
Un modelo de evaluación cognitiva basado en Tests Adaptativos Informatizados para el diagnóstico en Sistemas Tutores Inteligentes

Eduardo Guzmán y Ricardo Conejo

Departamento de Lenguajes y Ciencias de la Computación. E.T.S.I.Informática
Universidad de Málaga, Campus de Teatinos, 29071 Málaga, Spain
guzman@lcc.uma.es

Resumen. En este trabajo se aplican técnicas de Inteligencia Artificial al ámbito de la educación. Se define un modelo de evaluación cognitiva basado en la Teoría de los Tests Adaptativos Informatizados, a través de la cual se generan tests adaptados a las características personales de cada alumno. Además, se define un modelo de respuesta basado en la Teoría de Respuesta al Ítem. Esta teoría probabilística se utiliza generalmente en los tests adaptativos, para estimar el conocimiento del alumno, para decidir qué preguntas se mostrarán en el test y para determinar el número de preguntas de cada test. El objetivo del modelo de evaluación cognitiva presentado es ser utilizado dentro de un Sistema Tutor Inteligente como módulo de diagnóstico del conocimiento del alumno. Asimismo, este modelo ha sido implementado en un sistema de evaluación a través de la Web, el sistema SIETTE.

1. Introducción

La evaluación educativa es un proceso en el que se realizan inferencias sobre lo que el alumno sabe, basándose en evidencias derivadas de observaciones sobre lo que éste dice o hace en determinadas situaciones (Pellegrino et al., 2001). Representa una parte fundamental del aprendizaje, ya que ofrece una medida de lo que ya ha sido aprendido (Patel et al., 1998), así como un medio eficaz para identificar lagunas en el conocimiento de los alumnos. La necesidad de disponer de mecanismos de evaluación efectivos es esencial dentro de cualquier proceso de instrucción. A través de la evaluación, es posible identificar lo que el alumno sabe, así como sus puntos fuertes y puntos débiles; observar su propio proceso de aprendizaje, y decidir en qué dirección dirigir ese proceso de aprendizaje (Gouli et al., 2001). En McCormack y Jones (1997) se esbozan las líneas que deben ser seguidas durante el proceso de evaluación: Evaluar el conocimiento del alumno antes o durante una lección, estimular que el alumno contemple el material que ha estudiado, ofrecer al alumno la posibilidad de revisar los conceptos que ha estudiado, y facilitar la posibilidad de que éste pueda indicar si ha entendido los conceptos estudiados con suficiente claridad.

Uno de los campos de aplicación de la *Inteligencia Artificial* (IA) dentro del ámbito de la educación, es precisamente la evaluación. Una de las principales ventajas de un sistema de evaluación inteligente es que puede contribuir a reducir el tiempo y el esfuerzo empleado por un profesor en llevar a cabo la tarea mecánica y repetitiva de evaluar numéricamente a un alumno. Asimismo, el uso de sistemas de evaluación inteligente permite suministrar al alumno un refuerzo, es decir, una guía, inferida por sus resultados en la evaluación, que permita orientar al alumno en su instrucción. Esto equivale por tanto a una tutorización individual (Patel et al., 1998).

2. Los Sistemas Tutores Inteligentes

Gracias a la aplicación de técnicas de IA a los sistemas educativos surgieron los *Sistemas Tutores Inteligentes* (STI). Además de en la IA, estos sistemas se apoyan en otras dos áreas de conocimiento: la Psicología Cognitiva y la Investigación Educativa. Estos sistemas nacen como un intento de complementar (o incluso de suplir) la cada vez más difícil tarea de proporcionar a cada alumno una instrucción personalizada. Una definición completa de un STI es la dada por Guardia (1997):

Un STI es un SEAC que utiliza técnicas de IA, principalmente para representar el conocimiento, y dirigir una estrategia de enseñanza; y que es capaz de comportarse como un experto, tanto en el dominio de conocimiento que enseña (mostrando al alumno cómo aplicar dicho conocimiento), como en el dominio pedagógico, donde es capaz de diagnosticar la situación en la que se encuentra el estudiante, y de acuerdo a ello, ofrecer una acción o solución que le permita progresar en el aprendizaje.

En definitiva, un STI es un sistema que debe responder a tres preguntas fundamentales: *¿Qué es lo que se enseña?* *¿A quién se enseña (Cuáles son sus características)?*, y por último, *¿Cómo se lleva a cabo esa enseñanza?* La aplicación de la IA queda reflejada, principalmente, en dos aspectos del desarrollo de un STI: (a) Deben proporcionar una enseñanza individualizada en función de las necesidades del alumno en cada momento. (b) El orden y plan de interacción entre alumno y sistema nunca debe estar predefinido.

Los componentes principales dentro de un STI son:

- *Modelo del dominio*: Corresponde al *qué* se enseña. Contiene el conocimiento sobre la materia que debe ser aprendida. Un modelo del dominio será más potente cuanto más abundancia de conocimiento tenga (Anderson, 1988).
- *Modelo del alumno*: El uso de modelos del alumno en STI surge como consecuencia del hecho de que estos sistemas deben trabajar con información

incompleta, y por regla con un alto grado de incertidumbre sobre los alumnos (Mayo y Mitrovic, 2001). Representa el a *quién* se enseña: lo que el alumno conoce y lo que no conoce del dominio anterior. La estructura que almacena el estado de conocimiento del alumno es propiamente el modelo del alumno. El proceso de inferencia de las características internas del alumno a partir de la observación de su comportamiento se denomina *diagnóstico del alumno* (Van-Lehn, 1988).

- *Modelo de instrucción*: Corresponde al *cómo* se enseña. Constituye, por tanto, las estrategias de enseñanza o estrategias tutoriales. Es decir, cómo el sistema debe mostrar el material educativo al alumno.
- *Interfaz*: A través de ella, se lleva a cabo la interacción hombre-máquina. Es necesario un esfuerzo adicional en el desarrollo de esta parte de la arquitectura, haciéndola intuitiva y transparente a los ojos del usuario alumno.

3. El diagnóstico del alumno en los STI

Durante la instrucción, el canal de comunicación que se establece entre un alumno y un STI es muy restrictivo. El STI únicamente es capaz de medir el conocimiento de forma directa, mediante la monitorización de la interacción con el alumno. Según Burns y Capps (1988), el diagnóstico del alumno es una "aventura" de alto riesgo. La importancia de un diagnóstico certero es vital para el buen funcionamiento de un STI, ya que las estrategias tutoriales se deciden en función de la información que el sistema tiene sobre el estado en el que se encuentra el conocimiento del alumno. Las características fundamentales del diagnóstico del alumno hacen referencia a: (a) todas aquellas características observables del alumno que son almacenadas en función de medidas específicas; (b) las características internas que debe ser inferidas en base a la información almacenada, y que son importantes para el aprendizaje; y (c) el método utilizado para extraer esa información a través de la monitorización y el seguimiento del alumno.

La presencia de *incertidumbre* es un factor importante que frecuentemente lleva a errores en el diagnóstico del alumno. Esta incertidumbre aparece como resultado, en parte, de los errores y aproximaciones durante el proceso de análisis de los datos medidos, o bien es debida a la naturaleza abstracta de la percepción humana y a la pérdida de información resultante de su cuantificación (Grigoriadou et al., 2002).

Desde el punto de vista de la IA, la principal demanda exigida a un sistema de diagnóstico del alumno es el desarrollo de un método fiable. Este método debería ser capaz de analizar con efectividad, en la forma en la que el profesor lo haría, medidas del comportamiento del alumno. A partir de estas medidas, debería hacer estimaciones sobre sus características internas, actualizando el modelo del alumno de acuerdo con esto. El problema está en que los sistemas de diagnóstico

desarrollados se fundamentan: o bien en heurísticos, o los paradigmas que proponen son poco viables desde el punto de vista práctico, o solamente aplicables a dominios muy concretos.

Asimismo, otro de los problemas del diagnóstico del alumno es la *determinación inicial del estado de conocimiento*. Idealmente, sería necesario utilizar un método de inferencia que permitiera, tras una breve interacción con el alumno, obtener una primera medida de su conocimiento sobre el dominio. El problema principal reside en que las técnicas que se utilizan para este fin, o bien requieren bastante tiempo, o bien los modelos del alumno inicialmente inferidos son muy imprecisos. Una posible solución a este problema de inicialización del modelo es el uso de *pretests*. Los pretests, son tests que se llevan a cabo antes de que el alumno comience su instrucción en un STI. La idea es que, inicialmente, el alumno sea sometido a un test que determine su nivel de conocimiento inicial.

4. Antecedentes

4.1. Los Test Adaptativos Informatizados

Habitualmente los tests que se realizan para evaluar a un conjunto de alumnos se hacen sobre papel y lápiz. Desde finales de la década de los 80, y debido a la proliferación del uso de PCs, empezaron a realizarse tests sobre soporte electrónico. De esta forma, surgieron los denominados *Tests Administrados por Computador* (TAC). Entre las ventajas de este tipo de tests está la posibilidad que ofrecen al alumno de realizar tests en cualquier momento y en cualquier lugar, además de la disponibilidad inmediata de los resultados. Este tipo de tests permiten a su vez la inclusión de nuevos formatos de ítems (denominación que reciben las preguntas en este ámbito) con, por ejemplo, elementos multimedia (imágenes, video, audio, etc...); nuevos tipos de ítems (Parshall et al., 2000); y quizás la posibilidad de realizar evaluaciones de ítems de respuestas abiertas (de redacción o de respuesta corta) (Wise, 1999).

Una versión más sofisticada de los TAC son los *Tests Adaptativos Informatizados* (TAI) (Olea et al., 1999; van der Linden y Glas, 2000). Wainer (1990) define este tipo de tests de la siguiente forma:

La idea fundamental de un Test Adaptativo es imitar de forma automática el comportamiento de un examinador. Esto es, si un examinador le presenta al alumno un ítem demasiado difícil para él, éste dará una respuesta errónea, y por lo tanto, la siguiente vez, el examinador presentará una pregunta algo más fácil, y viceversa

En general, en los TAI, los ítems se muestran de uno en uno, y la presentación de cada ítem, así como la decisión de finalizar el test y la evaluación del alumno

se toman dinámicamente, basándose en las respuestas del alumno. El objetivo final de un test adaptativo es diagnosticar el nivel de conocimiento del alumno. En términos más precisos, un TAI es un algoritmo iterativo que comienza con una estimación inicial del conocimiento del alumno y continúa con los siguientes pasos (el proceso se ilustra en la Figura 1):

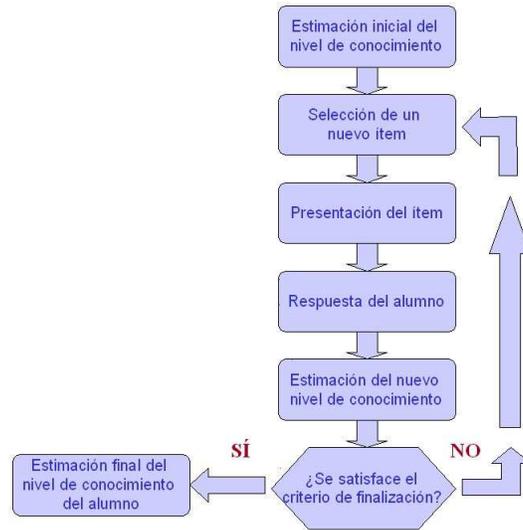


Figura 1. Diagrama de flujos de un test adaptativo (adaptado de Olea y Ponsoda (2001))

1. Todos los ítems del banco de ítems (que no han sido administrados todavía), son analizados para determinar cuál de ellos contribuye en mayor medida a una mejor estimación del conocimiento del alumno. Para ello, se utiliza el nivel de conocimiento estimado del alumno hasta ese momento.
2. El ítem es presentado al alumno.
3. En función de la respuesta dada por el alumno, se estima el nuevo nivel de conocimiento de éste.
4. Los pasos del 1 al 3 se repiten hasta que el criterio de finalización del test se satisfaga.

Los criterios para la selección del ítem que debe mostrarse al alumno en cada momento, la decisión de finalizar el test, así como la estimación del conocimiento del alumno se basan en procedimientos bien fundamentados. Estos procedimientos pueden ser controlados por medio de parámetros, con los que es posible definir la precisión requerida. La selección de ítems y el criterio de parada del test son, por tanto, adaptativos. El número de ítems de un TAI no es fijo, y a cada examinando se le mostrará una secuencia diferente de ítems, e incluso diferentes ítems.

Los elementos principales a la hora de desarrollar un TAI son:

- Un *modelo de respuesta* asociado a los ítems: Este modelo describe el comportamiento probabilístico del alumno cuando responde a un ítem en función de su nivel de conocimiento. Cuando se mide el nivel de conocimiento, el resultado obtenido debe ser independiente de la herramienta utilizada para llevar a cabo la medición. En la mayoría de los casos, se utiliza como modelo de respuesta un modelo basado en la *Teoría de Respuesta al Ítem*.
- Un *banco de ítems*: Es uno de los elementos más importantes de un TAI, ya que cuanto mejor sea la calidad del banco, los tests adaptativos serán más precisos. En cuanto al número de ítems que debe tener un buen banco, muchas son las recomendaciones, aunque en general, se considera que como mínimo debe tener 100 ítems.
- El *nivel de conocimiento inicial*: Es muy importante llevar a cabo una buena estimación inicial del nivel de conocimiento del alumno, ya que ésta determinará la duración final del test. Se pueden utilizar diferentes criterios: la media de los niveles de conocimiento de los alumnos que hayan realizado el test con anterioridad, creación de un perfil y utilizar la media de los alumnos que sean clasificados con ese perfil (Thissen y Mislevy, 1990), etc.
- El *criterio de selección* de ítems: El mecanismo de adaptación propio de los TAI se encarga de seleccionar el siguiente ítem que debe mostrarse al alumno en cada momento. Esta decisión se toma en función del nivel de conocimiento estimado (obtenido a partir de los ítems propuestos al alumno con anterioridad). Seleccionar el mejor ítem, es decir, elegir aquel ítem más informativo desde el punto de vista de la estimación, mejora la precisión del test, y reduce el número de ítems de éste.
- El *criterio de finalización*: Cuando se satisface este criterio, el test debe finalizar. Diferentes criterios pueden utilizarse en función del objetivo final del test. El criterio más apropiado desde el punto de vista de la adaptación es aquél que finaliza el test cuando la precisión en la estimación del nivel de conocimiento del alumno es menor o igual que un cierto umbral predefinido. Otros criterios no adaptativos utilizados son: llegar al máximo número de ítems permitidos en un test, haber consumido el tiempo requerido para completar el test, etc.

Diversos autores (Mislevy y Almond, 1997; van der Linden y Glas, 2000) han indicado las ventajas que aportan los TAI frente a los tests tradicionales de papel y lápiz:

- La ventaja principal es que son más eficientes, es decir, el número de ítems requerido para estimar el nivel de conocimiento del alumno es menor, y como consecuencia también el tiempo empleado en su realización.
- Las estimaciones realizadas por un TAI son más precisas y no están sesgadas.

- Los TAI se ajustan a las características personales de cada examinando. Esto conlleva también una mejora de la motivación del alumno, ya que las preguntas son adecuadas al nivel de conocimiento del alumno. Por este motivo, el alumno no se ve frustrado porque las preguntas que aparecen en el test son muy difíciles para él, ni por el contrario se aburre con preguntas que le parecen excesivamente fáciles.
- Se reduce la ansiedad del alumno durante la realización del test, ya que cada alumno dispone exactamente del tiempo que él mismo requiere para la finalización del mismo.
- Los tests son más seguros, ya que cada a cada alumno le es suministrado un conjunto y un número diferente de ítems.

Todas estas razones hacen que, por regla general, entre un TAI y el mismo test en formato de papel y lápiz, los alumnos prefieren la primera opción (van der Linden y Glas, 2000).

Por el contrario, los TAI presentan algunas desventajas. Una de las más importantes es la seguridad, ya que los examinandos podrían memorizar los ítems de un test y dar esta información a futuros examinandos. Es por tanto necesario, disponer de un gran banco de ítems, así como de técnicas para el control de la exposición de los ítems y para la detección de ítems comprometidos. Por otro lado, los ítems deben estar bien calibrados. Sin embargo, para llevar a cabo esta tarea, es necesario disponer de un número considerable de sesiones de tests realizadas por alumnos (de forma no adaptativa), y esto no siempre es posible. Además, las técnicas de calibración de ítems, por lo general suelen ser computacionalmente costosas. Asimismo, algunos alumnos pueden sentirse discriminados frente a otros por el mero hecho de haber respondido a diferentes ítems y tests con diferente número de ítems, que el resto de sus compañeros. Finalmente, mediante el uso de tests adaptativos sólo es posible, en cada test, inferir un único nivel de conocimiento del alumno. Esto implica que, al final del test, se obtendrá una única calificación global.

5. La Teoría de Respuesta al Ítem

La *Teoría de Respuesta al Ítem* (TRI) (van der Linden y Hambleton, 1997; Embretson y Reise, 2000) se encarga de modelar el conjunto de procesos relacionados con la respuesta de un alumno a un ítem (Thissen, 1993). La TRI se apoya en dos principios fundamentales (Hambleton et al., 1991): los resultados obtenidos por un alumno en un test pueden ser explicados mediante un factor denominado *nivel de conocimiento* o *rasgo latente*, que puede ser medido mediante valores numéricos inicialmente desconocidos. Además, la relación entre los resultados del alumno en el test y la respuestas del alumno a un cierto ítem puede ser descrita

mediante una función monótona creciente denominada *Curva Característica del Ítem* (CCI). La CCI es el elemento básico de la TRI, el resto de elementos de esta teoría dependen de la CCI. Representa la probabilidad condicional de que un alumno con un cierto nivel de conocimiento (θ) responda correctamente a un ítem. Esta función debe ser conocida para cada ítem del test (se obtiene mediante el proceso de calibración), y se expresa mediante una función: $f : (-\infty, \infty) \rightarrow [0, 1]$ en la que el rango de valores fluctúa en la escala en la que se mide θ .

Existen multitud de modelos basados en la TRI. En todos ellos, la probabilidad de una respuesta correcta depende de la aptitud del examinando θ , y de los parámetros que caracterizan al ítem (Martínez Arias, 1995). El modelo más comúnmente utilizado es un modelo dicotómico basado en una función logística de tres parámetros (Birnbaum, 1968).

$$P(u_i = 1|\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-1.7a_i(\theta - b_i)}} \quad (1)$$

θ normalmente asume valores reales entre -4 y 4. Los tres parámetros restantes son: **(a) Factor de discriminación** (a_i): Es un valor proporcional a la pendiente de la curva. Un valor alto indica que la probabilidad de que un alumno con un rasgo latente estimado mayor que la dificultad del ítem acierte es mayor. **(b) Dificultad** (b_i): Corresponde al valor de θ para el cual la probabilidad de responder correctamente al ítem es la misma que de responder de forma incorrecta. **(c) Factor de adivinanza** (c_i): Es la probabilidad de que un alumno sin conocimiento ninguno responda correctamente al ítem, cuando selecciona la respuesta de forma completamente aleatoria.

La TRI se aplica a la teoría de los TAI como modelo de respuesta para, entre otras cosas, inferir el nivel de conocimiento del alumno. Uno de los métodos basados en la TRI para la estimación del conocimiento del alumno es el *método de estimación bayesiana* de Owen (1969, 1975). En este método, se calcula, aplicando el teorema de Bayes, la distribución de probabilidad a posteriori del conocimiento del alumno ($P(\theta|u)$). Es decir,

$$P(\theta|u) = \prod_{i=1}^n P_i(u_i = 1|\theta)^{u_i} (1 - P_i(u_i = 1|\theta))^{1-u_i} P(\theta) \quad (2)$$

donde $u = u_1, u_2, \dots, u_n$ es el patrón de respuesta de un alumno a los n ítems del test, y $u_i = 1$, indica que el alumno ha respondido correctamente al ítem preguntado en la posición i -ésima. $P(\theta)$ representa la distribución de probabilidades del conocimiento del alumno a priori.

La aplicación del método bayesiano proporciona como resultado una función de distribución del conocimiento del alumno. Existen dos modelos bayesianos en función de cómo se estime el nivel de conocimiento del alumno a partir de esa distribución. Cuando se utiliza la moda de la distribución a posteriori como

nivel de conocimiento estimado, el modelo se denomina *Estimación bayesiana del Máximo a Posteriori* (MAP). Si el valor que se toma es la media o esperanza matemática, se denomina entonces *Estimación bayesiana de la Esperanza a Posteriori* (EAP).

Además de para estimar el nivel de conocimiento del alumno, la TRI se utiliza en los TAI para determinar el ítem que debe mostrarse al alumno en cada momento, es decir, como criterio de selección de ítems; y para determinar si el test debe finalizar.

6. Planteamiento

En las secciones anteriores se han puesto de manifiesto las ventajas que suponen el uso de tests adaptativos. Este tipo de tests podrían utilizarse para el diagnóstico en STI, ya que proporcionan métodos bien fundamentados. Aún así, el uso de tests adaptativos no parece solventar el problema de la estimación inicial del conocimiento del alumno, ya que sólo son capaces de inferir un único valor numérico que representa el nivel de conocimiento del alumno. El problema se agrava aún más, ya que, aunque estos tests permiten valorar globalmente el nivel de conocimiento del alumno, los mecanismos de selección de ítems *per se* no son capaces de asegurar que la selección de ítems sea balanceada. Es decir, cuando se administra un test en el que se evalúan diversos conceptos simultáneamente, no se puede garantizar que los ítems que aparecen constituyan una distribución homogénea de ítems por concepto.

Los *modelos de evaluación cognitiva*, en general, persiguen un objetivo más ambicioso que simplemente clasificar linealmente a los alumnos tal y como hacen los modelos de respuesta basados en la TRI. Los *modelos de evaluación cognitiva* proporcionan una lista de habilidades u otros atributos cognitivos que el alumno puede poseer, en función de las evidencias proporcionadas por las tareas que éste realiza (Junker y Sijtsma, 2001a). En resumen, estos modelos intentan tener en cuenta y medir los procesos cognitivos y las estrategias seguidas para responder a ítems dicotómicos o politómicos (Junker y Sijtsma, 2001b).

Tomando como base la teoría de los TAI, se propone un modelo de evaluación cognitiva mediante tests adaptativos. El modelo propuesto intenta aplicar la teoría de los TAI al campo de los STI. El objetivo principal es dotar a los STI de un mecanismo para la inferencia del estado de conocimiento del alumno. Este mecanismo deberá garantizar que los resultados obtenidos en la evaluación son verdaderos indicadores del conocimiento del alumno.

En esta sección se procederá a describir este modelo. Inicialmente se describirá el modelo de respuesta (basado en la TRI) que se ha desarrollado con este fin, para posteriormente estudiar las características del modelo de evaluación cognitiva.

6.1. Un modelo de respuesta discreto basado en la TRI

La gran mayoría de los modelos de respuesta basados en la TRI son modelos continuos. Es decir, el nivel de conocimiento se suele medir utilizando como rango los números reales. Este tipo de modelos son más precisos debido a la precisión inherente a la escala de medición. A pesar de tener esta característica, desde el punto de vista puramente práctico, son más difíciles de utilizar. Por esta razón, en este modelo, se ha preferido utilizar una discretización, que permite representar las curvas características y la distribución del conocimiento estimado del alumno como vectores. Los valores que puede tomar el nivel de conocimiento son números naturales entre 0 y el número de niveles de conocimiento menos 1. Sea K el número de niveles de conocimiento, 0 representa la ausencia total de conocimiento, y $K - 1$ el conocimiento absoluto. El número de niveles de conocimiento K lo define el profesor según la granularidad deseada en la evaluación. Como resultado, las operaciones de estimación del conocimiento del alumno, selección del ítem que debe ser mostrado y de finalización del test se llevan a cabo más eficientemente.

En general, los modelos TRI aplicados a los TAI son modelos dicotómicos. En los modelos dicotómicos los ítems son evaluados en dos categorías: correctos o incorrectos. Por el contrario, los modelos politómicos permiten clasificar la respuesta del alumno en diversas categorías. El objetivo con el que fueron inicialmente concebidos era modelar ítems de opción múltiple, de forma que se constatará la influencia de la respuesta seleccionada por el alumno, más que en si ésta ha sido correcta o incorrecta. De esta forma, cada respuesta es tratada independientemente. Diversos investigadores (Embretson y Reise, 2000) han puesto de manifiesto que los modelos de la TRI politómicos son más informativos que los dicotómicos. El ser más informativos conlleva además la ventaja de que se pueden llegar a obtener estimaciones más precisas con un número más reducido de ítems (Hontangas et al., 2000).

En general, en los modelos de respuesta politómicos además de la CCI, se define una curva característica por cada respuestas posible al ítem. Estas curvas se denominarán *Curvas Características de Respuesta* (CCR). Las CCR, análogamente a las CCI, sirven para caracterizar cada respuesta del ítem, y al igual que las CCI deben ser estimadas. Las CCR pueden ser descrita como la porción de una población de alumnos con un determinado nivel de conocimiento que, cuando se les administra el ítem, seleccionan esa respuesta.

A pesar de su objetivo inicial, en la actualidad estos modelos se aplican principalmente a ítems en los que las respuestas están ordenadas en categorías, por ejemplo, en test de personalidad. El motivo principal por el que este tipo de modelos han sido descartados dentro de los tests adaptativos, es que la calibración de sus ítems es mucho más costosa que en los modelos dicotómicos. Si en estos

últimos es necesario disponer de una muestra poblacional de al menos 100 examinados que realizaron el test de forma no adaptativo, en el caso politómico, el tamaño de la muestra recomendado asciende a un mínimo de 3000 individuos.

El modelo presentado es también un modelo politómico, en el que cada respuesta tiene asociada su CCR. Para solventar el problema de la calibración, se utiliza un método de estadístico basado en el *suavizado núcleo* (Ramsay, 1991). La idea principal del *suavizado núcleo* es obtener una estimación no paramétrica de las curvas características tomando una media ponderada en cada punto de evaluación, en la que los pesos se determinan mediante una función núcleo (Ferrando, 2004). Con esta técnica, la muestra poblacional necesaria para la calibración es considerablemente inferior a la requerida en los modelos politómicos. Asimismo el proceso de calibración es mucho más eficiente, y los resultados obtenidos son equivalentes a la calibración mediante otros algoritmos (Douglas, 1999).

6.2. Un modelo para el diagnóstico en STI basado en Tests Adaptativos

La arquitectura del modelo de diagnóstico propuesto (Figura 2) combina los elementos principales de un sistema para la generación de TAIs, con los elementos necesarios que debe tener un módulo de diagnóstico para STIs. En ella se pueden distinguir los siguientes componentes:

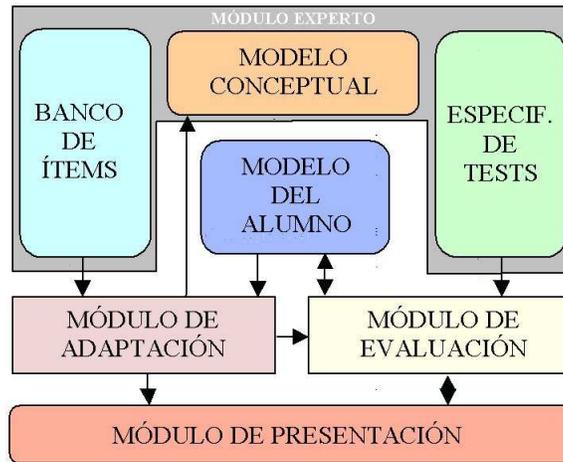


Figura 2. Arquitectura del modelo de diagnóstico cognitivo

- *El módulo experto:* Contiene todo el conocimiento aportado por el experto (en este caso, el profesor). Se compone a su vez, tres modelos: Un *modelo (o mapa) conceptual*, similar al modelo del dominio de un STI, pero en el

que únicamente se almacenan los conceptos y las relaciones entre éstos; un *banco de ítems*, que contiene los ítems para llevar a cabo el diagnóstico; y por último, un conjunto de *especificaciones de tests*, que no son más que guías de evaluación definidas por los profesores en las que se expresan los conceptos que van a ser evaluados y los parámetros que caracterizan esas sesiones de evaluación.

- *El modelo del alumno*: Al igual que en la arquitectura de un STI, este modelo se encarga de almacenar toda la información disponible sobre el alumno. Es, por tanto, la representación del alumno dentro del modelo de diagnóstico.
- *El módulo de adaptación*: Este módulo, a partir de la información almacenada en el modelo del alumno, a partir de los ítems disponibles en el banco de ítems, y en función de los parámetros de configuración del test, se encarga de seleccionar qué ítem debe ser mostrado al alumno en cada momento de la sesión de diagnóstico. Es el encargado asimismo de decidir cuándo debe finalizar la sesión. Para más información sobre el funcionamiento de este módulo véase (Guzmán y Conejo, 2002).
- *El módulo de evaluación*: Durante una sesión de evaluación, y una vez que el alumno responde a un ítem, este módulo lleva a cabo la función de inferir el nuevo conocimiento del alumno en el concepto que evalúa el ítem que acaba de ser mostrado, y actualiza el modelo del alumno.
- *El módulo de presentación*: Es el encargado de interactuar con el alumno. Por un lado muestra a éste el ítem que ha sido seleccionado en el módulo de selección. Una vez que el alumno ha respondido al ítem, este módulo se encarga de enviar la respuesta al módulo de evaluación.

6.2.1. El módulo experto

El modelo o módulo experto en los STI contiene el conocimiento que tiene el experto sobre el dominio. Este conocimiento se expresa mediante una red de conceptos que permite representar el conjunto de conceptos de que deben estudiar los alumnos durante el proceso de aprendizaje. Para poder llevar a cabo diagnósticos del alumno, es necesario disponer de los instrumentos necesarios. En nuestro caso, estos instrumentos son los ítems y las especificaciones de los tests. De esta forma, el modelo experto se compone a su vez de tres partes: el *modelo conceptual*, el *banco de ítems* y las especificaciones de los *tests*.

- *El modelo conceptual*: En los criterios tradicionales de enseñanza, las materias (cursos o asignaturas) suelen estructurarse en partes. Cada una de estas partes (conceptos o temas) puede estructurarse a su vez en subpartes (subconceptos o subtemas), y así sucesivamente. De esta forma, se obtienen jerarquías con granularidad variable en forma de árbol. La granularidad de un dominio hace referencia al nivel de detalle o perspectiva desde la que los conceptos pueden

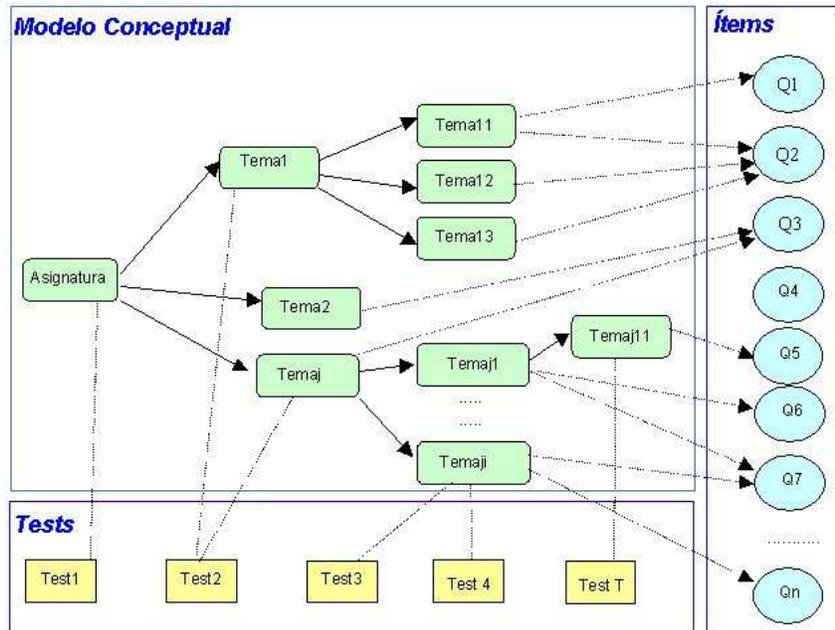


Figura 3. Relación entre los elementos del modelo experto

ser vistos (Greer y McCalla, 1994). En este caso, el grado de granularidad lo determina el profesor-experto, que es el encargado a su vez del diseño de la jerarquía arbórea. Estas jerarquías son denominadas *currículos*, y suelen representarse mediante una *red semántica*. Una *red semántica* no es más que un grafo acíclico donde los nodos son las partes en las que se dividen las materias, y los arcos representan relaciones entre esas partes. El nivel de granularidad vendrá determinado por la precisión requerida en la evaluación. Asimismo se asumirá que los conceptos de un nivel de la jerarquía están relacionados con el nivel inmediatamente anterior mediante relaciones de agregación, es decir, relaciones "es-un".

En la Figura 3 se muestra una representación gráfica de esta estructuración del dominio. Como se puede apreciar, la asignatura puede ser considerada como un concepto estructurado en otros subconceptos, y por tanto es posible obtener una valoración global del conocimiento del alumno en ella. Según Reye (2002) los conceptos son elementos del currículo que representan piezas de conocimiento o habilidades cognitivas que el alumno podría adquirir. De esta forma, consideramos los conceptos como elementos susceptibles de ser evaluados. Los nodos finales (nodos hoja) del modelo corresponden a conceptos único o a un conjunto de conceptos indiscernibles de cara a la evaluación.

- *El banco de ítems:* Los ítems se utilizan como medio de diagnóstico del conocimiento del alumno en los conceptos del modelo conceptual. Cada ítem permitirá evaluar un único concepto, y se dirá que un ítem *evalúa directamente* (o proporciona una evidencia directa sobre el conocimiento del alumno en) un concepto, cuando las respuestas a ese ítem proporcionan una evidencia sobre el conocimiento del alumno en el concepto. Por ejemplo, en la figura 3 los ítems Q_1 y Q_2 evalúan directamente al concepto $Tema_{11}$.

En virtud de la relación que se establece entre los conceptos que componen el modelo conceptual, si un ítem Q evalúa directamente un concepto C_{ijk} , ese ítem también suministra alguna evidencia sobre el conocimiento del alumno en el concepto que precede al concepto C_{ijk} , es decir, en el concepto C_{ij} . Aplicando el mismo razonamiento, el ítem también aporta evidencias sobre el conocimiento del alumno en el concepto padre de C_{ij} , esto es, en el concepto C_i . Por último, se ha mencionado que la asignatura completa puede ser considerada una agregación de sus conceptos hijos. Por este motivo, el ítem Q también proporciona evidencias sobre el conocimiento del alumno global de la asignatura. Los ítems podrán evaluar directamente conceptos hoja (esto es, conceptos que no representan agregaciones de otros conceptos), conceptos intermedios, o incluso al concepto raíz de la jerarquía (la propia asignatura). Sin embargo, cuando un ítem evalúa directamente a un concepto intermedio, no podemos inferir directamente que también evalúe a sus descendientes.

Según la TRI, cada ítem permite evaluar un rasgo latente, y esta relación viene determinada por la CCI, en caso de un modelo de respuesta dicotómico, o bien por las CCR, cuando se aplica un modelo politómico. Extrapolando esta característica al modelo propuesto, si se considera cada concepto como un rasgo latente susceptible de ser evaluado, cada ítem deberá tener asociado un conjunto de CCR por cada concepto que evalúe (en el caso de un modelo dicotómico, una CCI por cada concepto). Sea n la profundidad del concepto evaluado directamente por un ítem Q , este ítem deberá almacenar un número de CCR igual al producto del número de respuestas multiplicado por n , es decir, un conjunto de CCR (una por cada respuesta) por cada concepto que represente una agregación directa o indirecta del concepto evaluado directamente. El número de ítems de cada concepto debe estar balanceado, es decir, todos los conceptos deben tener un número similar de ítems. Esto contribuirá a mejorar la precisión de los criterios de selección y finalización adaptativos. Es responsabilidad del profesor que construye el módulo experto de una determinada asignatura, asegurar que sus conceptos tienen un número similar de ítems de tal forma que el diagnóstico inferido por el modelo sea llevado a cabo de forma correcta.

- *Los tests:* Para poder evaluar a los alumnos, el modelo propuesto permite definir tests. Un test en sí no es más que una sesión de evaluación. Su objetivo

es obtener una estimación del nivel de conocimiento del alumno sobre uno o varios de los conceptos del currículo.

Estableciendo una analogía con los STI, Chua Abdullah (2003) señala cuáles son las preguntas fundamentales a las que debe responder un sistema de evaluación, esto es, los elementos que deben ser tenidos en cuenta para poder realizar evaluaciones efectivas. Aunque ella únicamente indica tres consideraciones, desde el punto de vista de la evaluación adaptativa, es necesario añadir un cuarto punto. En resumen, los elementos indispensables que deben ser conocidos a priori para evaluar correctamente son enumerados a continuación:

1. *Qué evaluar*, esto es, qué conceptos del modelo conceptual van a ser evaluados.
2. *A quién evaluar*, es decir, cómo es el alumno al que se va a evaluar. Esta información estará contenida en su modelo del alumno.
3. *Cómo evaluar*, es decir, por un lado, qué criterio de evaluación va a ser utilizado, esto es, cómo va a ser inferida la calificación del alumno a partir de su actuación en el test; el nivel de granularidad de la evaluación: en cuantos niveles de conocimiento va a ser evaluado el alumno, de cuantos conceptos, si se trata de una evaluación completa o por el contrario de una evaluación agregada; y por último, cómo van a ser secuenciados los elementos de evaluación (los ítems), es decir, qué criterio de selección de ítems se va a utilizar.
4. *Cuándo finalizar la evaluación*, ya que en criterios de evaluación adaptativos es necesario determinar a priori cuando se va a considerar que la estimación del conocimiento del alumno es suficientemente precisa.

En el modelo propuesto, este conjunto de consideración son representadas mediante parámetros de configuración del test. Los tests se definen, principalmente, en función del concepto o conceptos que se deseen evaluar.

En un test se podrán evaluar directamente y de forma simultánea tantos conceptos como se desee, con la única restricción de que entre esos conceptos no debe existir ninguna relación de agregación, es decir, no debe existir ningún camino entre los conceptos en la red semántica del modelo conceptual. Esta restricción debe satisfacerse para todos los conceptos del test dos a dos. Es necesario reseñar también que no existe una relación directa entre tests e ítems. La relación entre ambos se establece a través de los conceptos del modelo conceptual.

Asimismo, en cada tests, debe indicarse el criterio de selección de ítems que se utilizará, cómo se inicializará el modelo del alumno, cuándo debe finalizar el test, cómo se evaluará al alumno, así como el número de niveles de conocimiento en que éste será evaluado.

Por último, cuando se determina que un concepto va a ser evaluado en un test, hay que decidir también la granularidad del proceso de evaluación. En

este sentido, pueden utilizarse dos modos de evaluación: *agregada*, en el caso en que sólo se requiera la evaluación de ese nodo del currículo; o *completa*, que indica que es necesaria una evaluación exhaustiva de todos los nodos del subárbol cuya raíz es este tema.

6.2.2. El módulo de evaluación

Supóngase que el profesor define un test de más de un concepto, por ejemplo, el test $Test_2$ (Figura 3). Como resultado final, se obtendrá una estimación del nivel de conocimiento sobre el concepto $Tema_1$, y otra estimación del conocimiento sobre el concepto $Tema_j$. Los ítems que podrían formar parte del test son todos aquéllos que evalúan esos conceptos, o cualquiera de sus conceptos descendientes. Si se seleccionara por ejemplo, el ítem Q_5 , éste proporciona evidencias sobre el conocimiento del alumno no sólo en el concepto $Tema_j$, sino también en los conceptos $Tema_{j1}$ y $Tema_{j11}$. Como consecuencia, en el mismo test, como efecto colateral, se estimará el nivel de conocimiento del alumno en estos conceptos. Al final del test, se obtiene una estimación por cada uno de los conceptos sobre los que el profesor definió el test, y una por cada uno de los conceptos que descienden de éstos.

La estimación del conocimiento del alumno tras responder a un ítem, se realiza aplicando el método bayesiano anteriormente presentado, pero adaptado al modelo de respuesta politómico que se ha definido. Por este motivo, la ecuación 2 ha sido modificada y reducida al sumatorio de los productos entre la CCR de la respuesta seleccionada en cada ítem, por la estimación inicial. Para más información véase (Guzmán y Conejo, 2004a).

7. Implementación: El sistema SIETTE

Este modelo ha sido implementado y probado, dando como resultado el sistema SIETTE (Conejo et al., 2004; Guzmán y Conejo, 2004b). SIETTE (*Sistema Inteligente de Evaluación mediante Tests para TeleEducación*) permite la generación de tests y su elicitación a través de interfaces Web. Este sistema puede funcionar bien sea como una herramienta de evaluación independiente, o formando parte de otros sistemas web educativos, especialmente en sistemas de enseñanza adaptativos en los que SIETTE puede desempeñar el rol de sistema de diagnóstico del conocimiento del alumno.

Este sistema es multilingüe, que actualmente está disponible en español y en inglés, aunque es posible traducirlo fácilmente a otros idiomas gracias a la generación dinámica de sus interfaces. Permite a los profesor incluir diversos tipos de ítems, incluyendo además una biblioteca para la construcción de ejercicios y un mecanismo para la generación automática de ítems a partir de plantillas.

La arquitectura de SIETTE comprende dos partes principales. Por un lado, el *espacio de trabajo de los alumnos*, y por otro lado, el *entorno de edición*. El *espacio de trabajo de los alumnos* es el lugar donde los alumnos pueden realizar tests. El componente principal de esta parte es el generador de tests, que se encarga de construir tests para los alumnos. Dos interfaces diferentes pueden ser utilizadas para acceder a los tests generados:

- El *aula virtual*: Es donde los alumnos realizan tests para autoevaluarse, y donde los profesores suministran tests a los alumnos para evaluar su nivel de conocimiento.
- La *interfaz para las conexiones externas*: A través de esta interfaz SIETTE puede funcionar como herramienta de diagnóstico en otros sistemas educativos hipermedia adaptativos.

El *entorno de edición* es un conjunto de herramientas creadas para los profesores. Estas herramientas permiten la creación y actualización de contenidos, así como el análisis de las sesiones de tests realizadas por los alumnos. Este entorno se compone de las siguientes herramientas:

- El *editor de tests*: Por medio de esta herramienta se actualizan los contenidos de la base de conocimiento.
- La *herramienta de análisis de resultados*: Esta herramienta permite analizar las sesiones de tests realizadas por los alumnos.
- La *herramienta para la calibración de ítems*: A través de esta herramienta, y tomando como entrada las sesiones de alumnos que han realizado tests de forma no adaptativa, se pueden calibrar las curvas características de los ítems.

El sistema es accesible a través de la dirección <http://www.lcc.uma.es/siette>.

8. Conclusiones

Se ha definido un modelo de respuesta basado en la TRI. La discretización facilita el cálculo de las estimaciones del conocimiento del alumno. El proceso utilizado para la calibración de los ítems se basa en la aplicación de técnicas estadísticas de suavizado núcleo. Como resultado, se obtiene un método de calibración, que a diferencia de los métodos comúnmente utilizados, requiere una muestra inicial menor de tests realizados de forma no adaptativa.

Utilizando el modelo de respuesta anterior, se ha definido un modelo de diagnóstico cognitivo para STIs en el que se adapta la TAI para evaluar simultáneamente diversos conceptos. Se ha definido un modelo conceptual para sistemas de diagnóstico que combina una jerarquía de conceptos con un banco de ítems y especificaciones de tests. Posee un mecanismo de actualización del

conocimiento en múltiples conceptos en virtud de las relaciones que se pueden establecer entre los nodos del modelo conceptual del currículo.

El modelo de diagnóstico ha sido implementado en una herramienta para la Web: el sistema SIETTE. Es un sistema que puede funcionar de forma autónoma para autoevaluación o evaluación; o incluso puede funcionar como módulo de diagnóstico dentro de un STI sobre la web, gracias a un protocolo que permite, de forma simple, que ambos sistemas interactúen. Ha sido integrado en sistemas como TAPLI (Millán et al., 2003), y actualmente está siendo integrado en el sistema LeActiveMath (Melis et al., 2001) como parte de un proyecto europea del Programa Marco. Permite no sólo la realización de tests adaptativos, sino también la realización de tests convencionales, facilitando de esta forma el soporte necesario para poder llevar a cabo la calibración de los ítems. Ha sido utilizado en la práctica en la E.T.S.I. Informática en diversas asignaturas como medio de evaluación.

Referencias

- Anderson, J. (1988). The expert module. En *Foundations of intelligent tutoring systems* (pp. 21–54). Lawrence Erlbaum Associates, Inc.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's mental ability. En *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Burns, H. y Capps, C. (1988). Foundations of intelligent tutoring systems: An introduction. En M. Polson y J. Richardson (Eds.), *Foundations of intelligent tutoring systems*. Lawrence Erlbaum Associates, Inc.
- Chua Abdullah, S. (2003). *Student modelling by adaptive testing - a knowledge-based approach*. Tesis doctoral no publicada, University of Kent, Canterbury.
- Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez de la Cruz, J. L. y Ríos, A. (2004). Siette: a web-based tool for adaptive testing. *Journal of Artificial Intelligence in Education*, 14, 29–61.
- Douglas, J. (1999). *Asymptotic identifiability of nonparametric item response models* (Technical Report No. 142). University of Wisconsin. Department of Biostatistics and Medical Informatics.
- Embretson, S. y Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Ferrando, P. (2004). Kernel smoothing estimation of item characteristic functions of continuous personality items: An empirical comparison with the linear and the continuous-response models. *Applied Psychological Measurement*, 28(2), 95–109.
- Gouli, E., Kornilakis, H., Papanikolaou, K. y Grigoriadou, M. (2001). Adaptive assessment improving interaction in an educational hypermedia system. En

- Human computers interaction. panhellenic conference with international participation* (pp. 217–222).
- Greer, J. y McCalla, G. (1994). Granularity-based reasoning and belief revision in student models. En J. Greer y G. McCalla (Eds.), *Student modelling: The key to individualized knowledge-based instruction* (Vol. 125, pp. 39–62). New York: Springer Verlag.
- Grigoriadou, M., Kornilakis, H., Papanikolaou, K. y Magoulas, G. (2002). Fuzzy inference for student diagnosis in adaptive educational hypermedia. En *Methods and applications of artificial intelligence. lecture notes in artificial intelligence* (Vol. 2308). New York: Springer Verlag.
- Guardia, B. (1997). *Asesores inteligentes para apoyar al proceso de enseñanza de lenguajes de programación*. Tesis doctoral no publicada, Instituto Tecnológico y de Estudios Superiores de Monterrey, Universidad de Ciudad de México.
- Guzmán, E. y Conejo, R. (2002). Simultaneous evaluation of multiple topics in siette. En S. Cerri, G. Gouardères y F. Paraguacu (Eds.), *Proceedings of the 6th international conference on intelligent tutoring systems (its 2002). lecture notes in computer science* (pp. 739–748). New York: Springer Verlag.
- Guzmán, E. y Conejo, R. (2004a). A model for student knowledge diagnosis through adaptive testing. En J. C. Lester, R. M. Vicari y F. Paraguaçu (Eds.), *Proceedings of the 7th international conference on intelligent tutoring systems (its 2004). lecture notes in computer science* (pp. 12–21). New York: Springer Verlag.
- Guzmán, E. y Conejo, R. (2004b). A brief introduction to the new architecture of siette. En P. De Bra y W. Nejdl (Eds.), *Proceedings of the iiith international conference on adaptive hypermedia and adaptive web-based systems(ah 2004). lecture notes in computer science* (pp. 405–408). New York: Springer Verlag.
- Hambleton, R. K., Swaminathan, J. y Rogers, H. (1991). *Fundamentals of item response theory*. Sage publications.
- Hontangas, P., Ponsoda, V., Olea, J. y Abad, F. (2000). Los test adaptativos informatizados en la frontera del siglo xxi: una revisión. *Metodología de las Ciencias del Comportamiento*, 2(2), 183–216.
- Junker, B. y Sijtsma, K. (2001a). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258–272.
- Junker, B. y Sijtsma, K. (2001b). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement*, 25(3), 211–220.

- Martínez Arias, R. (1995). *Psicometría: Teoría de los test psicológicos y educativos*. Madrid: Síntesis, S.A.
- Mayo, M. y Mitrovic, A. (2001). Optimising its behaviour with bayesian networks and decision theory. *International Journal of Artificial Intelligence in Education*, 12, 124–153.
- McCormack, C. y Jones, D. (1997). *Building a web-based education system*. Wiley Computer Publishing.
- Melis, E., Andres, E., Bündenbender, J., Frischauf, A., Gogvadze, G., Libbrecht, P., Pollet, M. y Ullrich, C. (2001). Activemath: A generic and adaptive web-based learning environment. *International Journal of Artificial Intelligence in Education*, 12, 385–407.
- Millán, E., , García-Hervás, E., Guzmán, E., Cruz, J. P. de-la y Rueda, A. (2003). Adaptation and generation in a web-based tutor for linear programming. En G. Gauthier, C. Frasson y K. VanLehn (Eds.), *Proceedings of the 3rd international conference on web engineering (icwe 2003). lecture notes in computer science*.
- Mislevy, R. y Almond, R. (1997). *Graphical models and computerized adaptive testing* (Informe Técnico No. 434). Center of the Study of Evaluation (CSE).
- Olea, J. y Ponsoda, V. (2001). *Tests adaptativos informatizados* (Curso de doctorado). Programa de Metodología del Comportamiento.
- Olea, J., Ponsoda, V. y Prieto, G. (1999). *Tests informatizados: Fundamentos y aplicaciones*. Pirámide.
- Owen, R. (1969). *A bayesian approach to tailored testing* (Research Report No. 69-92). Educational Testing Service.
- Owen, R. (1975). A bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, 70(350), 351–371.
- Parshall, C., Davey, T. y Pashley, P. (2000). Innovate item types for computerized testing. En W. van der Linden y C. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 129–148). Dordrecht (NL): Kluwer Academic Publishers.
- Patel, A., Kinshuk y Russell, D. (1998). A computer based intelligent assessment system for numeric disciplines. *Information Services and Use*, 18(1-2), 53–63.
- Pellegrino, J., Chudowsky, N. y Glaser, R. (2001). *Knowing what student knows: The science and desing of educational assessment*. National Academy of Science.
- Ramsay, J. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, 56, 611–630.

- Reye, J. (2002). A belief net backbone for student modelling. En S. Cerri, G. Gouardères y F. Paraguacu (Eds.), *Proceedings of the 6th international conference on intelligent tutoring systems (its 2002). lecture notes in computer science* (p. 596-604). New York: Springer Verlag.
- Thissen, D. (1993). Repealing rules that no longer apply to psychological measurement. En N. Frederiksen, R. Mislevy y I. Bejar (Eds.), *Test theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Thissen, D. y Mislevy, R. (1990). Testing algorithms. En H. Wainer (Ed.), *Computerized adaptive testing: A primer* (pp. 103–136). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- van der Linden, W. y Glas, C. (2000). *Computerized adaptive testing: Theory and practice*. Kluwer Academic Publishers.
- van der Linden, W. y Hambleton, R. (1997). *Handbook of modern item response theory*. New York: Springer Verlag.
- VanLehn, K. (1988). Student modeling. En M. C. Polson y J. Richardson (Eds.), *Foundations of intelligent tutoring systems* (pp. 55–76). Lawrence Erlbaum Associates Publishers.
- Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Lawrence Erlbaum.
- Wise, S. (1999). Tests autoadaptados informatizados: fundamentos, resultados de investigación e implicaciones para la aplicación práctica. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados: Fundamentos y aplicaciones* (pp. 189–203). Pirámide.