

PROPUESTA DE DESARROLLO DE TEST INFORMATIZADO ADAPTÁNDOLO A LAS RESPUESTAS DEL USUARIO

José Clares López
jclares@us.es

Universidad de Sevilla (España)

Este trabajo trata de acercarse a los test adaptativos informatizados (TAI). Se hace una propuesta para conocer de una forma inteligente al nivel que posee el usuario en el conocimiento de una determinada materia. El objetivo es hacerle el menor número de preguntas y determinar antes a su nivel de conocimientos. Para ello se parte de la opinión del usuario en lo referente al grado de conocimiento que él piensa que posee.

Palabras clave: Test Adaptativos Informatizados; Tutores inteligentes; diagnóstico; propuesta de investigación.

This work tries to approach the Computerized Adaptive Test (CAI). A proposal is made to know an intelligent form the level that has the user in the knowledge of a certain matter. The objective is to do the smaller number to him of questions and to determine before at its level of knowledge. For it part of the opinion of the user in the referring thing to the knowledge degree that he thinks that he has.

Key words: Computerized Adaptive Test; Intelligent tutors; Diagnosis; investigation proposal.

1. Introducción

La investigación básica y aplicada en el campo de la medición educativa están en un periodo bueno de desarrollo como lo demuestra las publicaciones de revistas especializadas o congresos que se celebran anualmente como se puede ver en American Education Resech Association (AERA) o la National Coouncil on Meassurement in Education (NCME) (Neira, 2001).

También hay que resaltar el gran interés que se da en la actualidad por la objetividad y el rigor en la medición educativa. Se tiene asumido que es necesaria una medición educativa rigurosa para poder conseguir una educación lo más eficientemente posible. La medición nos va a dar el resultado de las pautas

que hemos llevado a cabo en el proceso educativo.

La investigación en la tecnología psicométrica ha sido muy abundante en las dos últimas décadas, impulsándose la construcción y el análisis de instrumentos de medida en educación y en el resto de las ciencias sociales, apareciendo una serie de modelos englobadas en la llamada Teoría de Respuesta al ítem. (T.R.I).

La T.R.I nos va a aportar soluciones al problema de que para comparar a dos personas en una variable teníamos que aplicarle el mismo tests, cosa que ahora no ocurre con la T.R.I, debido a que sus modelos expresan en la misma métrica puntuaciones obtenidas por pruebas distintas. Esto es, que los modelos

del ámbito de la Teoría de Respuesta a Ítem generan mediciones invariables respecto a las pruebas que se hayan usado para conseguir estas mediciones, así la medición se ajusta al sujeto mediante la adaptación de la prueba a las peculiaridades de la persona a la que queremos evaluar. Esta característica unida al gran desarrollo tecnológico que ha permitido la generación de rápidas computadoras que han llevado a un importante avance en el campo de la medición educativa. Estamos hablando de los Tests Adaptativos Informáticos (Neira 2001), tutores inteligentes,...

2. Test adaptativos informatizados

Aplicar un test que esté al nivel del sujeto que se le pasa produce una serie de ventajas tanto para el que se examina, ya que reduce sus niveles de frustración al realizar una prueba adaptada a su nivel; como para el examinador que además de tener corregida la prueba automáticamente se reduce significativamente el tiempo de aplicación. Por otra parte, los que aumentan son los problemas, tanto en los algoritmos necesarios para el movimiento por los diferentes ítems, como por los cálculos que tiene que llevar a cabo.

A los tests referidos al criterio, se les ha dado un nuevo impulso con la Teoría de la Respuesta a los Ítems, especialmente en lo referente a la fiabilidad, mediante la función de información ya que podemos establecer un punto de corte más exacto minimizando los errores al clasificar las personas según dominen, o no, el criterio que se valora.

Se ha potenciado también con la aparición de esta teoría la construcción de los ítems llegando estos a ser unidades de medición y análisis en prejuicio de los tests.

El profesorado debe aumentar su preparación tecnológica en el campo de la evalua-

ción, lo que llevaría a una mejora de los alumnos evaluados y por consiguiente en el proceso educativo.

Se está hablando últimamente de evaluación auténtica como una evaluación alejada de situaciones estandarizadas dando así mayor realismo al proceso evaluativo y acercamiento a lo que se pretende medir (Neira 2001).

Se habla también de evaluación de rendimiento (Performance Assessment) y el portafolios. La primera hace referencia a la evaluación de ejecuciones concretas del examinado sin utilizar pruebas estandarizadas. En el portafolio el alumno presenta una carpeta con una serie de trabajos necesarios para poder demostrar que domina una determinada materia o campo.

Estos últimos tipos de evaluaciones presentan diversos problemas como por ejemplo la fiabilidad y validez, frente a las pruebas estandarizadas, además de elevado coste en tiempo y dinero que suponen.

El tema de la investigación en este campo se ha desarrollado mucho, las investigaciones son cada vez más profundas y concienzudas. Por ejemplo el estudio en el que se intenta explicar las diferencias como "error" debido al procedimiento de Abad, Olea, Real y Ponsoda (2002).

La evolución de las tecnologías y su acercamiento al usuario final ha supuesto un gran desarrollo para el campo de los tests y de los cuestionarios informáticos. En este sentido Hontangas (1999) hace una revisión bastante exhaustiva y establece una clasificación de los tests y cuestionarios informatizados en función de las prestaciones y funciones de los mismos. Aquí sólo haremos referencia a una parte de ellos, y a título ilustrativo, concretamente a los dos grupos que les otorga mayores prestaciones y al grupo de los test adaptativos informatizados.

Grupo 1: Programas con mejores prestaciones y más valorados:

1. C-Quest (Cogent Computing Corporation, 1997)
2. CAT Software System (Computer Adaptative Technologies, 1994)
3. ESTEST (Engineering Software Associates, 1998)
4. Fas-Test (Assessment Systems Corporation, 1998)
5. MicroCAT (Assessment Systems Corporation, 1995)
6. Micro Test Professional (Chariot Software Group, 1998)
7. Percepción (Question Mark Corporation, 1998)
8. Question Mark for Windows (Question Mark Corporation, 1997)
9. The Examiner (Examiner Corporation, 1998)

Grupo 2: Programas con un conjunto de prestaciones suficientes:

1. BANKIT (Molina, 1997)
2. Interactive Assesor (EQL International, 1996)
3. LXRTTest (Logic eXtension Resources, 1994)
4. Quiz Factory (Learning Ware, 1997)
5. Test Generator (Fain & Company, 1997)
6. Test Buidier (Top Grade Sofware, 1997)
7. Win WTE (WICAT, 1996)
8. WinAsk Pro (SmartLite Software, 1997)

Grupo test Adaptativos. Además de estos grupos considera otros grupos, entre los Test Adaptativos informatizados:

1. MicroCAT (Assessment Systems Corporation, 1995)
2. CAT Software System (Computer

Adaptative Thecnologies, 1994)

3. Question Mark for Windows (Question Mark Corporation, 1997)
4. Perception (Question Mark Corporation, 1998)
5. UCAT (Linacre, 1987)
6. CATsys (De la Torre y Vispoel, 1991)
7. METRIX Engine (Renom, 1992)
8. ADTEST (Ponsoda, Olea y Revuelta, 1994)

Olea, Ponsoda y Prieto (1999) presentan al final de su libro aplicaciones reales. Hacen también García, Gil y Rodríguez (1998), una propuesta de análisis de los TAI, en cuanto a sus posibilidades y limitaciones. Como por ejemplo la forma la valía de un test en el análisis de las habilidades cognitivas, su aplicación en la mejora de los procesos de enseñanza y aprendizaje

Uno de los aspectos más importantes en la construcción de un TAI, es el banco de ítems, ya que es en ellos en los que, en definitiva, se basa el diagnóstico. Renom, Doval y Sellés (1998) proponen una rutina, desarrollada por Renom y Martínez (1995), que ellos denominan de autoarranque, mediante la cual calibran los ítems del banco evitando trabajar con muestras muy elevadas de sujetos, lo que haría más costosa la construcción del banco.

Por otra parte, la forma de introducir los ítems en la base de datos puede variar, en función de la programación llevada a cabo. El programa ADETEST, por ejemplo, permite que el usuario introduzca su propio banco de ítems. Así el programa lee dos ficheros en formato ASCII conteniendo el banco de ítems y los parámetros psicométricos. Además es posible fijar la longitud del test y el tiempo de respuesta por ítem. El programa funciona en MS-DOS y es particularmente útil para realizar prácticas en clase y que los alumnos puedan crear sus propios tests GIMPSE (2002). En este

sentido es importante la conducta de elección de la dificultad de los ítems por parte de los sujetos en los TAIs, aspecto que es objeto de estudio (Hontanga, Olea y Ponsoda, 1998).

Molina, San Martín y Pareja (1998), insisten en la importancia de la construcción del ban-

co de ítems, y proponen y diferencian tres grandes bloques de tareas a la hora de implementar un sistema de este tipo. Estas tareas tienen identidad propia y reconocida y son:

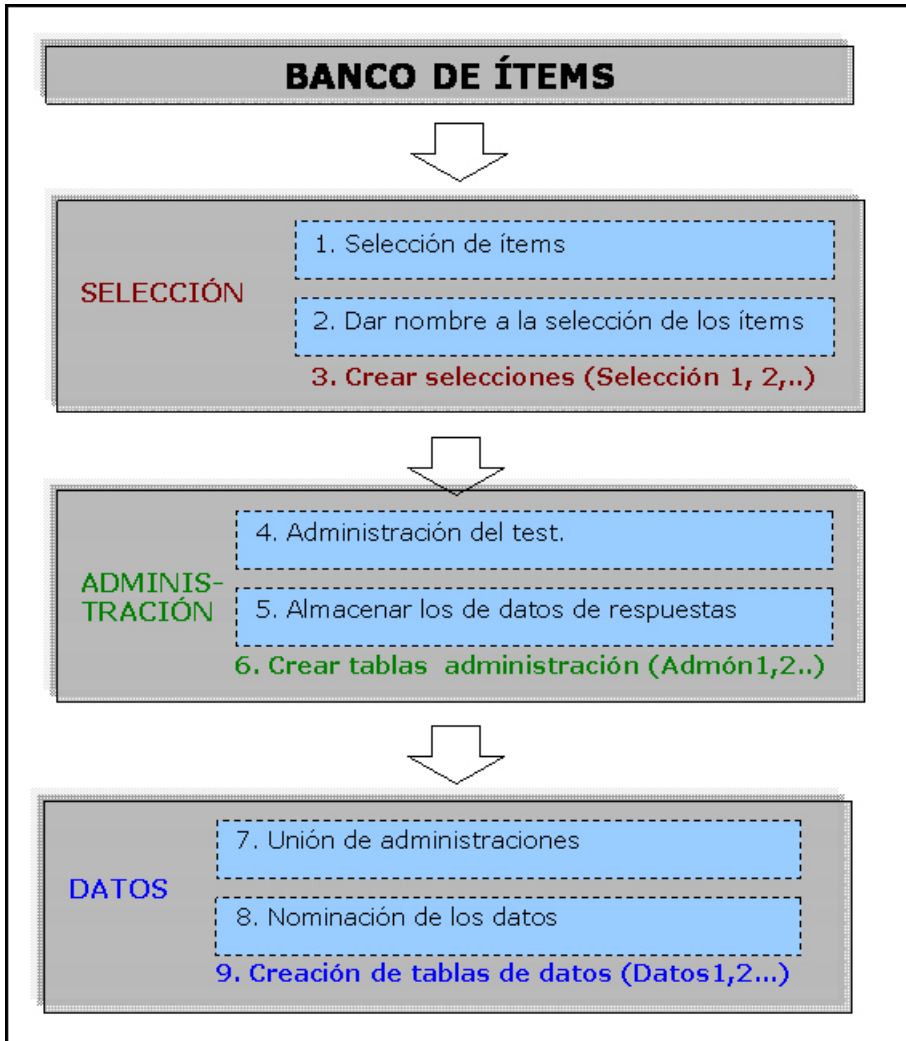


Figura 1. Selección, administración y análisis de los datos.
Adaptado de Molina, Sanmartín y Pareja. (1998).

- Almacenamiento y gestión de la información básica de los ítems.
- Recogida y gestión de las respuestas dadas.
- Construcción del propio banco de ítems.

Estos autores proponen una serie de actividades conducentes al desarrollo del citado banco, estas tareas consentirían en:

- 1) Selección de los ítems que se quieran administrar en forma de tests.
- 2) Registro y gestión de la selección realizada.
- 3) Exportación de la selección de los ítems para administrarlos mediante un programa.
- 4) Importación de los datos de las respuestas dadas a los ítems seleccionados.
- 5) Registro y gestión de la administración de los ítems realizados.
- 6) Configuración del conjunto de datos a analizar psicométricamente.
- 7) Registrar el conjunto de datos para su posterior análisis.

En la Figura 1 aparecen las diferentes fases con sus tablas correspondientes (de selección, de administración y de datos) que nos proponen Molina, San Martín y Pareja (1998). Estas fases tiene el objetivo de llevar a cabo un análisis de los ítems, con sucesivas y diferentes aplicaciones para conocer su comportamiento y características, con el fin de utilizar aquellos que sean más convenientes. La selección de los ítems para su administración se puede llevar a cabo de dos formas; o por selección ítems a ítems o mediante alguna selección automática.

La principal adaptación del esquema de Molina, Sanmartín y Pareja (1998) está la de cambiar los pasos 4 y 5, que en un principio eran tareas de exportación de la selección de

los ítems y la importación de los datos de los resultados, a administrar en el propio sistema la prueba, administrando los test (paso 4), y almacenado los resultados del mismo (paso 5).

Hay un aspecto que tiene sobre el que siempre se manifiesta una atención especial cuando se diseña un test de este tipo. Se trata de hacer una selección adecuada de los ítems de manera que se minimicen los problemas que puedan surgir en su administración. El diseño de un contenido equilibrado es una preocupación cuando se elabora una prueba adaptativa automatizada (CAT) (Leung, Chang, & Hau, 2003). Estos autores han estudiado modelos como CAT obligado (CCAT), el CAT obligado modificado (MCCAT), y el modelo multinomial modificado (MMM). En los estudios que llevaron a cabo, de los tres métodos, fue el MMM en el que aparecían algún ítems con una tasa de exposición excesiva. Anteriormente, en 2001, estos autores habían estudiaron dos métodos para controlar la sobre y subexposición de ítems, con alto y bajo valor discriminativo respectivamente. Estos métodos fueron el a-stratificado (ASTR) y el b-bloqueo (BASTR). Ponsoda y otros (2004), hacen una revisión, entre otros aspectos, de la investigación en la selección de los ítems.

Otro aspecto que también se ha tenido en cuenta es el referente al tiempo de respuesta de los ítems. En los estudios llevados a cabo se había constatado que se utilizaba más tiempo en responder a los ítems erróneos que a los que se respondían correctamente. Esta idea es rebatida por Hornke (2000) en la que en un amplio estudio no encuentra diferencias significativas de tiempo al contestar ítems correctos con incorrectos.

3. Propuesta de investigación

Cuando tratamos de aplicar o llevar a cabo la elaboración de test, desde la perspectiva del constructor del mismo, nos planteamos la forma de acercarnos a los niveles que tiene el alumno, para no tener que llevar a cabo un excesivo número de preguntas. Con carácter de propuesta de investigación nos plantearemos una forma de aproximarnos a la construcción de un instrumento que se intente adaptar al sujeto que lo va a realizar.

El primer punto del que debemos partir es la existencia de un banco de ítems lo suficientemente amplio y bien estructurado, como para evitar clásicos problemas como puede ser el conocimiento por parte del usuario de las preguntas propuestas. Olea, Ponsoda y Prieto (1999) en uno de sus capítulos de su libro hacen referencia al soporte y construcción del banco de datos para el test informatizado. Ese banco de ítems los vamos a ordenar por difi-

cultad de menor a mayor, y a distribuir de forma teórica en tres grandes grupos en función del conocimiento que se tenga de la materia.

Para acercarnos lo más rápidamente posible al nivel de conocimientos que tiene el usuario, y perder así tiempo y preguntas que pueden ser algo frustrantes para el estudiante, le vamos a pedir que nos diga el nivel que piensa que tiene con relación a la materia o contenido que estamos tratando. Para ello le daremos la opción de que seleccione un nivel de los que le proponemos.

Aunque tenemos tres grandes grupos de ítems en función de su dificultad, para ofrecer al estudiante que seleccione su nivel le vamos a proponer, además de los niveles en que hemos delimitado el nivel de los ítems, Bajo, Medio, Alto, le vamos a dar otra opción más, Novel, que representaría la ausencia de conocimiento por parte del alumno en esa materia. (Figura 2).

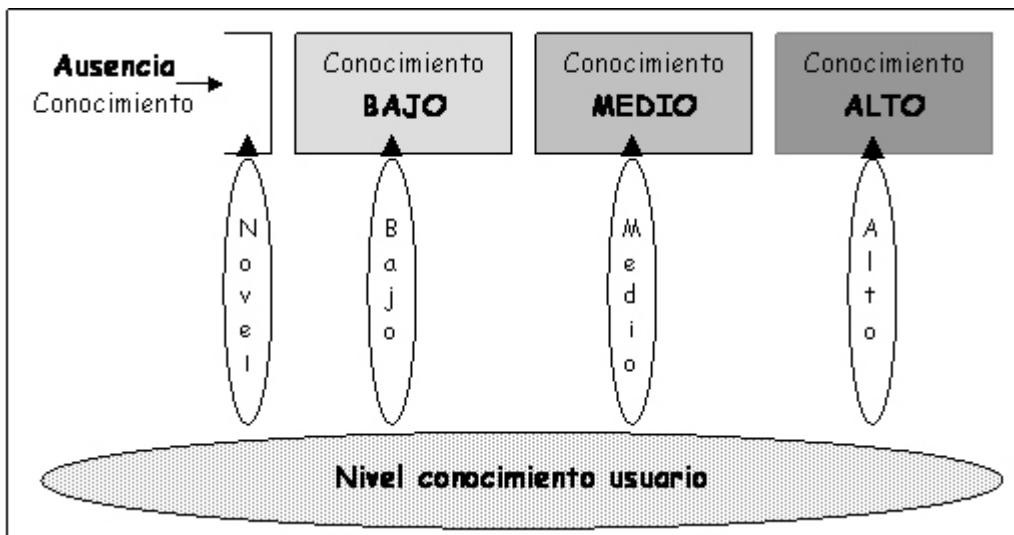


Figura 2. Niveles de elección del usuario de su propio conocimiento.

Una vez delimitada la opción por parte del usuario del nivel que piensa que tiene, se procederá según los casos de la siguiente manera:

A. Conocimiento Novel o ausencia de conocimiento. En este caso sólo hay que constatar que el usuario no conoce nada y se le

harían hasta tres preguntas del Bloque de conocimientos Bajo de las situadas entre las más fáciles.

B. Conocimiento Bajo. En este caso situaríamos el nivel de inicio de las preguntas en la parte intermedia del bloque de conocimientos bajos.

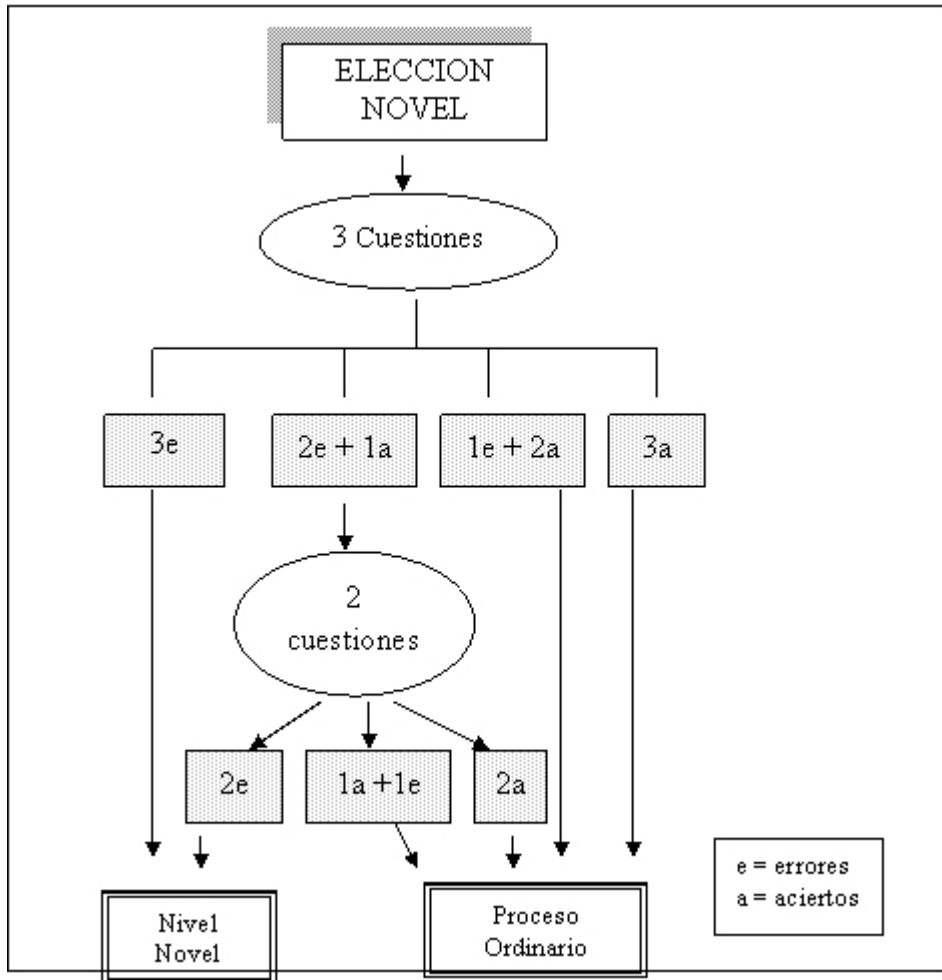


Figura 3. Proceso de evaluación en la elección de Nivel novel.

C. Conocimientos Medios. Aquí partiríamos para iniciar los ítems en la parte central del bloque de conocimientos Medios.

D. Conocimientos Altos. Igual que el anterior, pero estando situados en el bloque de conocimientos altos y en su parte intermedia.

Hasta ahora no hay mayor dificultad. El lugar de inicio va a depender de la elección que el usuario haga con respecto a sus conocimientos. Donde tenemos que iniciar la toma de decisiones es una vez realizadas las primeras cuestiones en el nivel elegido.

En el primer apartado, del Conocimiento Novel, tenemos que comprobar que no tiene conocimientos de la materia, por lo que con proponerle tres cuestiones situadas en el comienzo del nivel inicial y recibir respuestas erróneas, damos por hecho que no tiene ningún conocimiento. (Figura 3)

Si por el contrario responde afirmativamente a alguna de ellas, se harían dos preguntas más, y si el resultado de éstas es erróneo se podría pensar que el acierto no es debido al conocimiento, sino otras causas, dejando a efectos de la nueva instrucción la ausencia de este. Pero si acierta alguna de las dos o las

dos ya entraríamos en la dinámica ascendente de preguntas y respuestas hasta determinar su nivel exacto, y significaría que la idea plasmada por el usuario de que no sabe nada de la materia en cuestión no es del todo correcta, o bien sabe más de lo que en realidad cree saber.

Cuando el apartado elegido es Nivel Bajo, Medio o Alto, se procederá de la siguiente manera. Se harán tres preguntas del nivel medio del apartado que se haya seleccionado, como se indica en la Figura 5.

A modo de ejemplo lo haremos sobre el Nivel Bajo, pero es extrapolable al Nivel Medio y Alto. Empezamos planteando tres cuestiones seleccionadas de la zona intermedia del nivel elegido. Las opciones posibles que pueden salir ante la cuestión son:

- 3 aciertos
- 2 aciertos + 1 error
- 1 acierto + 2 errores
- 3 errores

Ahora el tema es la reacción del programa ante las respuestas del usuario. Como hemos comentado antes, estamos situados sobre la mitad del bloque que queremos diagnosticar

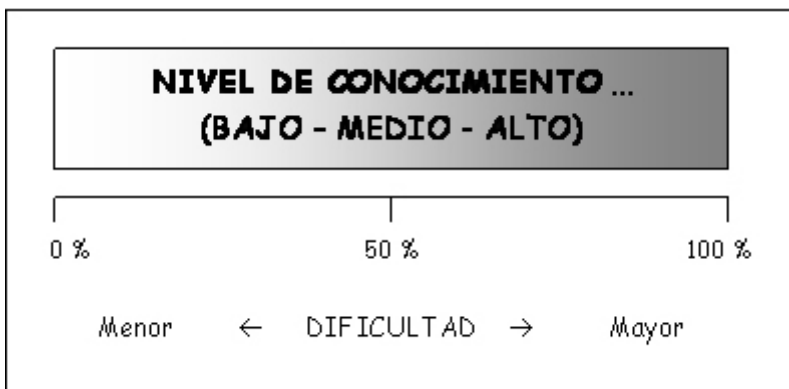


Figura 4. Distribución de la dificultad de conocimientos dentro de un nivel.

(Nivel Bajo), sobre el 50 % de las cuestiones. Consideraremos el 0 % en la parte de las cuestiones más sencillas y la parte del 100 % las más difíciles del nivel del que se trate. Figura 4

En el caso de que sean los tres aciertos, para hacer las nuevas preguntas nos situaremos sobre el 75 % del banco de preguntas en el que estamos. Es decir, aumentamos un 25 % del nivel. Si partimos de dos aciertos y un error entonces nos situaríamos un 10 % hacia la parte con mayor dificultad, es decir un 60 %. Si por el contrario lo que tenemos es un acierto y dos errores, nos situaremos un 10 % por la parte de abajo, es decir en el 40 %. Y en la última opción, si tenemos tres errores nos situaremos un 25 % hacia la parte con menor dificultad. Es decir, nos situaríamos en el 25 % del bloque.

Los resultados anteriores podríamos esquematizarlos de la siguiente manera.

Resultados	Situación inicial	Desplazamiento porcentual	Situación tras los resultados
3 aciertos	50 %	+ 25 %	75 %
2 aciertos + 1 error	50 %	+ 10 %	60 %
1 acierto + 2 errores	50 %	- 10 %	40 %
3 errores	50 %	- 25 %	25 %

Así, después de responder a las tres primeras cuestiones, y situarse en el nivel porcentual que le correspondiese, tendríamos que iniciar de nuevo el ciclo para la siguiente fase, y seleccionar otras tres cuestiones del nivel alcanzado, y aplicar el mismo desplazamiento porcentual que para la fase inicial.

Se nos puede plantear el problema práctico que a la hora de que el programa seleccione las preguntas moviéndose por los porcentajes, puedan aparecer preguntas repetidas al aumentar un 25, 50,.. % o disminuir. También

se puede dar el caso de que otras preguntas no aparezcan nunca. Es lo que Revuelta, Ponsoda y Olea (1998) llaman del "control de tasa de exposición" de los ítems, refiriéndose a la frecuencia con que son utilizados cuando se administra un test. Un ítem con una tasa alta de exposición corre el riesgo de que sea demasiado conocido y pierda validez. Estos autores presentan tres métodos para controlar la tasa de exposición de un ítem en los TAI (Test Adaptativos Informatizados) que serían: añadiendo un componente aleatorio (al método de selección del ítem de máxima información); controlar la tasa de frecuencia de cada ítem y el último control sería mezclar los dos anteriores.

También pensamos que podría ser interesante, completando la propuesta anterior, para prevenir esta problemática generar de forma aleatoria un número extraído, por ejemplo, de una consulta a la hora del sistema y sumando o restando (en función de si el número acaba

en par o impar) otro número aleatorio de entre 0 y 6 a ese porcentaje asignado. De esta manera habría muchas menos posibilidades de que esa circunstancia tuviese lugar.

Con estas precauciones se evitaría, además de la repetición de un mismo ítem, que dos alumnos contiguos, con un nivel conocimientos similar, estuviesen recibiendo la misma pregunta a la vez, lo que sería poco recomendable si estuviesen los alumnos en un aula de informática.

El esquema propuesto de movimiento por-

centual de dificultad se mantiene hasta que no se produzcan accesos de 25 % tanto hacia zona de mayor dificultad, como de menor. Esto es, que no haya tres aciertos o tres errores seguidos. En principio podemos tomar como tres los intentos para determinar la zona de conocimientos en la que se encuentra, aunque será la experimentación la que delimitará el número de ítems necesarios para establecer el nivel de conocimiento.

Con el desarrollo de las fases puede ocurrir que nos situemos fuera del nivel del que partimos, bien por la parte de abajo o por la superior. Por la parte de abajo llegaría al 0 %, y por la de arriba al 100 %, cuando esto ocurra sólo habría que entrar en el nivel correspondiente y seguir efectuando los diferentes posicionamientos en función de las respuestas del usuario.

El último apartado que nos ocupa es cono-

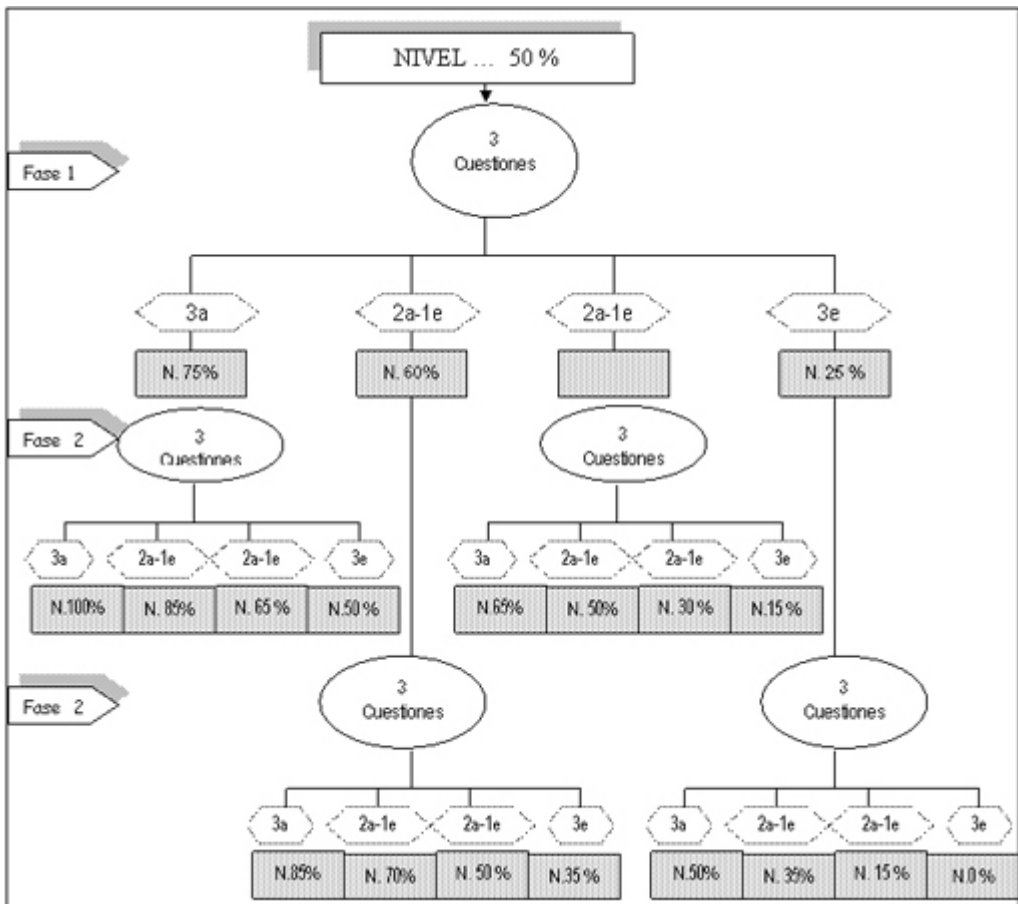


Figura 5. Detalle de las fases en la elección del nivel de conocimientos en función de las respuestas.

cer el nivel concreto de conocimientos que posee el estudiante. Para su establecimiento vamos a hacer algunas propuestas para la experimentación. Las propuestas están centradas en varias formas de acceder a la puntuación intermedia, tomando las fases en las que se sitúan en las últimas etapas del test. Podríamos tomar para el cálculo algunas de las siguientes formas:

1ª. Tomando las últimas cinco fases de preguntas, es decir los 15 ítems, se determina el menor y el mayor y se calcula el centro. Este será el nivel de conocimientos.

2ª. Situándose en las fases en las que no se dan aciertos o errores totales, es decir que no se aciertan ni se fallan las tres, sino que se acierta uno o dos, o se aciertan una o dos, se obtiene el ítem situado en una puntuación intermedia.

Esta es una propuesta para la investigación sobre la forma de selección, por parte del programa informático, de los ítems de una manera inteligente en orden a reducir el tiempo y el número de ítems, entre otros elementos. Propuesta que deberá adaptarse y completarse con los resultados de un proceso de investigación.

4. Bibliografía

ABAD, F. J., y otros. (2002). Estimación de habilidad y precisión en tests adaptativos informatizados y tests óptimos: un caso práctico. **Revista Electrónica de Metodología Aplicada**, 7, 1-20.

GARCÍA, E.; GIL, J. y RODRÍGUEZ, G (1998). La evaluación de tests adaptativos informatizados. **RELIEVE**, vol. 4, n. 2.

GIMPSE (2002) Software de libre acceso.

Grupo de Investigación en Medición Psicológica y Educativa. <http://www.uv.es/~hontanga/demos.html>

HORNKE, L. F. (2000). Item Response Times in Computerized Adaptive Testing. **Psicológica** 21, 175-189.

HONTAGAS, P. (1999) . Software para la construcción y administración de tests informatizados. En OLEA J.; PONSODA V. y PRIETO G. (Eds.), **Tests informatizados. Fundamentos y aplicaciones.** 251-286. Madrid: Pirámide.

HONTAGAS, P; OLEA, J. y PONSODA, V. (1998). Elección de la dificultad de los tests autoadaptados informatizados: un estudio piloto. **RELIEVE**, vol. 4, n. 2. http://www.uv.es/RELIEVE/v4n2/RELIEVEv4n2_3.htm.

LEUNG, C. K.; CHANG, H. H. y HAU, K. T. (2003). Computerized adaptive testing: A comparison of three content balancing methods. **Journal of Technology, Learning, and Assessment**, 2 (5). <http://www.jtla.org>

LEUNG, C. K.; CHANG, H. H. y HAU, K. T. (2001). Integrating Stratification and Information Approaches for Multiple Constrained CAT. **NCME Annual Meeting, April.** 11-13.

MOLINA, J. G.; SANMARTÍN, J. y PAREJA, I. (1998). Recogida y gestión de datos de repuesta en bancos de ítems: análisis de un sistema informático. **Relieve. Vol 4. n° 2.**

NEIRA, T. (2001). **La evaluación en el aula.** 2ª Edic. Oviedo: Nobel.

OLEA, J.; PONSODA, V. y PRIETO, G (Eds.) (1999) **Test Informatizados. Fundamentos y aplicaciones.** Madrid: Pirámide.

PONSODA, V. y otros (2004) Los tests adaptativos informatizados: investigación actual. **Metodología de las Ciencias del Comportamiento, Suplemento 2004**, 505-510

RENOM, J., DOVAL, E. Y SELLÉS, M. (1998).

Optimización de los TAI mediante el procedimiento de autoarranque. **Relieve**, vol. 4, n. 2. http://www.uv.es/RELIEVE/v4n2/RELIEVEv4n2_2.htm

RENOM, J. Y MARTÍNEZ, M. (1995) Construcción de bancos de ítems sin anclajes ni equiparaciones: el procedimiento de autoarranque. En M. ATO Y LÓPEZ-PINA, J.A. **IV Symposium de metodología de las Ciencias del Comportamiento**. Universidad de Murcia.

REVUELTA, J.; PONSODA, V. y OLEA, J. (1998). Métodos para el control de las tasas de exposición en tests adaptativos informatizados. **RELIEVE**, vol. 4, n. 2. http://www.uv.es/RELIEVE/v4n2/RELIEVEv4n2_4.htm