

SISTEMA INTEGRADO DE PROCESADO DE TEST

JOSÉ A. DÍAZ, ÁNGEL NEIRA, ALFREDO ALGUERO, JOSÉ A. L. BRUGOS,
VICTOR GARCÍA, LUIS ÁLVAREZ Y ENRIQUE SOLER*

Construir bancos de preguntas de forma rigurosa para comprobar la adquisición de conocimientos es una tarea difícil y compleja. En este trabajo, inicialmente, se analizan una serie de consideraciones metodológicas que el profesor debería tener en cuenta en su elaboración y a continuación se presentan varios mecanismos estándar para la organización y selección de ítems por computador. Con todo ello, se plantea un sistema informático para el procesado de test que integra las siguientes aplicaciones: a) un editor de ítems configurable para permitir la aplicación en libertad de las consideraciones metodológicas, b) un módulo para su catalogación y construcción de bancos de ítems, c) diversos sistemas para la generación de pruebas de acuerdo con los mecanismos anteriores, y d) una aplicación final para la realización de dichas pruebas por parte de los alumnos de forma dinámica e interactiva.

The elaboration of questions banks to check acquired learning rigorously is a difficult and complex task. In this paper, initially, they are analyzed various methodological considerations that the teacher should consider in their elaboration, and next, several standard mechanisms are introduced for items organization and selection by computer. With this, a computer system for testing is presented that integrates the following applications: a) a configurable item editor to allow the free use of the methodological considerations, b) a module for item banks construction and item classification, c) diverse systems for test exams generation by using the previous mechanisms, and d) a final dynamic and interactive application for testing the students.

1. Introducción

Osterlind define los Instrumentos de Evaluación como aquellos mecanismos de medida que proponen al alumno un estímulo y una forma prescriptiva de emitir una respuesta a partir de la cual se puede deducir la consecución de un objetivo de aprendizaje [1].

En este sentido, el recurso más tradicional o innovador y el indicio más pequeño que ayuden a un docente a comprender mejor a un alumno, o que posibiliten a un alumno a comprenderse mejor a sí mismo, deben considerarse como válidos, siempre que puedan estimarse en alguna medida

* JOSÉ A. DÍAZ, ÁNGEL NEIRA, ALFREDO ALGUERO, JOSÉ A. L. BRUGOS, VICTOR GARCÍA, LUIS ÁLVAREZ Y ENRIQUE SOLER son Profesores de la Universidad de Oviedo.

[2]. Conviene por tanto, esforzarse en obtener todos los indicios a través de todos los medios; los diversos modos de actuación de los alumnos servirán como indicadores del grado de consecución de los objetivos [3].

En concreto los Instrumentos de Evaluación pretenden provocar una respuesta definida del alumno que incluye procesos mentales sobre contenidos científicos. Se trata de diseñar situaciones para comprobar la consecución de objetivos de aprendizaje; construir preguntas, de manera precisa y breve, para poder deducir de las respuestas de los alumnos, conclusiones válidas y fiables sobre la adquisición de objetivos referidos a diferentes capacidades.

Ésta es una tarea difícil, entre otras razones porque con frecuencia los ítems se reducen a una simple frase que ofrece pocas oportunidades para describir un contexto que aclare su significado. Tampoco es fácil establecer la indispensable relación entre pregunta y objetivo que reduzca las fuentes de error y aumente la validez y fiabilidad de los resultados.

No es posible, ni se pretende con este trabajo, sustituir ni mermar, con un sistema informático, al profesor, en su labor creativa ni su experiencia en la elaboración y uso de las Pruebas Objetivas, tampoco se pretende, ni es competencia de estas líneas defender ni justificar el uso de dichas pruebas, es numerosa la bibliografía y los Centros especializados encargados de ello [3,4]. Sencillamente se constata que la mayor parte de la evaluación educativa actual se lleva a cabo dentro de lo que se denomina Teoría de la Respuesta a los Ítems. Este enfoque implica, entre otras cosas, la elaboración y uso apropiado de bancos de ítems organizados por materias y objetivos de conocimiento [5,6]. Estos bancos permiten la realización de pruebas adaptadas al alumno que se examina. Como afirma José Muñiz, “la época de un mismo examen para todos los sujetos ha terminado” [3].

Un breve repaso de los criterios metodológicos para una adecuada construcción de pruebas objetivas y una somera visión de lo que constituyen los fundamentos y estado del arte en el campo de la Teoría de la Respuesta a los Ítems en su aspecto computacional nos darán pie para presentar una aplicación informática para la edición y catalogación de ítems y la creación y realización de pruebas objetivas en vistas a su estudio para el planteamiento de nuevos criterios adaptativos.

Para ello, desde estas líneas, nos comprometemos con los lectores a la distribución de esta herramienta con el ánimo de conocer y evaluar los resultados de su aplicación en diferentes campos de trabajo y materias de conocimiento.

2. Pruebas objetivas. Criterios metodológicos de elaboración

Las pruebas objetivas (PO) tienen como denominador común que el alumno, previa elaboración mental acompañada a veces de una construcción material, elige la respuesta entre una serie de alternativas que se le proporcionan.

Profesores con experiencia han demostrado repetida y continuamente que un experto puede transformar en PO casi todos los exámenes que se ponen en nuestras aulas; y que las PO, lejos de disminuir la capacidad discriminativa de los ítems y la fiabilidad de las calificaciones, las aumentan.

La relación entre el proceso de construcción de un test y su validez es compleja. A continuación, se enumerarán algunas normas para construir ítems como un marco para facilitar la construcción de ítems de calidad que proporcionen interpretaciones válidas [3].

El proceso de construcción de cualquier formato de ítem, que debe proponerse de tal manera que un profesor preparado pueda seguirlo y repetirlo, debe empezar por la explicación de lo que se va a comprobar, es decir, la variable que se quiere medir. Como en un contexto académico se hará una valoración criterial, esta variable es un objetivo de aprendizaje explicado a través de unos indicadores observables. Aunque es ésta una condición obvia, es una de las más difíciles de cumplir, por lo que es frecuente que los profesores la ignoren o afirmen que los objetivos los tienen implícitamente en su mente; pero es imprescindible también explicitarlos antes de establecer las preguntas. Posteriormente, supuesto el objetivo, a la hora de formular las preguntas, en la práctica suele ser más fácil usar los materiales de enseñanza (libros, apuntes, etc.) como fuente de ideas que deducirlas directamente de los objetivos de aprendizaje.

Conseguir el máximo de relación entre el ítem y el objetivo evaluado, relación que influye directamente en la validez y fiabilidad de los resultados obtenidos, no es fácil. Determinarla requiere, además de dominar la materia y la expresión, un gran conocimiento de los aspectos psicológicos de la capacidad medida.

Las PO presentan diversos formatos, como son el Verdadero/Falso (V/F), V/F múltiples, Respuesta doble, Respuesta limitada, etc. Las que se van a utilizar en esta aplicación son fundamentalmente las PO de múltiple elección de respuestas.

2.1. Múltiple elección de respuestas

Los ítems de Múltiple Elección de Respuesta (MER) son pruebas en las que el alumno debe sólo seleccionar una de entre un conjunto de posibilidades limitas.

En su forma más completa constan de unas Instrucciones; un Dibujo, Gráfico o Texto con la información necesaria; una base, que presenta un problema, con una frase incompleta o preferentemente una pregunta; una serie de alternativas, es decir, Distractores o respuestas plausibles incorrectas, y la Respuesta correcta. Las alternativas pueden ser frases completas, fragmentos de oraciones o incluso palabras aisladas o números.

Con frecuencia no son necesarios tantos detalles, por lo que el ítem se presenta en forma abreviada, con la Base y las alternativas (a veces también se les llama distractores, aunque es mejor reservar este término para las respuestas incorrectas).

BASE: De las actividades que realiza el profesor, ¿cuál pertenece a la evaluación?

DISTRACTORES: a) Seleccionar contenidos.
b) Formular objetivos.
c) Elegir metodología.

REPUESTA: d) Recoger datos del proceso educativo.

Existe una serie de variedades, entre las que citamos las siguientes:

2.1.1. Escoger la única respuesta correcta

Es la variedad más popular y aconsejable. Ejemplo:

¿Qué es la neviza?

- a) Polvo de nieve caído en el mismo día.
- b) Una variante del hielo cristalizado.
- c) Nieve comprimida que se hace hielo.
- d) La abrasión del hielo cristalizado.

Entre muchas de las preguntas que se pueden hacer es preferible formular no una única respuesta correcta (dado que a veces es difícil, dentro de los límites razonables de un ítem MER, formular esa única respuesta absolutamente válida), sino la mejor respuesta, la más correcta o la más completa en la gradación de respuestas válidas que se presenta. Así pues, las alternativas pueden contener alguna corrección o relevancia, aunque una debe ser más completa o más relevante. Las menos correctas o menos completas, que aisladamente se podrían considerar como verdad, en relación con las otras alternativas en un formato MER se tratan como incorrectas. Aunque en un ítem MER de única respuesta correcta se sobreentiende que es la *mejor* entre las propuestas, este extremo es mejor explicitarlo; por ejemplo:

¿Qué afirmación caracteriza mejor la poesía de San Juan de la Cruz?

- a) Es una poesía escrita en liras.*
- b) Su poesía utiliza sabiamente las aliteraciones y otros recursos sonoros.*
- c) Es una poesía amorosa, profunda y mística.*
- d) Su poesía es un diálogo incontaminado con este mundo.*

2.1.2. Eliminar la opción incorrecta

Las negaciones resultan especialmente atractivas a los examinadores porque son fáciles de formular añadiendo un *no* a los materiales de instrucción. Sin embargo, las preguntas formuladas negativamente son confusas, ya que requieren que el alumno elija una respuesta que no es verdad o no es característica, por lo que se aconseja evitarlas, en la medida de lo posible.

Ejemplo:

Se dice que un cuerpo NO está cargado cuando tiene:

- a) exceso de electrones.*
- b) defecto de electrones.*
- c) un determinado potencial.*
- d) igual número de protones que de electrones*

2.2. Proceso de construcción de los ítems

El proceso de construcción de los ítems se puede resumir en los siguientes pasos:

Objetivo: Como en cualquier situación de evaluación, es indispensable fijar claramente y tener a la vista el objetivo que se va a valorar.

Proposiciones: como se dijo al hablar en general de las PO, los MER se construyen mejor y más adecuadamente a partir de ideas recogidas en los materiales de instrucción.

Formulación de ítems: Para desarrollar ítems MER a partir de proposiciones es necesario:

- Formular una pregunta o una afirmación incompleta que encierre claramente un problema sobre la proposición; es la Base del ítem.
- Proporcionar una respuesta aceptable a la pregunta, expresada con pocas y bien escogidas palabras.
- Proporcionar varias respuestas plausibles pero incorrectas a la pregunta; son los distractores.

Instrucciones: hay que distinguir las instrucciones generales de las de cada ítem. Es especialmente importante que las instrucciones generales, que se refieren de manera uniforme a toda la prueba, sean claras; por ejemplo: tiempo disponible, dónde y cómo responder, criterios de calificación, etc. Los alumnos se adaptan y acostumbran con rapidez.

Presentación: es importante comprobar la frecuencia del lugar en que se coloca la respuesta correcta. Cada profesor tiene a colocarla en un lugar determinado. Esto da ventajas al *experto* en exámenes. Para evitarlo, cada posición de la respuesta correcta debe tener una misma proporción, de tal modo que, en conjunto de la prueba, una no aparezca más a menudo que otra. Se pueden hacer varias versiones de las mismas pruebas con sólo colocar los distractores en distinto orden.

Normas: nos centramos, sencillamente, en aquellas sugerencias específicas para formular y presentar buenos ítems MER, que en ciertos aspectos son comunes a todas las PO. La mayoría de ellas reflejan conclusiones prácticas a las que han llegado expertos en construirlos, y en algunas se añaden justificaciones racionales. Sin embargo, sólo unas pocas han sido contrastadas en investigaciones rigurosas, siempre difíciles de hacer en esta materia. No suelen ser grandes los efectos de obviar alguna de estas sugerencias; sin embargo, los profesores llegan a formular mejores

preguntas si conocen y siguen estas sugerencias que si las desconocen o desprecian.

2.2.1. Construcción de la Base

La función de la base del ítem es familiarizar al alumno con el problema que se le propone, por lo que el propósito de su sintaxis es comunicar significados explícitos tan eficazmente como sea posible. Pocos textos escritos se leen tan atentamente para comprender su significado, explícito o implícito, como las PO; se miran con *lupa*.

Es difícil exponer los criterios para construir buenos ítems MER. En cualquier caso, no se pueden reducir a comprobar una lista de normas, una por una. La competencia para construirlos, además del dominio de la materia, el rigor en el uso del vocabulario y de la expresión escrita, las estrategias instruccionales y las técnicas complejas de producción de PO, suponen una comprensión profunda de la capacidad de los alumnos para interpretar y responder los distintos ítems y una gran originalidad para crear situaciones de evaluación adecuadas.

Estos criterios se pueden reducir a que las preguntas sean claras y pertinentes. Algunas sugerencias pueden ser: las características que deben tener las ideas que se recogen en los ítems MER, las características de la información con que se explicitan y otras características complementarias.

Ideas: Se resumen las características que deben tener los datos y conceptos que se vierten en los ítems MER.

- **Novedades:** Es deseable evitar en los exámenes el uso de las mismas preguntas y problemas que se usaron durante la explicación. Por el contrario, introducir preguntas y situaciones problemáticas nuevas, incluso imposibles, refuerza al alumno que intenta comprender lo que se le enseñó y penaliza al alumno superficial.
- **Pregunta directa:** Con frecuencia es preferible una pregunta directa, expresada en forma interrogativa, ya que no sólo ayuda a presentar al alumno un problema más específico, sino que también puede centrar la intención del examinador para evitar distractores irrelevantes que no vengan a cuento. Hay que expresar la base del ítem tan directa, exacta y simplemente como sea posible, evitando lo innecesariamente complejo.
- **Respuesta definida:** Es su forma ideal, la respuesta a un ítem MER debe ser lo suficientemente correcta, de manera que no

admira diferencias de opinión, ni se fundamente en afirmaciones cuya validez puedan contradecir expertos competentes en la materia. Sin embargo, no hay que evitar preguntas importantes por el solo hecho de que no haya una respuesta completamente correcta; se trata de la variedad de escoger la mejor respuesta antes mencionada.

Información completa: El segundo grupo de sugerencias se refiere a la necesidad de dar al alumno claramente toda la información que necesita para responder a la pregunta.

- Base autosuficiente: Es lo más determinante y característico de los buenos ítems MER. El enunciado de la Base debe incluir una idea completa. Si se eliminan los distractores de un ítem MER, automáticamente debe transformarse en una Respuesta Abierta.
- Condiciones: Es necesario introducir todas las condiciones para hacer que la respuesta sea correcta. Pero, a veces, para asegurarse de que el ítem tenga una respuesta correcta, los examinadores reducen sus dimensiones hasta tal punto que el problema se diluye.
- Introducciones: Abunda en la norma anterior. A veces son necesarias frases de introducción para describir la escena o el contexto. Tales frases difieren de los preámbulos instruccionales y escaparates retóricos que se mencionan al hablar de suprimir lo irrelevante.

Si son necesarias muchas ideas descriptivas y cualificadoras, combinar las descripciones con las preguntas en una misma oración probablemente hace la base del ítem más compleja. Se puede conseguir una expresión más clara presentando estas descripciones e informaciones previas en frases de introducción separadas; en otra se hace la pregunta.

- Suprimir lo irrelevante: Aunque se puede aprender mucho durante un examen, la inclusión deliberada de materiales de enseñanza en las situaciones de evaluación puede reducir su eficacia, sin que aumente su valor instructivo.

Con mayor razón, deben evitarse los detalles y preámbulos retóricos que sólo sirven de escaparate o escenario y no ayudan al alumno a comprender la pregunta que se le hace.

Otras características: Se añaden por último otras sugerencias para lograr que el ítem sea claro y pertinente.

- **Pistas inintencionadas:** Este aspecto se recoge de nuevo al hablar de la homogeneidad de los distractores. Hay que evitar indicios que puedan contribuir a sugerir la respuesta correcta. Es fácil eliminar estas pistas si el profesor está sobre aviso. Sin el conocimiento adecuado, se puede adivinar la respuesta asociando elementos de la Base con una de las respuesta, o identificando algún elemento común en las alternativas para elegir la que no lo tenga. En algunos casos, palabras clave de la Base, o sus sinónimos, se repiten o están relacionadas con términos de la respuesta correcta.
Las preguntas deben ser también independientes, de manera que el enunciado de una pregunta no tiene que dar pistas para responder otra.
- **Negaciones:** Las negaciones resultan especialmente atractivas a los examinadores porque son fáciles de formular añadiendo un *no* a los materiales de instrucción. Sin embargo, las preguntas formuladas negativamente son confusas, ya que requieren que el alumno elija una respuesta que no es verdad o no es característica; por lo que se aconseja evitarlas, en la medida de lo posible. si es preciso incluirlas, se debe subrayar la expresión negativa para atraer la atención del alumno.
- **Dificultad dosificada:** Cualquier pregunta que sea excesivamente fácil o excesivamente difícil para un grupo de alumnos no discrimina sus niveles relativos de logro.
Los distractores simplistas perjudican a los mejores alumnos, ya que sus conocimientos adicionales interfieren con la respuesta que se puede dar a partir de la mera información que proporciona el ítem. El alumno sólo puede adivinar aproximadamente el grado de conocimiento adicional que se le presupone. Se llama inhibición proactiva a la situación en que el alumno, para solucionar un ítem, aporta conocimientos más allá de la intención del profesor. Este factor supone una discriminación negativa en contra de los mejores alumnos, que se debe tener en cuenta a la hora de construir ítems.

2.2.2. Construcción de las Respuestas

Una vez terminada la Base del ítem, hay que escribir la respuesta correcta así como un conjunto de distractores o respuestas incorrectas. El

propósito de los distractores es discriminar entre los alumnos que dominan y no dominan una materia. Para ello, el distractor debe ser una alternativa lógica y plausible que parezca tener relación con la pregunta. El alumno inteligente descartará por ridículos los distractores demasiado alejados de la pregunta; con ello se reduce el número de distractores reales. Se pueden emplear varias técnicas para preparar distractores plausibles, sin proporcionar pistas que ayuden al alumno no preparado a contestar correctamente; entre otras, se pueden citar los siguientes:

- Usar afirmaciones verdaderas que no contestan directamente a la pregunta.

- Emplear expresiones familiares estereotipadas que tengan alguna asociación con los términos usados en la pregunta y que parezcan atractivas a los alumnos simplemente por un recuerdo vago de haber oído aquellas palabras alguna vez.

- Definir la clase de cosas a la que deben pertenecer todas las respuestas alternativas. Por ejemplo, si se pregunta lo que enfría un refrigerador eléctrico, la clase de posibles respuestas se define como “cosas que pueden causar el enfriamiento” y los distractores serían: hielo, aire en movimiento, expansión de un gas, etc.

- Utilizar los errores cometidos por los alumnos en exámenes clásicos (RA) y de respuesta limitada (RL); aunque a veces no compensa por el trabajo que esto supone.

He aquí algunas sugerencias para generar buenos distractores en ítems MER, agrupadas por su homogeneidad interna y externa, y completadas con otros recursos que suelen utilizar los expertos.

Homogeneidad interna: Puesto que las opciones intentan ser respuestas a la misma pregunta, deben estar relacionadas con ella. Por tanto, hay que vigilar que los distractores y la respuesta correcta sean homogéneos internamente.

- Contenido: Una falta de paralelismo en las alternativas puede llevar a los malos alumnos a la respuesta correcta; por tanto, los distractores deben ser relativamente semejantes por su contenido. Desafortunadamente, este no es siempre el caso.
- Matizar los datos: Cuanto más homogéneas sean las respuestas, más difícil y, consecuentemente, más discriminante será el ítem. Por el contrario, una manera de transformar un ítem difícil para hacerlo más fácil es proporcionar varios datos en las alternativas.

- Distractores plausibles: Es obligado evitar alternativas jocosas, frecuentes entre los exámenes que los profesores proponen a sus propios alumnos, ya que suelen ser alternativas desperdiciadas por absurdas.
- Concisión: La brevedad de las respuestas simplifica la tarea del alumno eliminando una fuente irrelevante de dificultades. Las respuestas breves tienden también a centrar la atención en las diferencias esenciales entre las alternativas. En igualdad de condiciones, el ítem MER con respuestas más breves son preferibles a las demasiado complicadas; aunque no hay que sacrificar la importancia y significado de las preguntas para ganar brevedad en las respuesta.

Cuando las respuestas son largas y complejas, los alumnos tienen dificultad en comprender y mantener en la memoria las diferencias esenciales entre las alternativas.

Homogeneidad externa: Los distractores deben ser también homogéneos externamente.

- Elaboración gramatical: Es necesario cuidar la corrección gramatical (género y número); el simple descuido ortográfico o gramatical puede disminuir la eficacia de la pregunta. La forma interrogativa en la Base, preferible a las frases incompletas, ayuda a no introducir incoherencias gramaticales. Conviene que todas las alternativas tengan la misma complejidad y que su extensión sea aproximadamente la misma. Hay una tendencia a escribir las respuestas correctas de forma más cuidadosa e inclusiva que los distractores. Dar todas las pistas relevante, es decir, las útiles para los alumnos preparados, y evitar las irrelevantes es una importante destreza necesaria para construir ítems MER. Sin embargo, no es censurable incluir deliberadamente pistar en los distractores para equivocar a los alumnos *expertos* en contestar PO, pero mal preparados en la materia.
- Elementos Comunes: Hay que evitar reiteraciones de palabras innecesarias e introducir los elementos comunes en la base, sin repetir palabras en los distractores. Esto coincide con la norma de formular una idea completa en la Base. Cuando todas las respuestas empiecen por las mismas palabras, hay que considerar la posibilidad de incluirlas en la Base.

Conviene también eliminar los artículos en las alternativas que no sean frases completas, ya que pueden dar pista no deseables. Cuando sean homogéneas, hay que incluir las unidades de las respuestas en la base.

- Escalas: Si la pregunta exige una respuesta cuantitativa, es necesario escribir respuestas que representen intervalos a lo largo de una misma escala. Hay cuestiones en que es más importante la relación, o el conocimiento del orden de la magnitud a nivel general, que el valor preciso cuantitativo.
- Distractores no específicos: Cuando al profesor no se le ocurren distractores en número suficiente, un recurso común para completarlos es añadir como alternativa final *todas las anteriores* o *ninguna de las anteriores*.

El uso de *todas las anteriores* como respuesta correcta es adecuada sólo si todas las alternativas precedentes son respuestas enteramente correctas; ya que conociendo que una sola de las anteriores es incorrecta, queda descartada. Y al contrario, se puede intuir como respuesta correcta si el alumno conoce la mayoría de los distractores.

La respuesta *ninguna de las anteriores* es particularmente útil para ítems aritméticos o de ortografía en donde lo correcto y erróneo son inequívocos.

Recursos: Finalmente se señalan otros recursos y estrategias para lograr buenos distractores.

- Explicaciones añadidas: Enriquecer las preguntas de manera que se deban contestar, además de con *si* o *no*, con una explicación añadida.
- Combinar elementos: Usar combinaciones de elementos para obtener varias alternativas.
- Distractores en líneas diferentes: Las diferentes partes del ítem, Instrucciones, Base y Alternativas, deben presentarse en la misma página. De todas maneras, es más conveniente escribir los distractores en líneas diferentes, y no uno a continuación de otro.
- Número de distractores: Cuantos más distractores, más discriminará el ítem; pero los buenos distractores no se nos ocurren fácil e inmediatamente, y hay que tener en cuenta que cada distractor nuevo será más débil. [7] recomiendan más ítems

con menos distractores, que menos con más distractores. Aunque no hay un número mágico de alternativas, la práctica común aconseja escribir 4 opciones o, lo que es lo mismo, una respuesta y 3 distractores para cada ítem.

3. Fundamentos de los Test Adaptativos por Computador

Un test adaptativo se puede definir como un proceso de administración de test en el que los ítems que el alumno realiza son resultado de su actividad y las respuestas a ítems administrados previamente.

A diferencia de un test convencional, cuya máxima precisión se encuentra en aquellos niveles de habilidad cercanos a su dificultad promedio, un test adaptativo permite evaluar con la misma precisión a todos los individuos, independientemente del nivel de habilidad en que se encuentren. El único requisito es que los ítems que se van a administrar deben tener parámetros de dificultad y discriminación adecuados al nivel de habilidad donde se encuentra el alumno que se quiere evaluar.

Un test adaptativo producirá, por tanto, medidas de mayor calidad que un test convencional con un menor número de ítems. Esta mejora se refleja en un mayor incremento en la precisión de la medida en todos los niveles del rasgo del alumno que se está analizando, lo que a su vez implica una mejora de las características psicométricas propias de todo test: fiabilidad y validez [8].

Los mecanismos estándar para la selección de ítems pueden ser clasificados en dos grupos generales: los preestructurados y los de paso variable y la principal diferencia entre ellos es precisamente la adaptabilidad que proporcionan.

Los procedimientos preestructurados: realizan la selección de los ítems de una forma determinística y prefijada en el test. Constituyen los primeros intentos de test adaptativo y se basan en la utilización de algoritmos con criterios preestablecidos de selección de ítems para obtener las baterías de cuestiones. Se pueden citar entre los más representativos:

- El de dos etapas: básicamente consiste en la aplicación consecutiva de dos test, el primero, denominado test de rutas, cumple la función de asignar al alumno un nivel de rendimiento,

posteriormente se le aplicará el segundo test supuestamente adaptado a su nivel, [9].

- El piramidal: aquí los ítems están agrupados en niveles con el mismo índice de dificultad. Si un alumno responde correctamente, se procede a contestar a un ítem de más dificultad y si falla a uno más fácil, [10].
- El de nivel flexible: su estrategia de aplicación requiere un conjunto estructurado de ítems dispuestos por su nivel de dificultad. Cada alumno comienza el test a partir de un ítem de dificultad media, si la respuesta es correcta, se pasa al de más dificultad, y si no lo es, se desciende, procediendo así sucesivamente, [11].
- El adaptativo por estratos: esta estrategia parte de la estratificación de los niveles de dificultad de los ítems, y dentro de cada estrato, se disponen de mayor a menor discriminación, de forma que el alumno siempre responderá en primer lugar a los ítems mas discriminativos de cada estrato. Cuando el alumno responde a uno de los ítems, si acierta, se pasa al estrato siguiente, administrando el ítem con mas discriminación que no haya sido presentado, y si falla, al correspondiente del estado anterior [12].

Las objeciones que se pueden realizar a estos mecanismos en su conjunto, muchos de ellos todavía en uso, se basan esencialmente en que las estrategias de selección permanecen constantes a lo largo de todo el examen. Esto implica, desde el punto de vista de la adaptación, y por tanto, de la efectividad, una limitación muy importante, pues, aunque al principio se hace necesario presentar una secuencia de ítems sucesivos con un grado de dificultad muy diferenciado, dada la poca información que se tiene del alumno, precisado el nivel de éste, estas diferencias deben reducirse a medida que avance el test, adaptándose por tanto la secuencia de presentación a la propia actividad del alumno.

Los de paso variable: permiten una selección interactiva de forma prácticamente ilimitada a través del test. Merecen ser destacadas, como base del presente desarrollo, dos metodologías iniciales:

- La de la máxima información: conceptualmente es muy sencilla, se trata de escoger aquel ítem que maximizar la información cuando el estudiante responde. Para seleccionar dicho ítem, se tienen en cuenta, la calificación obtenida por el estudiante en ese momento y los parámetros que definen el ítem. Respondido el

test, se recalcula la calificación del alumno para ser utilizada en la selección del siguiente ítem o presentada a la finalización del examen, [13].

- El método Bayesiano: el segundo método básico en la selección de ítems es el Bayesiano, en él, cada examinado empieza el test con un nivel preestimado inicial y con un grado de confianza asociado. Éstos son manejados como la media y varianza de una distribución normal a priori. Cuando se responde a un ítem, se recalculan los niveles según la fórmula de Bayes, como probabilidades a posteriori a partir de la respuesta obtenida y de los valores a priori anteriores. A su vez, un método de selección bayesiana elige el ítem que reducirá en un mayor grado la varianza a posteriori, es decir, reducirá mas el error de la estimación, [14, 15].

A partir de estos mecanismos fundamentales se han desarrollado para casos prácticos multitud de variaciones modificaciones y ajustes. Tres de ellas que merecen ser especialmente referenciadas:

- El examen Auto Adaptado: por la que se permite al alumno participar en la selección de los ítems, minimizando de esta manera su ansiedad y maximizando, cuando está debidamente motivado, su rendimiento. Para ello, es necesario que los ítems estén preestructurados en grupos de dificultad al estilo de los tests adaptativos por estratos. Ésto, a su vez, posibilita también la participación del profesor para influir, de acuerdo con su experiencia, en el examen a realizar, [16].
- La segmentación del test: este método propone la organización de las baterías de ítems en clusters especializados por contenido o área de conocimiento. Estos segmentos pueden ser fácilmente revisados por los expertos, calificándolos en cuanto a dificultad y materia, eliminando interdependencias y estandarizándolos [17].
- Modelo con lógica difusa: considerando que los métodos probabilísticos adolecen de ciertos problemas como la asignación arbitraria de prioridades a priori, la no representatividad y el efecto de contexto en [18] hemos planteado un modelo alternativo basado en la lógica difusa considerando interpretaciones imprecisas del nivel de información de los ítems y sus distribuciones a priori.

4. Aplicación: Sistema Integrado de Procesado de Test

4.1. El Editor de Ítems


Con esta aplicación se pretende extender la construcción de los Ítems de tipo Test a lo que la mayoría de las personas que trabajan en estos campos están acostumbradas, es decir, en la mayoría de las aplicaciones que permiten al usuario editar o responder a Ítems de tipo Test se nos presenta el ítem de la siguiente forma base:

ENUNCIADO
ALTERNATIVA A) ALTERNATIVA B) ALTERNATIVA C) ALTERNATIVA D)

Pudiendo en algunos casos añadir algún elemento gráfico, que en la mayoría de los casos siempre irá situado entre el enunciado y las alternativas.

ENUNCIADO
GRÁFICO
ALTERNATIVA A) ALTERNATIVA B) ALTERNATIVA C) ALTERNATIVA D)

Ejemplo:


<p>¿Qué indica esta señal?</p> 
<ul style="list-style-type: none"><input type="checkbox"/> Acceso prohibido a toda clase de vehículos<input type="checkbox"/> Circulación prohibida en ambos sentidos<input type="checkbox"/> Circulación prohibida en sentido inverso al que la vemos

En la cual el usuario tendrá que seleccionar entre las distintas alternativas que se nos presentan y con el cursor o ratón marcar la que considera correcta.

Lo que se pretende con esta aplicación es ofrecer al usuario la posibilidad de poder construir la pregunta a su antojo, es decir, sin limitar ni el espacio del que dispone en un principio para el enunciado, sin limitar el número de alternativas que puede tener el Ítem ni su posición o localización dentro del Ítem. Se pretende que sea el propio usuario el que construya sus Ítems sin tener en cuenta otros impedimentos, ni formatos ya predefinidos.

A modo de ejemplo, con esta aplicación, se podría construir el ítem de las siguientes formas:


¿Qué indica esta señal?




- a) Acceso prohibido a toda clase de vehículos
- b) Circulación prohibida en ambos sentidos
- c) Circulación prohibida en sentido inverso al que la vemos

¿Qué indica esta señal?

- a) Acceso prohibido a toda clase de vehículos
- b) Circulación prohibida en ambos sentidos
- c) Circulación prohibida en sentido inverso al que la vemos



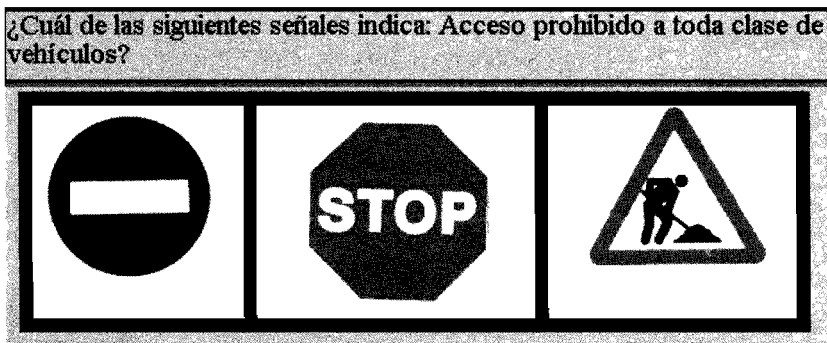
¿Qué indica esta señal?



- a) Acceso prohibido a toda clase de vehículos
- b) Circulación prohibida en ambos sentidos
- c) Circulación prohibida en sentido inverso al que la vemos

En definitiva, de lo que se trata es de que el usuario pueda estructurar sus Ítems como considere más oportuno. Podríamos considerar que el usuario puede “pintar” sus propios Ítems de la forma que desee, para ello dispone de una serie de herramientas que le van a permitir realizar estos pasos.

Una de las principales ventajas de la aplicación es que las posibles soluciones no tienen por que estar organizadas una en cada línea, ni con un determinado encabezamiento, ni siquiera las posibles soluciones o alternativas tienen que ser elementos de tipo texto. Podremos hacer que las posibles soluciones puedan ser elementos gráficos e incluso podemos hacer que las posibles soluciones se encuentren dentro de un elemento gráfico. Esto hace que se pueda expandir el campo en el que pueden aplicarse los Ítems de tipo test. Esta faceta se puede ver mejor en el siguiente Ítem:



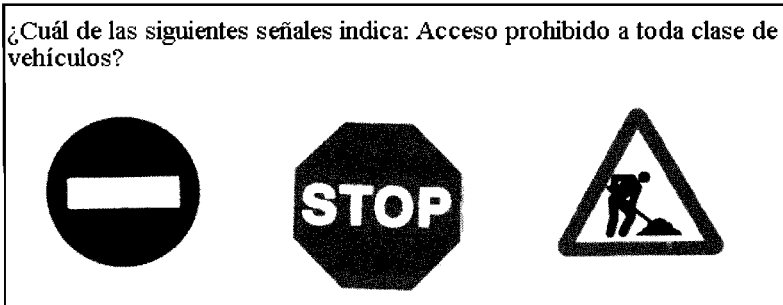
Con lo que el usuario final del Ítem deberá seleccionar una de las figuras.

Otro de los aspectos que pretende mejorar esta aplicación es la posibilidad de escribir varios formatos de texto, es decir, en la creación del Ítem se puede cambiar la fuente de letra que se utiliza en cualquier momento, con lo que podrá tener partes del texto con una determinada fuente y otras partes con otra fuente para diferenciar algunos aspectos del Ítem, así como diferentes tamaños, colores, etc. Esta faceta puede, de nuevo, ampliar aún más el campo de aplicación de los Ítems, ya que no solo las posibles soluciones se encuentran en formato texto, sino que podremos incluir las imágenes como soluciones al Ítem y además hacer que el usuario se sienta más interesado, ya que con ello podremos hacer los Ítems más visibles, más agradables a la vista y tratar que el usuario final de los Ítems

no caiga en la monotonía mientras realiza determinadas pruebas al presentarse estas siempre de forma similar.

La aplicación puede manejar dos tipos de elementos de texto. El tipo general, que nos permite introducir el texto del Ítem como si estuviéramos ante un procesador de texto común, y un tipo de texto gráfico, que nos va a permitir introducir texto en cualquier zona de la pantalla visible para el usuario final, sin tener relación con el texto normal del Ítem. Con este segundo tipo de texto podremos señalar o realzar determinados aspectos de un gráfico, etc.

Además de los elementos de tipo texto, se añade la posibilidad de poder incluir en los Ítems dos “capas” de figuras simples que nos permitirán realzar los Ítems construidos, bien remarcando o sirviendo de fondo para las posibles respuestas, como añadir elementos simples de dibujo (líneas, círculos y rectángulos). La primera capa se situará por debajo del texto del Ítem, por lo que tenemos la posibilidad de cambiar el color de fondo del Ítem o de algunas partes del Ítem. La segunda “capa” irá tanto por encima del texto como de los gráficos que necesitemos y nos puede servir para “tapar” determinados elementos de los aspectos gráficos que no deseamos que sean visibles, así como poder realizar el “efecto de marco” para las figuras y el texto. Ejemplo:



Vistos los resultados, con muy poco aprendizaje se pueden llegar a crear Ítems que sean llamativos para el usuario final.

Otro de los aspectos que no hemos tratado hasta ahora es la forma en que seleccionamos las posibles alternativas del Ítem. Una vez realizado el Ítem, deberemos, con la ayuda de una herramienta de la aplicación, “marcar” o delimitar las posibles alternativas que se le presentarán al usuario. Con esta herramienta iremos delimitando todas las posibles soluciones que el usuario que construye el ítem irá viendo como rectángulos

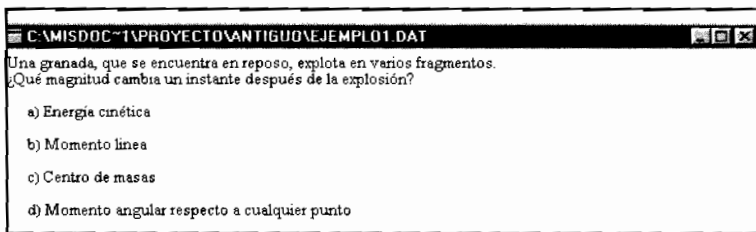
que se sitúan encima de las alternativas con líneas discontinuas (este efecto no será apreciado cuando se presente al usuario final el Ítem), y una vez marcadas todas las posibles alternativas, con otra herramienta de la aplicación seleccionaremos con un simple clic del ratón la alternativa o alternativas correctas, ya que la aplicación permite que los Ítems tengan más de una respuesta correcta para el caso en que debamos de señalar dos o más opciones válidas.

Teniendo en cuenta que en la mayoría de las materias se pueden realizar Ítems para realizar pruebas o aprendizaje con estos elementos queda muy abierto los campos en los que se puede utilizar la aplicación, ya sea en Matemáticas, Física, Álgebra, Lógica, Lenguaje, Historia, etc. Además como elemento añadido a la aplicación podremos incluir dentro del Ítem elementos de sonido, con lo que, por ejemplo, podremos mostrar al usuario una parte del Ítem, junto a las posibles soluciones y el encabezado del Ítem podría ser un elemento que incluye sonido, es decir, se puede grabar con cualquier programa de grabación de sonido que tienen actualmente los equipos multimedia el encabezado del Ítem, con lo que se puede ampliar el campo donde se pueden utilizar estos Ítems, por ejemplo en Música, en la cual se puede presentar al usuario un Ítem preguntándole por la clase de Instrumento que emite un determinado sonido, que evidentemente el usuario podrá escuchar y después seleccionar la alternativa que crea conveniente. Por poner un ejemplo, imaginémosnos un Ítem que tenga un encabezado como el siguiente:

¿Qué clase de instrumento emite este sonido?

Junto con las figuras de varios instrumentos musicales, y mientras el usuario estudia el Ítem puede escuchar el instrumento en cuestión, pudiendo posteriormente el usuario seleccionar de los instrumentos que se le presenten el que considere oportuno.

Hasta aquí hemos realizado un repaso por las características más importantes con las que cuenta la aplicación para construir los Ítems, sin embargo y para que se vean más claramente que la aplicación permite construir Ítems de una gran variedad de temas, les mostramos los siguientes ejemplos:



C:\MISDOC\1\PROYECTO\ANTIGUO\EJEMPLO1.DAT

¿Cómo se modifican las lecturas del voltímetro y de los amperímetros A_1 y A_2 al cerrar el interruptor I?

	V	A_1	A_2
A)	No cambia	Aumenta	Disminuye
B)	No cambia	Aumenta	No cambia
C)	Aumenta	Disminuye	Aumenta
D)	Disminuye	Aumenta	Disminuye

Y una vez visto estos ejemplos de Ítems que se pueden construir con la aplicación daremos una vista general del aspecto de la aplicación:

r de Ítems

Texto Gráficos Sonido Catalogo Test Ayuda

Herramientas

C:\MISDOC\1\PROYECTO\ANTIGUO\EJEMPLO1.DAT

¿Cómo se modifican las lecturas del voltímetro y de los amperímetros A_1 y A_2 al cerrar el interruptor I?

	V	A_1	A_2
A)	No cambia	Aumenta	Disminuye
B)	No cambia	Aumenta	No cambia
C)	Aumenta	Disminuye	Aumenta
D)	Disminuye	Aumenta	Disminuye

172,105

4.2. Catalogación de los Ítems

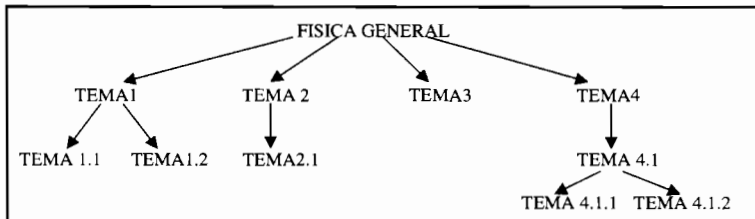
Otro de los aspectos importantes de la aplicación es que incorpora un apartado para la ordenación y catalogación de los Ítems según las categorías o temas con los que el ítem esté relacionado. Mediante la catalogación de los Ítems se pueden ir desarrollando un “árbol” con los diferentes temas para poder ordenar de una manera más eficiente las preguntas. Posteriormente esta ordenación puede ser utilizada para obtener una serie de Ítems que cumplan determinados requisitos, formando una batería de preguntas que pueden ser utilizadas a modo de examen o autoevaluación.

A modo de aclaración de lo que podríamos generar con esta catalogación veremos el siguiente ejemplo, en el cual se da una serie de nombres genéricos a los temas que el usuario es libre de cambiar y poner a su antojo, ya que la aplicación permite la entrada de frases identificativas de los temas, sin tener que atenerse a ninguna numeración en particular.

CATALOGO: FÍSICA

TEMAS: FÍSICA GENERAL
TEMA 1
TEMA 1.1
TEMA 1.2
TEMA 2
TEMA 2.1
TEMA 3
TEMA 4
TEMA 4.1
TEMA 4.1.1
TEMA 4.1.2

Evidentemente de esta forma resulta cuanto menos dificultosa o difícil de ver las relaciones que existen entre los temas. Con la ordenación de los Ítems dentro de los catálogos y relacionando mediante la aplicación los diferentes temas, se puede llegar a construir un jerarquía o gráfico de relaciones como el siguiente:



Cabe destacar el hecho de que mediante la aplicación estas relaciones se pueden cambiar de una forma sencilla, así como añadir evidentemente nuevos temas que sirvan como subtemas de los ya introducidos o añadir temas de mayor peso o temas padres que pueden enlazar estos subgrafos con otros subgrafos.

Un hecho a destacar es que la aplicación permite que un mismo Item pueda ser catalogado dentro de distintos temas e incluso ser catalogado dentro de diferentes catálogos, con lo que se da un amplio margen al usuario para que organice sus propios Ítems.

La importancia de los catálogos es clave para la aplicación, ya que además de permitir una clasificación de los Ítems que pueda servir al usuario, la aplicación hace uso de dichos catálogos para generar las pruebas de Test.

4.3. Creación y realización de pruebas

Una vez que ha sido creado y catalogado un banco de ítems, el siguiente paso es la realización de la prueba por parte del alumno. Para ello, el profesor dispone de una serie de opciones que le permiten seleccionar ítems o grupos de ítems de forma manual, de acuerdo con sus necesidades y preferencias en función a los alumnos a los que va dirigido, o bien de forma aleatoria.

En la actualidad la aplicación puede generar tres tipos de Test:

- Test con Ítems directamente seleccionados por el profesor.
- Test generados aleatoriamente de entre los que están asociados a un determinado catálogo.
- Test que están compuestos por Ítems que se seleccionan aleatoriamente dentro de unos determinados temas que el usuario puede indicar a la aplicación de entre los disponibles en el catálogo.

En la realización de estos Test, además del número de Ítems que los componen y los temas que se incluyen, se tienen que introducir los parámetros de puntuación para que posteriormente se pueda evaluar a una determinada persona. Estos parámetros incluyen la puntuación que se obtiene por Ítem contestado correctamente, incorrectamente o no contestado.

Evidentemente, una vez generados estas pruebas de tipo Test, se pueden utilizar para la evaluación o realizar una autoevaluación de un determinado alumno. Para ello la aplicación incluye un programa que

permite cargar estas pruebas de tipo test y que el usuario realice dichas pruebas, para que al finalizar la aplicación nos de un informe con el resultado obtenido según los parámetros que el test incluya.

5. Direcciones de los autores

Para recibir la aplicación presentada basta con ponerse en contacto con: *Ángel Neira*
Departamento de Informática. Campus de Viesques.
Universidad de Oviedo, 33204-Gijón.
Tfn: 98518 2481 e-mail: neira@correo.uniovi.es

6. Referencias

1. Osterlind, S.J. (1998). *Constructing test items*. Boston: Kluwer Academic Publishers.
2. Vaquero, A. (1990). *La enseñanza asistida por computadora*. En Métodos informáticos aplicados a la Psicología. Eds. Algarabel, S.; Sanmartín, J. Pirámide.
3. R. Neira, T. (Coord.) (En prensa). *La Evaluación en el Aula*. Oviedo: Nobel.
4. Hanrahan, H.E. Effective examining using short questions. *International Journal of Electrical Engineering Education*, Vol. 29/3. 1992.
5. García-Mendoza, A. y Soler, E. (1994). *Evaluación de Aprendizajes. Banco de Pruebas. Tomo 1. Física*. I.C.E. Aula Abierta, Monografía 23.
6. Pereda, M.A. y Soler, E. (1994). *Evaluación de Aprendizajes. Banco de Pruebas. Tomo 2. Química*. I.C.E. Aula Abierta, Monografía 24.
7. Pereda, M.A. y Soler, E. (1994). *Evaluación de Aprendizajes. Banco de Pruebas. Tomo 2. Química*. I.C.E. Aula Abierta, Monografía 24.
8. López, J.A., Ato, M., Sánchez, J. y Veladrino, A. (1990). *Test y diagnóstico psicológico por computador*. En Métodos informáticos aplicados a la Psicología. Eds. Algarabel, S.; Sanmartín, J. Pirámide.
9. Angoff, W.H. & Huddlesron, E.M. (1968). *The multi-level experiment. A study of a two-level test system for the College Board Scholastic Aptitude Test*. Princeton, NJ: Educational Testing Service.
10. Hansen, D.N. (1969). An investigation of computer-based science testing, pp.209-226. En Atkinson & Wilson (Ed.). *Computer-assisted instruction*. New York: Academic.
11. Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.
12. Weiss, D.J. (1973). *The stratified adaptive computerized ability test*. RR No. 73-3. Univ. Minnesota. Dep. Psychology.

13. Lord, F.M. (1977). A broad-range tailored test of verbal ability. *Applied Psychological Measurement 1*.
14. Owen, Roger J. (1975). A Bayesian Secuential Procedure for Quantal Response in the Context of Adaptive Mental Testing. *Journal of the American Statistical Association*, pp. 351-356.
15. Millán, E., Trella, M., Pérez de la Cruz, J.L. y Cornejo, R. (1999). *Uso de redes bayesianas en test adaptativos computerizados*. Congreso Nacional de Informática Educativa. Puertollano.
16. Rocklin, T. (1989). *Individual differences in item selection in computerized self-adapted testing*. American Educational Research Association. San Francisco.
17. Wainer, H. & Kiely, G.L. (1987). Item clusters and computerized adaptive testing. *Journal of Educational Measurement*, 24. pp. 189-205.
18. Brugos, J.A.L., Neira, A., García, V. y Alguero, A. (1999). *Un modelo de verificación de aprendizajes mediante tests adaptativos con lógica difusa*. Congreso Nacional de Informática Educativa. Puertollano.