

# Inferencia estadística del conocimiento del alumno en entornos de evaluación basados en actividades procedimentales no acotadas

Jaime Gálvez<sup>1</sup>, Eduardo Guzmán<sup>1</sup> y Ricardo Conejo<sup>1</sup>

<sup>1</sup> Dpto. de Lenguajes y Ciencias de la Computación, Universidad de Málaga,  
Bulevar Louis Pasteur, 35, Campus de Teatinos,  
29071 Málaga, España  
{jgalvez, guzman, conejo}@lcc.uma.es

**Resumen.** La Teoría de Respuesta al Ítem (TRI) es un mecanismo estadístico que se emplea con éxito, desde principios del siglo XX, para inferir el conocimiento del alumno a través de tests. Sin embargo, para la evaluación de actividades procedimentales, las técnicas bien fundamentadas existentes, en general son complejas y se aplican a dominios muy bien acotados. En este artículo se describen cómo a través de un conjunto de técnicas que hemos desarrollado, basadas en la TRI, es posible inferir el conocimiento declarativo del alumno a través de actividades de tipo procedimental. Describimos cómo estas técnicas han sido empleadas, con estudiantes universitarios, en un dominio procedimental no acotado: la programación orientada a objetos.

**Palabras clave:** Modelado del alumno, evaluación del alumno, sistemas tutores.

## 1 Introducción

Desde el nacimiento de los primeros sistemas de enseñanza, las técnicas de Inteligencia Artificial (IA) han sido utilizadas para intentar dotar a estos sistemas con las capacidades necesarias para emular el comportamiento de un tutor humano [1]. La representación del conocimiento, el modelado del alumno y el diagnóstico cognitivo (o inferencia del conocimiento del alumno) son sólo algunas de las capacidades implicadas en el desarrollo de sistemas educativos. La IA se emplea en este ámbito para ofrecer al estudiante un proceso de enseñanza individualizado, acorde a sus necesidades y adaptándose a ellas para maximizar la influencia positiva en el mismo.

Desde la perspectiva del modelado del alumno, en la literatura (cf. [2]) pueden encontrarse diversas técnicas para abordar el problema de representar (y diagnosticar) características del alumno, tales como su nivel de conocimiento, sus errores conceptuales, etc. Hoy en día, existen mecanismos de evaluación bien fundamentados basados en teorías estadísticas como la *Teoría de Respuesta al Ítem* (TRI) [3] que se emplea con éxito para inferir el conocimiento del alumno a través de un test. Sin embargo, para la evaluación de actividades procedimentales (como los problemas), las técnicas bien fundamentadas existentes, en general son complejas. En esta línea,

propuestas como el seguimiento de modelos [4] u otras basadas en redes bayesianas [5], requieren actividades procedimentales en las que el conjunto de pasos que el alumno puede realizar está muy acotado y bien definido a priori [6]; y aún así el modelado de estos dominios se convierte en una ardua tarea.

En estudios preliminares [7] evaluamos cómo, empleando técnicas bien fundamentadas, podemos realizar una evaluación del conocimiento del alumno en un dominio acotado: la aplicación del algoritmo de optimización Simplex. En este artículo se describen estas técnicas basadas en la TRI, y se evalúa su idoneidad, esta vez en un dominio en el que el espacio de soluciones de las actividades no está acotado. La hipótesis de la que partimos es que, usando las técnicas propuestas, la inferencia del conocimiento declarativo del alumno es equivalente a la que proporcionaría un test basado en la TRI. La ventaja que ofrece nuestra propuesta es que el número de problemas requerido para la evaluación es notablemente menor que la cantidad de pregunta de test que se necesitarían para lograr un diagnóstico similar.

El artículo se estructura de la siguiente forma: En la siguiente sección se describe la técnica de modelado del dominio denominada, *modelado basado en restricciones* (MBR), que utilizamos en nuestra propuesta. A continuación, explicaremos los fundamentos de la TRI. En la sección 4 se presentan las ideas claves de este trabajo. La sección 5 se centra en explicar las características más relevantes de la herramienta educativa que hemos desarrollado. Posteriormente, en la sección 6, se describe el experimento realizado con la herramienta educativa para verificar nuestra hipótesis y los resultados obtenidos. Finalmente se presentan las conclusiones y el trabajo futuro.

## **2 Modelado basado en restricciones**

El uso del MBR en los entornos de enseñanza contribuye a mejorar el proceso de aprendizaje del alumno, haciendo que éste aprenda de sus propios errores mientras que resuelve un problema de un determinado dominio. El MBR se basa en la teoría de Ohlsson [8] sobre el aprendizaje a partir de errores. Según ella, el aprendizaje es un proceso que consta de dos fases: En la primera, se detecta un error durante la realización de una actividad; y posteriormente, éste es subsanado. Los errores suelen tener lugar cuando los alumnos intentan resolver un problema, pero no poseen el conocimiento declarativo necesario, no siendo capaces de aplicarlo de forma procedimental.

Para detectar errores, los sistemas tutores basados en el MBR, generan una representación de la solución que el alumno está construyendo, la cual se actualiza conforme el alumno realice acciones a través de la interfaz del sistema. En el MBR el dominio viene representado por un conjunto de principios de forma que, la solución del alumno se compara con esa colección para inferir qué violaciones del dominio está produciendo la solución del alumno. Los principios que forma el dominio se consideran las unidades fundamentales de conocimiento en el MBR y vienen representadas mediante restricciones sobre el estado que deben satisfacer todas las posibles soluciones correctas de todos los problemas. Es decir, una solución correcta a un problema nunca generará una representación que viole alguna de las restricciones del dominio.

Según el MBR, cada restricción viene definida por un par ordenado de condiciones:  $R_c$ ,  $S_c$ ; donde  $R_c$  es la condición de relevancia, que se emplea para determinar el tipo de problemas y estados para los cuales la restricción es relevante, estos es, donde podría ser aplicada.  $S_c$  es la condición de satisfacción y contiene la condición de error asociada a un determinado principio que puede infringir la solución a un problema. Cuando la  $R_c$  de una restricción es verdadera, para un determinado estado de una solución a un problema, se dice que la restricción es significativa, desde el punto de vista pedagógico, y por consiguiente,  $S_c$  también debería ser verdadera. En caso contrario, la restricción ha sido violada, lo que implica que se ha detectado un error. Tras la detección del error, el modelo del alumno se actualiza y el sistema debería responder aplicando una acción correctiva que ayude al alumno a subsanar su error conceptual.

Los resultados obtenidos por diversos tutores basados en el MBR prueban la efectividad de este enfoque en tareas de enseñanza [9] y su idoneidad frente a otras propuestas similares. Sin embargo, Ohlsson y Mitrovic [10] han señalado que, para permitir que los sistemas que empleen el MBR puedan ayudar a la toma de decisiones pedagógicas, es necesario disponer de un modelo del alumno a largo plazo. En este sentido, la mayoría de las propuestas existentes basadas en el MBR (a excepción de algunas que incluyen redes bayesianas [11]), infieren el conocimiento del alumno como la proporción de restricciones que el alumno sabe. Sin embargo, este heurístico no posee características que son necesarias en un sistema de diagnóstico del conocimiento (y en general, para cualquier sistema de este tipo), como por ejemplo la invariancia. Por esta razón, este tipo de estimaciones basadas en heurísticos están fuertemente condicionadas por los problemas concretos que el alumno ha realizado.

### 3 Teoría de Respuesta al Ítem

La TRI, desarrollada por Thurstone [3], es la disciplina más popular, basada en técnicas estadísticas para medir, de forma cuantitativa, ciertos rasgos, tales como la inteligencia, habilidades, nivel de conocimiento en un determinado concepto, personalidad, etc. Esta teoría se basa en dos principios [12]: Según el primero, el conocimiento que posee un alumno en una pregunta (o *ítem*) de un test puede explicarse a través de un factor denominado *nivel de conocimiento*. El segundo principio establece que la relación entre la probabilidad de responder de forma correcta al ítem y el nivel de conocimiento del alumno puede describirse a través de una función monótona creciente denominada *Curva Característica del Ítem* (CCI). Cuanto mayor sea el nivel de conocimiento del alumno, mayor será la probabilidad de responder correctamente a la cuestión. Esta función es además el elemento central de la TRI. Una de las diversas funciones empleadas (y quizás la más popular) para modelar la CCI es la función logística de tres parámetros (3PL):

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

donde  $P(u_i = 1 | \theta)$  es la probabilidad de responder correctamente al ítem  $i$  dado el nivel de conocimiento del alumno  $\theta$ , el cual suele medirse utilizando una escala

continua entre  $[-3.0, \dots, 3.0]$ . Los tres parámetros que caracterizan a esta curva dependen del ítem y son:

- La *discriminación* ( $a_i$ ), que es un valor proporcional a la pendiente de la curva y que cuanto mayor es, más discierne el ítem entre los niveles de conocimiento inferiores y los superiores.
- La *dificultad* ( $b_i$ ), que corresponde al valor del nivel de conocimiento para el cual la probabilidad de responder de forma correcta es la misma que la de responder de forma incorrecta (sin tener en cuenta la respuestas seleccionadas de forma aleatoria).
- La *adivinanza* ( $c_i$ ), que mide precisamente la probabilidad que existe de que el alumno responda correctamente a la pregunta a pesar de no poseer el conocimiento necesario para ello, modelando por tanto aquellas situaciones en las que el alumno responde de forma aleatoria.

La popularidad de la TRI es consecuencia directa de la consistencia de sus resultados. En otras propuestas como la *Teoría Clásica de los Tests*, los resultados de la estimación del conocimiento del alumno dependen de la muestra de alumnos sobre la que se ha realizado el test y, por consiguiente, los resultados en el test no son comparables con otros obtenidos en diferentes tests. Por el contrario, los resultados obtenidos aplicando la TRI poseen varias propiedades como la invariancia, esto es, el nivel de conocimiento inferido a través de esta teoría no depende del test. Así, si en dos tests que se administran al mismo alumno, se evalúa el mismo concepto, los resultados que se obtendrán serán muy similares.

Para poder aplicar la TRI es necesario disponer de los valores de las CCI correspondientes a cada ítem del dominio. Para ello, es necesaria la denominada fase de calibración, que es un proceso estadístico, para el que es necesario disponer de datos de administración de esos ítems a una muestra poblacional de alumnos. A través de este procedimiento se infieren los parámetros que caracterizan cada CCI. La entrada a este procedimiento está formada por los resultados de aquellos alumnos que realizaron tests con las cuestiones cuyas curvas quieren inferirse.

## **4 Evaluación de actividades procedimentales combinando la TRI y el MBR**

A través de la TRI se consiguen realizar inferencias del conocimiento del alumno con características deseables como la invariancia. Sin embargo, en principio, la TRI es difícil de aplicar cuando se trata de evaluar actividades de carácter procedimental. De hecho, para poder llevar a cabo una evaluación análoga a la que realizaría un profesor de un problema resuelto por un alumno, pero empleando la TRI, sería necesario construir un número bastante considerable de ítems que se centraran en todos los aspectos que se pueden evaluar con un único problema.

Nuestra propuesta trata de solventar los inconvenientes que presentan tanto la TRI como el MBR, a través de un conjunto de técnicas de evaluación que combina ambas. El objetivo es intentar mejorar los heurísticos habitualmente empleados en el MBR para la actualización de los modelos a largo plazo del alumno. Esta mejora consiste en

introducir técnicas de inferencia del conocimiento del alumno inspiradas en los fundamentos de la TRI. De esta forma, las evidencias que el alumno proporciona sobre su conocimiento serán las acciones que éste lleva a cabo mientras que resuelve un problema. Estas acciones se traducirán en la violación (o no) de restricciones de un conjunto de restricciones a través de las cuales se expresa el dominio.

Si en la TRI los elementos que se utilizan para determinar el conocimiento del alumno son los ítems, en esta propuesta se emplean las restricciones. Así, cada restricción tendrá asociada una curva característica que hemos denominado *Curva Característica de la Restricción* (CCR). Esta curva tiene la forma opuesta a una CCI ya que, mientras esta última representa la probabilidad de responder correctamente (conocimiento), la primera representa la probabilidad de violar una restricción en un problema (detección de un conocimiento erróneo). Cuando se viola una restricción, esto implica una carencia de conocimiento y por tanto, la curva necesaria para representarlo debe ser monótona decreciente. Cuanto mayor sea el nivel de conocimiento del alumno, menor será la probabilidad de que la restricción se viole. En la TRI, sería equivalente a una respuesta incorrecta a un ítem.

La distribución del conocimiento del alumno  $P(\theta | \Phi, \tau)$  se calculará como el producto de las CCRs de aquellas restricciones que han sido violadas, con las curvas opuestas para aquéllas que, siendo relevantes para el problema, no han sido violadas. Esta forma de calcular el conocimiento del alumno se basa en los mecanismos de inferencia que emplea la TRI:

$$P(\theta|\phi, \tau) = \prod_{i=1}^m \prod_{j=1}^n [P(c_j|\theta)^{f_{ij}}(1 - P(c_j|\theta))]^{r_{ij}} \quad (2)$$

donde  $\Phi = p_1, p_2, \dots, p_m$  representa el conjunto de problemas resueltos por el alumno y  $\tau = c_1, c_2, \dots, c_n$  la colección de restricciones de este dominio.  $P(c_j|\theta)$  es la curva característica de la restricción  $c_j$ ;  $r_{ij}$  es una variable binaria que indica si la restricción es relevante en el problema  $p_i$ ; y  $f_{ij}$  es otra variable binaria que indica si la acción del alumno en el problema  $p_i$  ha provocado la violación de la restricción  $c_j$ .

En nuestros trabajos anteriores [13] se han utilizado curvas características discretas cuyos valores son pares nivel de conocimiento/ probabilidad, simplificando notablemente la exigencia de datos requerida por la TRI para la inferencia de las curvas características. En esta propuesta, las CCRs son también discretas y cada valor indica la probabilidad que existe de que un alumno con cierto nivel de conocimiento viole una restricción.

El resultado de aplicar la ecuación 2 es una distribución de probabilidades en la que por cada nivel de conocimiento se expresa la probabilidad de que el conocimiento del alumno corresponda a ese nivel. Para calcular el denominado nivel de conocimiento se pueden emplear dos estrategias: tomar el valor esperado (o media) de la distribución (Esperanza a Posteriori) o bien elegir la moda (Máximo a Posteriori).

## 5 El sistema OOPS

Para llevar a la práctica el modelo propuesto en la sección anterior, hemos utilizado una nueva versión del sistema OOPS (*Object Oriented Programming System*) [14], el cual incorpora fundamentos del MBR como filosofía base para su funcionamiento. OOPS se enmarca en el dominio de la programación Orientada a Objetos y permite simular el comportamiento de un tutor humano en la instrucción de los alumnos detectando las debilidades en su conocimiento y actuando para paliarlas. Esta herramienta permite a los alumnos construir programas orientados a objetos en el pseudolenguaje que se utiliza en las titulaciones técnicas de la ETSI en Telecomunicación de la Universidad de Málaga. Los ejercicios se basan en definir e implementar clases (atributos y métodos) de acuerdo con un enunciado proporcionado por el sistema. La construcción se hace de forma visual, empleando la técnica de arrastrar y soltar elementos de una barra de herramientas, en la que se encuentran las sentencias y demás elementos necesarios para elaborar una aproximación orientada a objetos de un problema.

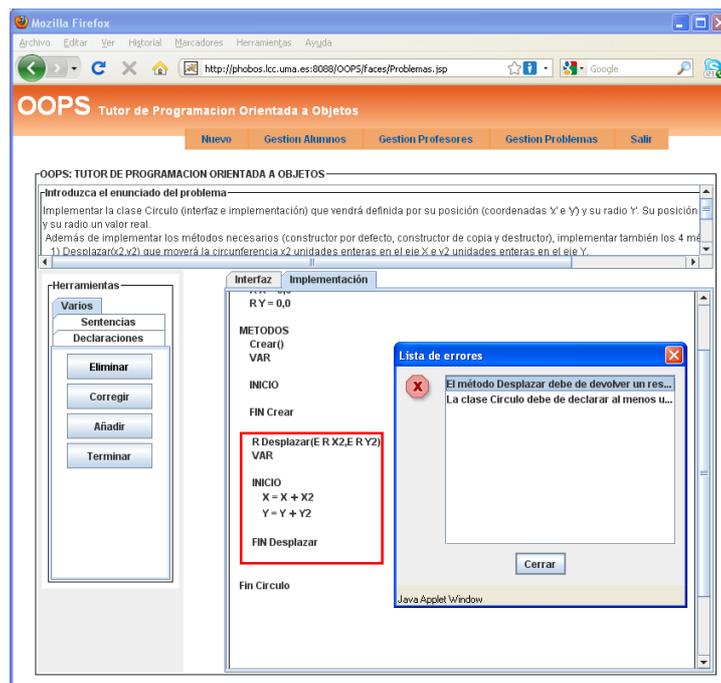


Figura 1. Interfaz del sistema OOPS.

Cuando el alumno decide corregir su solución, el sistema pone en marcha su motor de inferencia, el cual utiliza las restricciones definidas en el modelo de dominio para determinar los errores del alumno. Con estos errores, se actualiza el modelo del estudiante y, a excepción de que el sistema esté funcionando en modo de evaluación,

muestra mensajes de ayuda con el fin de que el alumno corrija sus errores conceptuales.

La arquitectura actual del sistema se basa en un marco de trabajo genérico [15] definido para la construcción de sistema de enseñanza basados en la realización de actividades de carácter procedimental. Por esta razón, OOPS está estructurado en los siguientes módulos:

1. El componente más externo, la *interfaz* (ver Figura 1), mediante la cual los profesores o alumnos pueden introducir o resolver problemas, respectivamente. A través de ella, el sistema puede configurarse para mostrar información adicional al usuario, como por ejemplo, mensajes de ayuda/refuerzo ante los errores cometidos.
2. La parte principal de sistema es el *módulo pedagógico*, en el que se sitúan los elementos necesario para estimar el conocimiento del alumno, tales como los algoritmos de estimación, un motor de inferencia JESS [16] para la identificación de las restricciones violadas; o aquellos que permiten controlar las acciones pedagógicas a realizar para instruir al alumno de la forma más adecuada, como el módulo de selección adaptativa de problemas.
3. El *modelo de dominio*, formado por el conjunto de restricciones y los problemas.
4. El *modelo del alumno*, que contiene las restricciones violadas y las satisfechas, las distribuciones con las estimaciones del conocimiento del estudiante y los *logs* de las acciones que llevó a cabo en cada problema.

## 6 Experimentación

Para evaluar la validez de nuestras técnicas de diagnóstico a través de actividades procedimentales no acotadas, llevamos a cabo un experimento con alumnos universitarios. El objetivo de la sesión era que los estudiantes probasen sus conocimientos de programación orientada a objetos a través de una sesión práctica llevada a cabo en un laboratorio de docencia. Ésta fue estructurada, a grandes rasgos, en dos fases: Primeramente, se administró un test usando el sistema web Siette [17] que proporciona evaluaciones basadas en la TRI. Posteriormente, se propusieron dos problemas a los alumnos para que los resolviesen usando el sistema OOPS.

El objetivo de esta sesión era que los datos recopilados, a través de un test por una parte y de problemas de programación por otra, fueran comparados para verificar si la inferencia del conocimiento de cada alumno, empleando las dos alternativas, proporcionaba como resultado datos similares.

### 6.1 Diseño del experimento

Con el fin de conseguir que los resultados de ambos sistemas fuesen comparables, el experimento se diseñó de forma que los mismos conocimientos que estaban siendo puestos a prueba en el test, fuesen presentados en los problemas prácticos a desarrollar. Para ello, en el sistema OOPS se eligieron dos problemas tipo que

evaluasen los conceptos básicos de la programación orientada a objetos, o lo que es lo mismo, dos problemas con los cuales se pudiesen evaluar las restricciones fundamentales asociadas a los principios del dominio. Usamos un subconjunto de 15 restricciones que fuese significativo en cuanto a la importancia de los conceptos contenidos, el cual fue definido por varios expertos en el dominio. En cuanto al test, el conjunto de preguntas a mostrar fueron diseñadas a partir del conjunto de restricciones previamente seleccionado. Cada restricción seleccionada tenía dos preguntas en el test que evaluaban el mismo concepto, lo que supone un total de 30 preguntas.

El experimento fue llevado a cabo en mayo del 2009 con estudiantes de Ingeniería Técnica en Telecomunicación de la Universidad de Málaga. Dichos alumnos habían recibido previamente clases en pizarra sobre los conceptos que se evaluarían en el experimento. Un total de 20 alumnos asistieron a la sesión práctica. Tras la realización del test (en el en ningún momento se mostraban las soluciones a las preguntas), los alumnos comenzaron a utilizar el sistema OOPS. Inicialmente se les propuso a todos un primer problema que les sirvió de entrenamiento en el manejo del sistema y que no se tuvo en cuenta para el análisis. Posteriormente, resolvieron dos problemas de programación a través de OOPS.

## **6.2 Análisis de los datos**

Una vez obtenidos los datos de ambos sistemas, incorporamos las características de la TRI a estos resultados preliminares utilizando la herramienta MULTILog [18], la cual es una de las más populares para este proceso. Primeramente calibramos las CCIs del test de acuerdo al modelo 3PL de la TRI. Para ello, usamos los resultados del test en una matriz de valores booleanos. En dicha matriz, requerida como entrada a MULTILog, para cada estudiante e ítem, se representaba si el concepto asociado se sabía (pregunta correctamente respondida) o no (respuesta incorrecta). Posteriormente, la calibración obtenida fue usada en MULTILog en conjunción con los resultados del test para obtener la estimación del conocimiento de cada estudiante.

De forma análoga, los datos obtenidos en OOPS fueron usados para calibrar las CCRs. En este caso, la matriz utilizada en MULTILog representaba, para cada estudiante y restricción, si el concepto asociado se sabía (no se había violado) o no (restricción violada) durante la resolución de los problemas. Las curvas obtenidas se utilizaron de nuevo con los datos de OOPS para generar una valoración del conocimiento del alumno.

Este proceso de evaluación fue cuidadosamente diseñado para que los resultados obtenidos fuesen homogéneos en cuanto a la fuente del conocimiento evaluada y en la naturaleza de los datos usados para estimar el conocimiento. Lo primero, mediante la ya mencionada correspondencia entre restricciones e ítems; y lo segundo, usando la matriz de entrada para MULTILog con el mismo significado de los valores: un valor booleano verdadero indicando el conocimiento del concepto relacionado y otro falso para conceptos no aplicados correctamente. Finalmente, es necesario mencionar que se utilizó MULTILog en ambos casos, para asegurar que las técnicas de calibración e inferencia que se empleaban eran las mismas.

### 6.3 Resultados obtenidos

Para llevar a cabo la comparación de ambas estimaciones, realizamos una prueba t de Student por pares, al 95% de confianza, sobre las dos evaluaciones. Este estadístico es comúnmente utilizado para comparar la diferencia entre dos poblaciones de tamaño reducido. La hipótesis nula del mismo es que la diferencia de medias de las poblaciones es cero. El análisis aportó un p-valor de 0.7972, que claramente sugiere que no podemos rechazar la hipótesis nula anterior y que, por tanto, no existe una diferencia significativa entre las evaluaciones obtenidas con OOPS, aplicando nuestras técnicas de inferencia que combinan la TRI y el MBR, y las obtenidas con Siete, aplicando el modelo 3PL.

## 7 Conclusiones y Líneas Futuras

Las técnicas descritas en este artículo aportan diversas ventajas a la evaluación de actividades de tipo procedimental: Las estimaciones sobre el conocimiento del alumno son invariantes, es decir, no dependen del conjunto de problemas que el alumno haya realizado; el grado de precisión de las estimaciones puede ser controlado; y, por último, se trata de un mecanismo de inferencia estadístico en el que las CCRs se estiman a priori empleando datos históricos de alumnos que realizaron los mismos problemas.

La hipótesis de la que hemos partido es que, usando técnicas estadísticas, inspiradas en aquéllas que se emplean normalmente para determinar el conocimiento en tests y, a través de actividades de tipo procedimental, podemos obtener diagnósticos similares a los que conseguiríamos a través de un test basado en la TRI. Los resultados de nuestro experimento en un dominio no acotado, sugieren que nuestra hipótesis es cierta. De esta forma, podemos llevar a cabo evaluaciones bien fundamentadas a través de un número reducido de actividades procedimentales que, si estas evaluaciones se hicieran empleando pregunta tipo test, para obtener resultados similares, necesitaríamos un número considerable de cuestiones.

Los tests basados en la TRI pueden administrarse de forma adaptativa, generando los denominados *tests adaptativos*, en los que cada pregunta que se muestra al alumno, se selecciona de forma dinámica, en función del nivel de conocimiento estimado. Además, el número de preguntas del test se decide también de forma dinámica, en función de la precisión que se desee alcanzar en las estimaciones. En este sentido, nuestras líneas futuras se centran en desarrollar un mecanismo análogo a los tests adaptativos que permita la administración individualizada de actividades procedimentales.

En este trabajo se ha estimado el conocimiento declarativo del alumno a través de actividades de tipo procedimentales. En la actualidad estamos trabajando en la extensión del modelo para evaluar también el conocimiento procedimental. Asimismo, puesto que la teoría presentada se ha estudiado en dos sistemas concretos, también es necesario extender el estudio a otros sistemas de evaluación que confirmen las hipótesis planteadas en estos experimentos.

**Agradecimientos.** Este trabajo ha sido co-financiado por el Ministerio de Ciencia e Innovación (TIN2007-67515) y por la Consejería de Ciencia, Innovación y Ciencia de la Junta de Andalucía (TIC-03243). Los autores quieren agradecer también al prof. David Bueno y a sus alumnos su buena disposición a participar en los experimentos descritos en este trabajo.

## Bibliografía

1. Brooks, R. A.: *Intelligence without representation*. Artificial Intelligence, 47, pp. 139–159 (1991).
2. Greer, J., McCalla, G.: *Student Modeling: The Key to Individualized Knowledge-based Instruction*. Springer-Verlag. pp. (1994)
3. Thurstone, L. L.: A method of scaling psychological and educational tests. Journal of Educational Psychology. Vol. 16, pp. 433–451 (1925)
4. Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W.: Cognitive modeling and intelligent tutoring. Artificial Intelligence, 42, pp. 7–49 (1990)
5. Conati, C., Gertner, A., VanLehn, K.: Using Bayesian networks to manage uncertainty in student modeling. User Modeling & User-Adapted Interaction, 12(4), pp. 371–417 (2002)
6. Hayes, J. R.: *Cognitive psychology: Thinking and creating*. Homewood, IL: Dorsey Press (1978)
7. Gálvez, J., Guzmán, E., Conejo, R., Millán, E.: Student Knowledge Diagnosis Using Item Response Theory and Constraint-Based Modeling. The 14th International Conference on Artificial Intelligence in Education (AIED 2009). Vol. 200, pp. 291–298, (2009)
8. Ohlsson, S.: *Constraint-based Student Modeling*. In Student Modeling: The Key to Individualized Knowledge-based Instruction. Springer-Verlag. pp. 167–189 (1994)
9. Mitrovic, A., Martin, B., Suraweera, P.: Intelligent Tutors for All: The Constraint-Based Approach. IEEE Intelligent Systems, IEEE Educational Activities Department, Vol. 22, pp. 38–45 (2007)
10. Ohlsson, S., Mitrovic, A.: Constraint-based knowledge representation for individualized instruction. Computer Science and Information Systems, 3, 1–22 (2006)
11. Mayo, M., Mitrovic, A.: Optimising ITS behaviour with Bayesian networks and decision theory. International Journal of Artificial Intelligence in Education. Vol. 12, pp. 124–153 (2001)
12. Hambleton, R. K., Swaminathan, H., Rogers, J. H.: *Fundamentals of Item Response Theory (Measurement Methods for the Social Science)*. Sage Publications, Inc (1991)
13. Guzmán, E., Conejo, R., Pérez-de-la-Cruz, J.L.: Adaptive Testing for Hierarchical Student Models. *User Modeling and User-Adapted Interaction*. Vol. 17, pp. 119–157 (2007)
14. Gálvez, J., Guzmán, E., Conejo, R.: A blended E-learning experience in a course of object oriented programming fundamentals. *Knowledge-Based Systems*, vol. 22, n. 4, pp. 279–286 (2009)
15. Gálvez, J., Guzmán, E., Conejo, R.: A SOA-Based Framework for Constructing Problem Solving Environments. The 8th IEEE International Conference on Advanced Learning Technologies. pp. 126–128 (2008)
16. Friedman-Hill, E.J.: JESS, The Java Expert System Shell. SAND–98-8206 (1997)
17. Guzmán, E., Conejo, R., Pérez-de-la-Cruz, J.L.: Improving Student Performance using Self-Assessment Tests. *IEEE Intelligent Systems*. Vol. 22, pp. 46–52 (2007)
18. Thissen, D.: Multilog: Multiple, categorical item analysis and test scoring using item response theory (version 5.1). Mooresville, In Scientific Software (1988)