

SIETTE: Un Sistema de Evaluación mediante Test basado en la Teoría de Respuesta al Ítem.

1. INTRODUCCIÓN

Uno de los mecanismos de evaluación más extendidos por su facilidad de corrección, es la realización de tests. La realización de tests tiene la ventaja de sistematizar la evaluación, por lo que ha sido ampliamente usada en Sistemas de Enseñanza Asistida por Ordenador (SEAO) y en Sistemas Tutores Inteligentes (STI; Brusilovsky y Miller, 1999). En los sistemas de enseñanza tradicional, el uso de tests de papel y lápiz tiene ciertos inconvenientes. La evaluación que se obtiene tras la realización de un test mediante papel y lápiz raramente tiene en cuenta más que el número de respuestas acertadas, y no la variabilidad en la dificultad de las preguntas, ni otros factores como la probabilidad de acertar una pregunta al azar, sin realmente saber la respuesta.

Por otra parte, en la última década, los SEAO y los STI han hallado en Internet, y más concretamente en la World Wide Web (WWW), un medio ideal de difusión. Desde el punto de vista de los STI, la WWW ofrece muchas ventajas en comparación con otros medios más tradicionales de difusión de programas: no requiere de instalaciones adicionales; se puede acceder desde cualquier lugar sin necesidad de transportar el software; distribución y actualización inmediata de contenidos; etc. Sin embargo, uno de los inconvenientes es la complejidad técnica que requiere tanto el desarrollo de sistemas personalizados como la creación de contenidos.

SIETTE es una aplicación eficiente de la teoría de Test Adaptativos Informatizados (TAI) y de la Teoría de Respuesta al Item (TRI) para la realización de tests a través de WWW. SIETTE ha sido diseñado como un componente autónomo, aunque puede también integrarse como parte de un STI. En concreto se prevé su integración en la arquitectura MEDEA (Trella, Conejo y Bueno, 2000).

En el siguiente epígrafe se presenta la arquitectura general del sistema SIETTE, sus principales componentes y funciones. Seguidamente en los apartados 3, y 4 se introducen brevemente los fundamentos teóricos de SIETTE, describiendo las características propias de la implementación en SIETTE. El apartado 5 trata sobre la integración de SIETTE como módulo de evaluación de un STI, centrándose en la utilización de ítems multidimensionales. Finalmente el apartado 6 presenta las conclusiones y las líneas de investigación en marcha.

2. ARQUITECTURA DEL SISTEMA SIETTE

SIETTE (Ríos, Pérez de la Cruz y Conejo, 1998) es un sistema de evaluación mediante TAI basado en la TRI. Utilizando un navegador estándar como interfaz gráfica de usuario, ofrece a los alumnos un aula virtual en el que realizar tests y permite a los profesores crear nuevos tests o añadir nuevas cuestiones a los tests ya creados.

La arquitectura del sistema SIETTE (Figura 1) contiene los principales componentes de un TAI agrupados en seis módulos principales:

- La *base de conocimientos*, Está compuesta por *el currículum o estructura del temario*, *las especificaciones de tests* y *el banco de ítems*, que constituye la colección de posibles cuestiones a presentar en un test, todas ellas calibradas con una serie de parámetros.
- El *generador de tests*. Es el encargado de seleccionar las preguntas a plantear al alumno, según las especificaciones del test y *el modelo temporal del alumno*.
- El *módulo de edición* permite a los profesores acceder a la base de conocimientos para almacenar las preguntas y respuestas y especificar *el currículum*, y *los tests* sobre los temas que se desea evaluar.

- El *módulo de comprobación y activación*. Para que los tests diseñados por el profesor pasen a disposición de los alumnos, sus datos deben ser validados. Sólo aquellas especificaciones de tests que cumplen los criterios mínimos de consistencia requeridos serán activadas por este módulo.

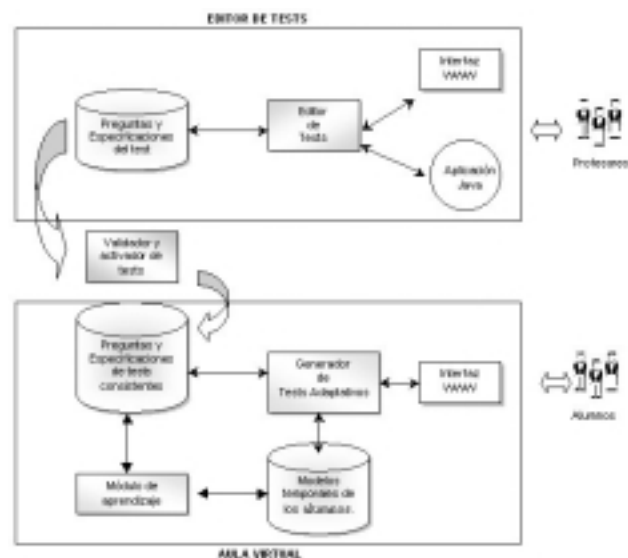


Figura 1: Arquitectura del sistema SIETTE.

- El *módulo de aprendizaje*. Se encarga de realizar una calibración de los parámetros de los items a partir de la información obtenida de forma empírica tras las sucesivas ejecuciones de los tests por parte de los estudiantes. Esta calibración es necesaria ya que los valores asociados a los items que introducen por los profesores, sólo se consideran aproximaciones iniciales (Conejo, Millán, Pérez de la Cruz y Trella, 2000).

3. ESTIMACIÓN DEL RASGO LATENTE.

Existen diversos métodos basados en TRI para estimar el valor del rasgo latente θ a partir de las respuestas del alumno a los items: método de máxima verosimilitud (Lord, 1980), bayesiano (Owen, 1975), etc.

SIETTE utiliza el método bayesiano, en el que la distribución de probabilidad del nivel de conocimiento del alumno con posterioridad a la respuesta a cada cuestiones del test se calcula aplicando el teorema de Bayes.

$$P(\theta/u) = P(\theta/u_1, \dots, u_n) = \frac{\prod_{i=1}^n P_i(\theta)^{u_i} (1 - P_i(\theta))^{(1-u_i)}}{\prod_{i=1}^n P(u_i)^{u_i} (1 - P(u_i))^{(1-u_i)}} P(\theta) \quad (1)$$

Finalmente, se toma como estimador el valor máximo de esta distribución, que como se puede apreciar resulta ser proporcional al producto de la función de verosimilitud $L(\theta/u)$ por la función de densidad a priori $P(\theta)$.

$$P(\theta/u) \propto L(\theta/u) P(\theta). \quad (2)$$

Aunque estas ecuaciones (1) y (2) pueden ser resueltas algebraicamente, el cálculo resulta complicado y computacionalmente costoso. Por otra parte para la mayoría de los STI no es necesaria una precisión muy grande en la evaluación. En general basta con un conjunto de valores discretos. En SIETTE se asume que el rasgo latente θ puede tomar solamente K valores discretos entre 0 y $K-1$. Se considera que cada curva característica del ítem (CCI) viene dada por un vector de K componentes $P_i(u|\theta) = (p_i(u|\theta=0), p_i(u|\theta=1), \dots, p_i(u|\theta=K-1))$. La estimación del valor latente se obtiene mediante el método bayesiano, que queda simplificado en este caso al producto vectorial de los vectores característicos de los ítems por el vector de densidad a priori, normalizado de forma que la suma de probabilidades sea 1 . Por lo que la fórmula (1) puede expresarse como:

$$\overline{P(\theta|u)} \propto \prod_{i=1}^n \left[\overline{P_i(\theta)}^{u_i} (\overline{1 - P_i(\theta)})^{(1-u_i)} \overline{P(\theta)} \right] \quad (3)$$

En principio, las CCI no requieren de parámetros para su definición, y por tanto la utilización de una función de distribución normal (Lord y Novick, 1968) o de una función logística (Rasch, 1960; Birnbaum, 1968) no es una elección que deba hacerse a priori, sino que puede estimarse sin restricciones a partir de las observaciones como veremos más adelante.

Otra característica muy interesante de la aproximación discreta es que el número de valores K que puede tomar el rasgo latente θ , en este caso el nivel de conocimiento, no necesita prefijarse totalmente al definir el test, sino que puede elegirse en el momento de iniciar cada test. Evidentemente, el número de valores de cualquier test no puede ser superior a los K que se almacenan para definir las CCI de cada ítem, pero dado que se trata de un producto vectorial, establecer un valor K' submúltiplo de K equivale a agrupar K/K' valores consecutivos de las CCI, sumando los valores, es decir, si las CCI están definidas como vectores de dimensión K : $\overline{P}(\theta) = (p_0, p_1, \dots, p_{k-1})$. y se desea evaluar a un alumno mediante K' niveles, las CCI que deben emplear serán: $\overline{P}'(\theta) = (p'_0, p'_1, \dots, p'_{k'-1})$, en donde

$$p'_j = K'/K \times (p_{j \times K/K'} + p_{j \times K/K'+1}, \dots, p_{j \times K/K'+K/K'-1}) \quad (4)$$

Esto ofrece una ventaja computacional considerable frente al método basado en distribuciones reales, en las que la complejidad computacional es sólo función del número de ítems del test.

Aunque no son estrictamente necesarios, SIETTE emplea tres parámetros por analogía con el modelo más extendido de la TRI, para la definición heurística de las CCI. Dado que SIETTE emplea solamente K niveles de conocimiento las CCI se aproximan mediante los valores de la curva real en puntos equidistantes del intervalo $[-(K-1)/2, (K-1)/2]$. SIETTE puede utilizar tanto aproximaciones basadas la función de

distribución normal como en la función logística. Por ejemplo, en este segundo caso el vector característico de un ítem en SIETTE $P(\theta) = (p_0, p_1, \dots, p_{K-1})$, puede definirse inicialmente mediante la función:

$$p_k = p_k(u=1|\theta=k) = c + (1-c) \frac{1}{1 + e^{-1.7a \left(\left(k - \frac{K-1}{2} \right) - b \right)}} \quad (5)$$

Puesto que SIETTE sólo emplea K valores de conocimiento, se exige que el parámetro de dificultad de todas las curvas tome también valores entre 0 y $K-1$.

También está implementado, aunque por el momento no se utiliza, un cuarto factor, al que se ha denominado *factor de distracción*, y que representa la probabilidad de que un alumno con conocimiento máximo falle la pregunta como consecuencia de una distracción. Este parámetro se trata análogamente al factor de adivinanza.

Además, SIETTE es capaz de modificar los vectores característicos de los ítems a partir de los datos obtenidos del uso del sistema. Esto modifica la forma de las CCI, que ya no pertenecerán a la familia normal o logística seleccionada inicialmente. Si se desea, se puede configurar SIETTE de manera que mantenga siempre vectores correspondientes a curvas normales o logísticas de la familia de tres parámetros, como se hace en la TRI clásica. Este proceso, inverso al anterior, se realiza mediante ajuste por mínimos cuadrados, entre el vector característico en uso, y la familia triparamétrica de curvas características del modelo clásico.

4. CRITERIOS DE SELECCIÓN DE ÍTEMS Y DE FINALIZACIÓN.

En SIETTE, el *criterio de selección* de preguntas es configurable para cada test a elección entre los siguientes: 1) *Bayesiano*, dada la distribución de conocimiento del alumno, para todas las preguntas no planteadas, se elige aquella que hace mínima la suma de varianzas a posteriori de las distribuciones resultantes del acierto o fallo del

ítem. 2) *Basado en máxima información*, selecciona al ítem cuya dificultad es más próxima en valor absoluto al valor más probable del nivel de conocimiento del alumno.

3) *Aleatorio*, esta opción permite desactivar el mecanismo de adaptación.

El *criterio de finalización* es igualmente configurable para cada test, seleccionando una combinación de entre los siguientes: a) Valor más probable de la distribución de conocimiento estimado superior a un cierto umbral. b) Varianza de la distribución de conocimiento estimado menor que un valor dado. c) Número mínimo y máximo de preguntas, para garantizar que en cualquier caso el test termina, y que se han realizado un mínimo número de preguntas que amortiguan el efecto de la estimación inicial del conocimiento del alumno.

SIETTE asume por defecto que la distribución de conocimiento a priori del alumno antes de efectuar el test es uniforme, es decir, todos los niveles de conocimiento son inicialmente equiprobables, para un alumno cuyo historial se desconoce. Otras opciones son también posibles, como considerar distribuciones binomiales centradas, o funciones de densidad obtenidas a partir de los datos de la población de alumnos que han realizado el mismo test hasta el momento. SIETTE guarda información sobre la distribución de conocimiento de un alumno concreto, y también puede usarla como punto de partida en sucesivas evaluaciones mediante el mismo test.

La evaluación multidimensional con ítems unidimensionales se realiza mediante la jerarquía de conceptos del currículum, evaluando cada nodo independientemente, y asumiendo herencia. Es decir, si un ítem pertenece a un nodo T_{ijk} también pertenece al nodo T_{ij} y al nodo T_i , y se evalúan todos ellos con la CCI unidimensional asociada a cada concepto.

5. INTEGRACIÓN EN SISTEMAS TUTORES INTELIGENTES.

La característica fundamental de los STI frente a los SEAO es que son sistemas que se adaptan al alumno ofreciendo una instrucción personalizada (Self, 1990). Tradicionalmente estos sistemas contienen un modelo del conocimiento de la materia a enseñar, un modelo del alumno y un planificador de instrucción, que decide cuál será el siguiente paso de instrucción. Para tomar estas decisiones, el planificador se basa en el modelo del alumno, eligiendo el siguiente tema a enseñar y el modo de enseñarlo de manera que la instrucción sea más efectiva. Este modelo frecuentemente está basado en el modelo del dominio, asignando un conocimiento a cada concepto que componen la materia a enseñar y se conoce como *modelo de overlay*. (Polson y Ricardson, 1988). El papel de la evaluación en un STI es, por tanto, la medición de los valores asignados a cada nodo que compone el modelo del alumno.

Hasta el momento se ha presentado SIETTE como un sistema capaz de medir un único rasgo. El uso de SIETTE como módulo de evaluación de un STI requiere un mayor detalle en la evaluación.

La base de conocimientos en SIETTE ha sido descrita brevemente en el apartado 2. Está compuesta por bases de conocimiento distintas para cada materia enseñada. Cada una esta formada por tres tipos de objetos (véase Figura 2):

- *Conceptos o temas*, que son los elementos en los que se descompone la materia a enseñar. Están estructurados jerárquicamente en temas y subtemas formando el *currículum*. SIETTE puede trabajar con un número indefinido de niveles en esta jerarquía. Cada nodo final corresponde a uno o a un conjunto de conceptos indiscernibles de cara a la evaluación. Los nodos intermedios de la jerarquía representan agregaciones de los subtemas de la jerarquía inferior según una

relación de pertenencia. El modelo del alumno asocia un nivel de conocimiento a cada tema, ya sea nodo terminal o intermedio. El currículum lo define el profesor y su mayor o menor detalle vendrá determinado por la precisión requerida en la evaluación. Se asume la independencia entre los valores del nivel de conocimiento de cualquiera dos nodos del currículum siempre que ninguno de ellos sea antecesor del otro.

- *Items*. Deben estar asociados explícitamente a uno o varios temas ya sean terminales o intermedios. Esta asociación indica que para responder correctamente al ítem es necesario el conocimiento de esos temas. La relación entre el nivel de conocimiento del tema y la respuesta al ítem viene dada por la CCI. Hasta el momento se ha considerado que cada ítem está asociado a un único tema, por lo que se obtienen CCI *unidimensionales*. En el caso en que sea necesario el conocimiento sobre varios temas para la resolución de un ítem, deben definirse CCI *multidimensionales*, que definan la probabilidad de contestar correctamente al ítem en función de la combinación de niveles de conocimiento de los temas necesarios. Se impone la restricción de que un ítem sólo puede estar asociado a varios temas en el caso en que estos sean hermanos en la jerarquía de temas. Este aspecto se tratará más adelante.
- *Tests*. Se definen en función del tema o temas a evaluar. Los ítems correspondientes a un test serán los necesarios para realizar la evaluación. No existe una asociación directa entre test e ítem, salvo a través de los temas. Se impone además la restricción de que un test sólo puede estar asociado a temas que sean hermanos en la jerarquía de temas. Al asociar un test a un tema pueden fijarse dos modos de evaluación: *agregada*, en el caso en que sólo se requiera

la evaluación de ese nodo del currículum; o *completa*, que indica que es necesaria una evaluación exhaustiva de todos los nodos del subárbol cuya raíz es este tema.

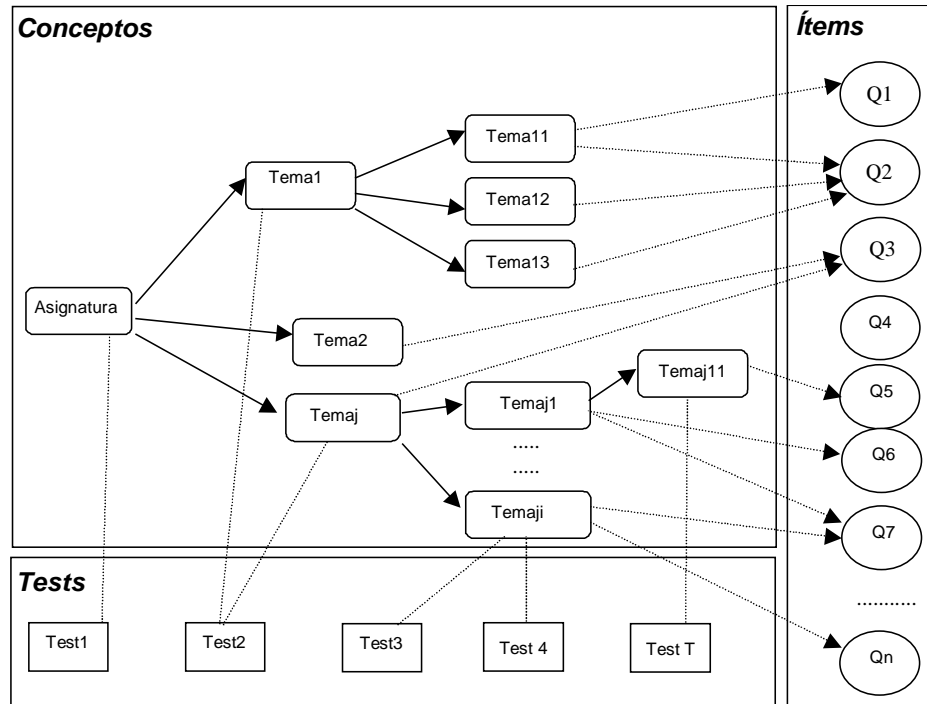


Figura 2 : Estructura de la base de conocimientos

5.1. Evaluación con ítems multidimensionales.

Hasta ahora se ha supuesto que la respuesta a un ítem viene condicionada solamente por el nivel de conocimiento de un único tema. Esto no siempre es así, es posible encontrar cuestiones cuya resolución depende de dos o más conceptos independientes. La presencia de estos ítems convierte a los modelos de evaluación en modelos multidimensionales.

En general, en el caso multidimensional, la CCI de un ítem i viene dada por la probabilidad condicionada de responder correctamente, dada una combinación de conocimientos de los s temas que influyen en la respuesta, es decir $P_i(\theta_1 \dots \theta_s) =$

$(P(u_i=1 \mid \theta_1 \dots \theta_s))$. En la TRI clásica, rara vez se consideran modelos de más de dos dimensiones dada la complejidad de resolución de las ecuaciones resultantes. En este caso las CCI vienen dadas por funciones basadas en la distribución normal (McDonald, 1996) o logística bidimensionales (Reckase, 1996). En general, se consideran familias de curvas con un menor número de parámetros de los que cabría esperar por la agregación. Por ejemplo, en el caso de la distribución normal, considerando solamente dos factores de dificultad y un único factor de adivinanza y discriminación se obtiene la familia de curvas características:

$$P(\theta, \theta') = c + (1 - c) \frac{1}{2\pi a^2} \int_{-\infty}^y \int_{-\infty}^x e^{-\frac{(x-b)^2 - (y-b)^2}{2a^2}} dx dy \quad (6)$$

Dadas las probabilidades condicionadas, para obtener los estimadores de los s niveles de conocimiento se aplica el procedimiento de máxima verosimilitud conjunta, o bien se calcula la densidad de la probabilidad a posteriori en el caso bayesiano $P_i(\theta_1 \dots \theta_s \mid u_1 \dots u_n)$, que es una función s -dimensional proporcional a:

$$P(\theta_1 \dots \theta_s) \times \prod_{i=1}^n P_i(\theta_1 \dots \theta_s)^{u_i} (1 - P_i(\theta_1 \dots \theta_s))^{(1-u_i)} \quad (7)$$

en donde $P(\theta_1 \dots \theta_s)$ es la probabilidad conjunta a priori.

En SIETTE el cálculo de la probabilidad a posteriori es relativamente simple, dado que las CCI se transforman en matrices s dimensionales de k componentes. En general se consideran matrices cuadradas, asumiendo que el número de niveles de conocimiento para cada rasgo es el mismo, lo que supone un total de k^s valores para cada curva. Claramente el tamaño las curvas es exponencial con el número de temas, pero desde un punto de vista práctico, el problema es abordable para valores de k^s que puedan ser procesados en un tiempo de respuesta razonable.

Los modelos del alumno necesarios para los STI requieren muchos más componentes. La inclusión de ítems múltiples no simplifica precisamente el problema, pero en ciertos casos, las características del ítem obligan a realizar una definición multidimensional para mantener la corrección del modelo de respuesta, basada en la hipótesis de independencia entre los temas de un mismo nivel en la jerarquía. En otro caso se estarían falseando los resultados.

Una primera restricción que debe imponerse para controlar la explosión combinatoria que se produce al utilizar ítems multidimensionales, es limitar el número de temas de los cuales pueda depender la respuesta a un ítem. Desgraciadamente, esta limitación sólo afecta a las CCI que se mantendrán menores de esa dimensión, pero no a la dimensión de la función resultante de la estimación de la probabilidad a posteriori, que vendrá dada, en el caso general, por el total de los temas que influyen en los n ítems del test. La condición de que los ítems multidimensionales sólo pueden estar asociados a temas que tengan un mismo antecesor en del currículum, garantiza al menos que la función podrá descomponerse en otras cuya máxima dimensión será menor que el número total de herederos.

Por ejemplo, sea el ítem Q_1 que depende de los temas T_1 y T_2 ; el ítem Q_2 que depende de los ítems T_2 y T_3 ; y el ítem Q_3 que depende de los temas T_3 y T_4 . Tras la aplicación de estos tres, se obtendrá una distribución conjunta de probabilidades para los niveles de conocimiento de los cuatro temas $P(\theta_1, \theta_2, \theta_3, \theta_4 | u_1, u_2, u_3)$. Sin embargo, si se elimina de la realización del test el ítem Q_2 , la evaluación puede realizarse de forma independiente para cada pareja de temas, es decir, se obtendrían dos distribuciones bidimensionales $P(\theta_1, \theta_2 | u_1)$ y $P(\theta_3, \theta_4 | u_3)$. Extrapolando esta idea, se puede limitar el número de dimensiones que se evalúan de forma conjunta mediante el análisis del

grafo de interferencias entre los temas que se establece, al considerar todos los posibles items, y aplicar un algoritmo de coloreado de grafos para eliminar del test los items necesarios para mantener el número máximo de dimensiones requerido. Si bien este algoritmo es a su vez exponencial, existen algoritmos eficientes que aunque no alcanzan el óptimo resultan aplicables en la práctica. Esta opción no está actualmente implementada en SIETTE, siendo otra línea de investigación abierta.

El uso de items multidimensionales es compatible con una evaluación jerárquica en paralelo de todos los temas del tests de forma simultánea. Por otra parte, siempre es posible sustituir los items multidimensionales por items unidimensionales asociados al tema inmediato superior en la jerarquía de conceptos. La diferencia es que en este caso se obtendría un modelo del alumno algo más tosco.

6. CONCLUSIONES.

En general todo el mundo está de acuerdo en que la aplicación de la informática al campo de la educación resulta muy útil. Pues bien, en este sentido creemos que el sistema de test por ordenador SIETTE aporta ventajas substanciales a los tests de lápiz y papel. Algunas de estas ventajas son: (1) menor número de preguntas para evaluar a un alumno, y (2) a igual numero de preguntas la evaluación adaptativa es mejor, supuestas las preguntas bien calibradas. SIETTE participa de estas ventajas ya que ofrece la posibilidad de realizar tanto test adaptativos como no adaptativos. Además, la implementación discreta de la TRI resulta computacionalmente eficiente y permite abordar en ciertos casos el problema de la multidimensionalidad.

Hay abiertas varias líneas de investigación relacionadas con SIETTE. El uso de múltiples dimensiones para la evaluación del nivel de conocimientos y su integración con el sistema MEDEA es una línea en sus comienzos. Otra línea muy interesante es la

utilización del mecanismo de preguntas y respuestas con fines tutoriales y no meramente evaluativos.

Actualmente se han desarrollado tests para diversas materias, relacionadas con las asignaturas que se imparten en la E.T.S.I. de Informática de la Universidad de Málaga. También se ha desarrollado un test de Piaget (Arroyo, Conejo, Guzmán y Wolf, 2001) para medir el desarrollo cognitivo en niños. Para el proyecto TREE, un proyecto de botánica para la identificación y clasificación de especies forestales, se ha desarrollado un test de reconocimiento de árboles (Ríos, Millán, Trella, Pérez de la Cruz Conejo, 1999). Se está desarrollando un nuevo test de Lógica y se han hecho pruebas de tests para otros dominios. La experiencia obtenida en el uso de la herramienta ha permitido mejorarla técnicamente, aunque el sistema actual (<http://www.lcc.uma.es/siette>) aún puede considerarse un prototipo.

REFERENCIAS

Arroyo, I., Conejo, R., Guzmán, E. y Woolf, B.P. (2001). An Adaptive Web-based Component for Cognitive Ability Estimation. En J.D. Moore, C. Luckhardt-Redfield, W. Lewis Johnson (eds.), Artificial Intelligent in Education: AI-ED in the Wired and Wireless Future (pp. 456-486). Amsterdam: IOS Press.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's mental ability. En Lord, F. M. y Novick, M.R. (eds.) Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Brusilovsky, P. y Miller, P. (1999). Web-based testing for distance education. En P. De Bra and J. Leggett (eds.) Proceedings of WebNet'99, World Conference of the WWW and Internet (pp. 149-154). Honolulu: AACE.

Conejo, R., Millán, E., Pérez de la Cruz, J. L. y Trella, M. (2000). An empirical approach to on-line learning in SIETTE. En Proceedings of the ITS 2000 (pp. 57-60). Montreal: Springer-Verlag.

Lord, F. M. y Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

McDonald, R.P. (1996). Normal-Ogive Multidimensional Model: En W. van der Linden y R. Hambleton, (eds.) Handbook of modern item response theory. New York: Springer.

Owen, R. J. (1975). A Bayesian sequential procedures for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association 70. (pp. 351-356).

Polson, M.C., Ricahrdson, J. (1998). Foundation of Intelligent Tutorial Systems. Hillsdale, N.J : Lawrence Earlbaum.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment test. Copenhagen: Danish Institute for Educational Research.

Reckase, M.D. (1996): A linear logistic multidimensional model for dichotomous items response data. En W. van der Linden y R. Hambleton (eds.) Handbook of modern item response theory. New York: Springer.

Ríos, A., Pérez de la Cruz, J.L. y Conejo, R. (1998). SIETTE: Intelligent Evaluation System using Test for TeleEducation. En 4th International Conference on Intelligent Tutoring System. ITS'98. Workshops papers. San Antonio, Texas, USA.

Ríos, A., Millán, E., Trella, M., Pérez de la Cruz, J. L. y Conejo, R. (1999). Internet Based Evaluation System. En Artificial Intelligence in Education AIED'99 (pp. 387-394). Le Mans.

Self, J. (1990). Theoretical foundation for Intelligent Tutoring Systems. AAAI/AI-ED.

Trella, M., Conejo, R. y Bueno, D. (2001). MEDEA, Una arquitectura basada en componentes para el desarrollo de sistemas tutores inteligentes en Internet. Actas de la Conferencia de la Asociación Española para la Inteligencia Artificial, CAEPIA/TTIA 2001 (2). (pp. 469-478). Gijón.