sujetos hábiles de entre todos los utilizados en las pruebas de evaluación de la asignatura.

Algunas de las informaciones que se pueden obtener de este modo a partir del banco de items son las siguientes. Como en el ejemplo recién visto, puede conocerse una determinada característica de un ítem como su dificultad o su capacidad para discriminar a los sujetos con amplio conocimiento y a los que no lo tienen. Puede interesar también obtener un listado de items atendiendo a un determinado contenido, que alcancen un determinado índice de dificultad, que discriminen mejor o peor a los examinandos con distintos niveles de conocimientos, o a que no muestren un comportamiento psicométrico adecuado, por ejemplo por presentar correlaciones negativas con la puntuación total de la prueba. En general, las consultas pueden utilizarse para listar a conveniencia items que presenten algún rasgo determinado, y a partir de su exploración tratar de abstraer inductivamente criterios generales y comunes para futuras elaboraciones.

CONCLUSIONES

El banco de items creado mediante el procedimiento sistemático descrito permite conocer en detalle los exámenes realizados, así como mejorar otros futuros e investigar sus aspectos de interés. A partir del conocimiento de las características de contenido de cada ítem y de sus propiedades métricas, el banco permite seleccionar a aquellos de sus items que vayan a componer una nueva prueba por cumplir con los requisitos que se deseen. Se está así en condiciones de hacer una elección a la carta de la prueba que se quiera construir; cuando las preguntas están referidas a casos prácticos, como en nuestros exámenes, los items del banco pueden ser aprovechados adaptándolos al contenido de cada nuevo caso que se construya. Adicionalmente, el conocimiento extraíble del banco de items posibilita la construcción de nuevos items con las características y propiedades que se deseen, ampliando así dicho banco. Para esa construcción futura puede ser útil avanzar en la categorización del contenido de los items. En la presente versión, el contenido de cada uno es especificado por las temáticas del programa de la asignatura, pero en una versión posterior esa diversidad podría ser agrupada en un número menor de categorías que facilitasen el análisis de los items existentes y la construcción de otros nuevos.

Una dificultad que conviene tener en cuenta en la elaboración de bancos de items es la necesaria equiparación de las propiedades de items y pruebas estimadas con distintas

muestras de sujetos de las sucesivas convocatorias. Este aspecto reviste especial importancia cuando se pretende reutilizar los mismos ítems en nuevas aplicaciones a muestras distintas de sujetos. De ahí la importancia de una calibración de sus propiedades precisa y relativamente independiente de las muestras utilizadas. Aunque en este trabajo no hemos referido este proceso, puesto que nuestro interés principal no era la reutilización de ítems, conviene tener en cuenta que existen distintos procedimientos técnicos para resolver la equiparación de las estimaciones (Navas, 1996).

A la vista de las posibilidades y beneficios que pueden obtenerse de los bancos de ítems para la educación superior, consideramos de interés su divulgación entre el profesorado universitario. Facilitaría analizar sistemáticamente las evaluaciones que se vienen realizando y también construir nuevas pruebas con una estimación de las propiedades psicométricas deseables. Cada banco tendría además la aplicación práctica inmediata de poder ser aprovechado por profesores de materias afines para desarrollar sus propias evaluaciones. Compartir esa información con profesorado de una misma materia permitiría establecer unos procedimientos de evaluación válidos relativamente homologables. Este intercambio no sólo facilitaría una homologación futura de la evaluación en el ámbito estatal, sino que también establecería las bases para responder con rigor a una previsible homologación que pueda ser planteada en el ámbito europeo.

Con tal perspectiva, sería más factible implicar en mayor medida a las instituciones en el proceso de evaluación universitaria, dejando de ser ésta un asunto particular de cada profesor o profesora. Una de las ayudas que podrían aportar como instituciones sería la de proveer de recursos técnicos que sobrepasan las posibilidades de una persona o incluso un departamento, como por ejemplo lectoras ópticas para el profesorado que deseara usarlas como medio de corrección de exámenes, y más aún cuando de esa corrección se derivan mayores facilidades para sistematizar el conocimiento y construcción de pruebas como en este trabajo venimos proponiendo.

Las posibilidades de los bancos de ítems se ven potenciadas además cuando por el modo en que estén contruidos facilitan el desarrollo de procedimientos que den respuesta a la necesidad de *renovar permanentemente* los elementos que lo conforman. Sucede que en gran parte de las evaluaciones ha de evitarse el uso reiterado de unos mismos ítems, dado que su posible difusión amenaza la validez de la evaluación de los conocimientos adquiridos por los sujetos. Para responder a este problema, se puede aspirar a generar sistemáticamente y de forma permanente nuevas preguntas adecuadas para la evaluación, logrando bancos de ítems virtualmente inagotables y válidos. Se puede aspirar a generar ítems incluso de manera automática mediante algoritmos definidos con las nociones propias del dominio de contenido considerado y del formato que se desee utilizar, ajustados además a las propiedades psicométricas buscadas. En este sentido existen intentos muy aprovechables como los modelos generativos de ítems (Bejar, 1993) y la incorporación tanto de principios teóricos de la temática estudiada (Prieto y Delgado, 1996; Embretson, 1998, 1999) como del formato de los ítems (Bartolomé 1993; Haladyna, 1994; Martínez, 2000; Martínez, Moreno y Carmona, en dictamen). Esta es de todos modos una línea de trabajo de mayor especialización que la que aquí presentamos como factible para gran parte del profesorado universitario actual, pero que es útil tener en perspectiva para un futuro no muy lejano.

REFERENCIAS

- BARBERO, I. (1996). Bancos de items. En **Psicometría**, J. Muñiz (coord.). Madrid: Universitas.
- BARTOLOMÉ, A. (1993). **Introducción a las pruebas objetivas**. Euroform: Material policopiado.
- BEJAR, I. I. (1993). A generative approach to psychological and educational measurement. En **Test theory for a new generation of tests**, N. Frederiksen, R. J. Mislevy e I. I. Bejar (eds.). Hillsdale, NJ: Lawrence Erlbaum.
- EMBRETSON, S. E. (1998). A cognitive design system approach to generating valid tests: application to abstract reasoning. **Psychological Methods**, 3 (3), 380-396.
- EMBRETSON, S. E. (1999). Generating items during testing: psychometric issues and models. **Psychometrika**, 64, 407-433.
- HALADYNA, T. M. (1994). **Developing and validating multiple-choice test items**. Hillsdale, NJ: Lawrence Erlbaum.
- LÓPEZ, J. y otros. (1995). **Elaboración de materiales para la docencia y evaluación universitarias: Aplicación de nociones psicológicas y metodológicas**. Memoria de investigación: ICE Universidad de Sevilla.
- MARTÍNEZ, R. J. (2000). **Diseño de items y niveles de interacción psicológica**. Tesis Doctoral no publicada: Universidad de Sevilla.
- MARTÍNEZ, R. J., MORENO, R. Y CARMONA, J. (2001, septiembre). Aproximación a la sistematización de los aspectos de formato de los items. Comunicación presentada al **VII Congreso de Metodología de las Ciencias Sociales y de la Salud**. Madrid.
- MARTÍNEZ, R. J., MORENO, R. Y CARMONA, J. (en dictamen). **Generación y corrección automáticas de items de una tarea de discriminación condicional de segundo orden**.
- MORENO, R., MARTÍNEZ, R. J. Y CHACÓN, S. (2000). **Fundamentos metodológicos en psicología y ciencias afines**. Madrid: Pirámide.
- MUÑIZ, J. (1998). **Teoría clásica de los tests**. Madrid: Pirámide.
- NAVAS, M. J. (1996). Equiparación de puntuaciones. En **Psicometría**, J. Muñiz (coord.). Madrid: Universitas.
- OSTERLIND, S. J. (1989). **Constructing test items**. Boston: Kluwer Academic Pub.
- PRIETO, G. Y DELGADO, A. R. (1996). Construcción de items. En **Psicometría**, J. Muñiz (coord.). Madrid: Universitas.
- ROID, G. H. Y HALADYNA, T. M. (1982). **Toward a technology of test-item writing**. New York: Academic Press.

ANEXO I

* Ejemplo de fichero de control para puntuar las notas en la asignatura mediante el paquete SPSSWIN.

* Puntúa como Acierto (1) o Error (0) según la clave de respuestas correctas (i1 a i30).

```
RECODE i1 (2=1) (SYSMIS=SYSMIS) (ELSE=0) INTO P1 .
RECODE i2 (4=1) (SYSMIS=SYSMIS) (ELSE=0) INTO P2 .
RECODE i3 (3=1) (SYSMIS=SYSMIS) (ELSE=0) INTO P3 .
(...)
RECODE i28 (4=1) (SYSMIS=SYSMIS) (ELSE=0) INTO P28 .
RECODE i29 (1=1) (SYSMIS=SYSMIS) (ELSE=0) INTO P29 .
RECODE i30 (1=1) (SYSMIS=SYSMIS) (ELSE=0) INTO P30 .
```

* Cuenta número de Aciertos (A), errores (E) y en blanco (B), y crea variable de CONTROL * de fallos en el conteo.

```
COUNT A = P1 TO P30 (1).
COUNT E = P1 TO P30 (0).
COUNT B = P1 TO P30 (MISSING).
COMPUTE CONTROL = (A + E + B)/30.
```

* Calcula la puntuación corregida por azar para 6 alternativas (PC), y ajusta el resultado a una escala de 0 a 10 (NUMERICA).

```
COMPUTE PC = A -(E/5).
COMPUTE NUMERICA = PC/3.
RECODE NUMERICA (LOWEST THRU 0 = 0).
```

* Asigna la NOTA cualitativa los alumnos en función de la NUMERICA obtenida.

```
STRING NOTA (A13).
DO IF (numerica < 5) .
RECODE nota (ELSE='SUSPENSO') .
END IF .
DO IF ((numerica >= 5) & (numerica < 7)) .
RECODE nota (ELSE='APROBADO') .
END IF .
DO IF ((numerica >= 7) & (numerica < 9)) .
RECODE nota (ELSE='NOTABLE') .
END IF .
DO IF ((numerica >= 9) & (numerica < 10)) .
RECODE nota (ELSE='SOBRESALIENTE') .
END IF .
DO IF (numerica = 10) .
RECODE nota (ELSE='M. HONOR') .
END IF .
```

NOTAS

¹ Este trabajo fue parcialmente financiado en la *Convocatoria de Ayudas a la Docencia para la Innovación* del ICE de la Universidad de Sevilla para el curso 2000-2001.

² Departamento de Psicología Experimental. Universidad de Sevilla. C/ Camilo José Cela s/n 41005 SEVILLA. Correo-e.: rmoreno@psicoexp.us.es; rmcervan@psicoexp.us.es

³ En el periodo indicado los profesores eran Salvador Chacón Moscoso y los autores de este trabajo