

**SISTEMA ADAPTATIVO INFORMATIZADO PARA APOYAR EL ÁREA DE
PROGRAMACIÓN EN LA UNIVERSIDAD DE CÓRDOBA**

***COMPUTERIZED ADAPTIVE SYSTEM TO SUPPORT THE AREA OF PROGRAMMING
IN THE UNIVERSITY OF CORDOBA***

Camilo Cervantes Salazar^{8*}, Daniel Salas Álvarez⁹

Recibido para publicación: 2 de marzo de 2016 - Aceptado para publicación: 14 de mayo 2016

RESUMEN

El presente trabajo de investigación tiene como propósito mostrar un sistema adaptativo (ADAPTIKE) orientado a la web que permita administrar tests s informatizados enfocados en el área de programación, para el cumplimiento de ésta meta se fundamentó en la teoría de respuesta al ítem, se determinó la especificaci conformó un banco de preguntas relacionados al área a evaluar, se determinó la arquitectura y la especificación de requisitos del sistema, posteriormente se desarrolló el sistema y la validación del mismo, con estudiantes y docentes del área de programación del programa de ingeniería de sistemas en la Universidad de córdoba.

Palabras claves: Diseño, construcción, validación, ensayos, máquina.

ABSTRACT

This research aims to develop a web oriented information system (ADAPTIKE) that is able to deliver computerized adaptive tests focused on the programming area. To fulfill this goal

^{8*} Ingeniero de sistemas, Universidad de Córdoba – Montería – Colombia
ccervantessalazar@correo.unicordoba.edu.co

⁹ Ingeniero de Sistemas, Magister en Informática, Director Grupo Socrates, Universidad de Córdoba – Montería – Colombia dajosallas@gmail.com

a bank of questions related to the mentioned area was created, the system requirements were determined from students and teachers who expressed their perceptions about assessment in their academic context and also a research in scientific journals and books about adaptive assessment models and item response theory. Subsequently, system development was made and operation tests were conducted.

Key words: computerized adaptive test, item response theory, adaptive assessment system.

1. INTRODUCCIÓN

En el contexto de una institución de educación superior, los medios para evaluar los conocimientos de los estudiantes es de gran importancia debido a que mediante este proceso se pueden verificar los resultados que evidencien la apropiación de conocimiento y el aprendizaje de los estudiantes.

En la Universidad de Córdoba, los procesos de evaluación a estudiantes se caracterizan por seguir la teoría clásica del test y se administran por medio de sistemas de información que permiten realizar evaluaciones online y también con tests a papel y lápiz.

Sin embargo, al revisar los resultados de éstos tests en el área de programación se observa una tendencia en la cual los estudiantes evaluados no logran obtener las calificaciones mínimas requeridas para aprobar el test.

A partir de esta observación se llevó a cabo un estudio en el cual buscaba conocer la percepción de estudiantes y docentes sobre la forma como se realizan los procesos de evaluación en la institución.

Por medio de encuestas y entrevistas, los estudiantes y docentes manifiestan que los tests tradicionales cumplen su papel como herramienta clave en la formación profesional pero que requiere mejorar en aspectos como la forma como se entrega el resultado del test, en la cual se indica una calificación que no permite conocer de forma detallada las fortalezas y debilidades evidenciadas a partir de dicho test, y la calidad de las preguntas del test, las cuales en ocasiones son demasiado complejas y no corresponden con los conocimientos que maneja el individuo evaluado.

De esta manera se evidencia la necesidad desarrollar una herramienta que aplique un modelo de evaluación diferente al tradicional. Se trata de los tests adaptativos informatizados, una forma de evaluación por computador que se adapta al nivel de conocimiento del estudiante evaluado eligiendo la siguiente pregunta a plantear basándose en la estimación del nivel de conocimiento del estudiante en cada momento e igualmente la decisión de finalizar el test.

Los tests adaptativos informatizados (TAI) denominados comúnmente CATs por sus siglas en inglés (Computer Adaptive Tests) es un tipo de test administrado por computador, donde la selección de la siguiente pregunta a administrar y la decisión de detener la prueba se realiza de forma dinámica basada en el perfil del estudiante, el cual es creado y actualizado durante la interacción con el sistema [1].

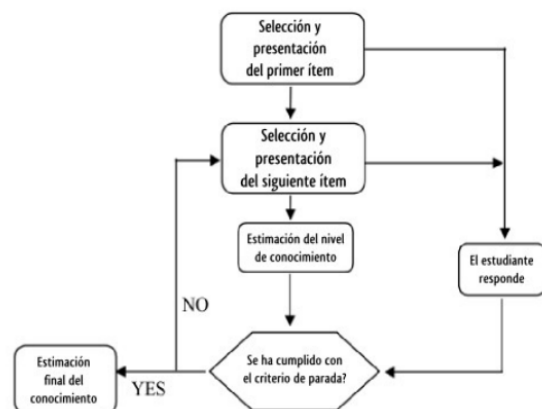


Figura 1. Algoritmo TAI

Los tests adaptativos resultan más convenientes que los tests tradicionales ya que tienden a ser más cortos, se presenta un aumento en la precisión del

nivel de conocimiento del evaluado en un aspecto determinado y el mismo test se muestra de diferentes maneras para cada individuo lo cual conduce a una disminución del fraude en la aplicación del test [2].

El origen de estos tests viene del investigador francés Alfred Binet [3] quien en 1905 introdujo las bases de la adaptabilidad en los tests mientras desarrollaba el test de inteligencia Binet. Este test fue diseñado para identificar el nivel de dificultad que es más apropiado para cada examinado y se caracteriza por usar bancos de ítems precalibrados, tener la capacidad de ajustar el nivel de dificultad inicial, aplicar reglas de selección de ítems y poseer una regla predefinida para la terminación del test.

A medida que las aplicaciones de las evaluaciones en diversos campos se volvieron prácticas comunes, surgió la necesidad de sistematizar dicho proceso para reducir costos y aprovechar el potencial de los recursos tecnológicos. Sin embargo, la informatización de los tests tuvo que esperar a que la tecnología fuera económicamente accesible para que su uso extensivo fuera un hecho factible.

En 1980, con el abaratamiento de los computadores personales se hizo posible que se desarrollaran los primeros tests informatizados, los cuales se caracterizan por aplicar el mismo test tradicional que se hace a papel y lápiz pero con las ventajas que proporciona la tecnología. En esa década la mayoría de tests estandarizados de Estados Unidos y Europa se hacían utilizando esta clase de tests.

Paralelamente, se abrió la posibilidad de implementar sistemas para la aplicación de tests adaptativos computarizados (TAI). El primer sistema comercial conocido para la aplicación de tests adaptativos fue el MicroCAT Testing System [4], desarrollado por Assessment Systems Corporation en 1984.

En 1990, se efectuó la primera evaluación adaptativa computarizada a gran escala en EE.UU. El departamento de defensa norteamericano desarrolló el sistema CAT-ASVAB [5], el cual es la versión adaptativa de la prueba ASVAB usada para el enlistamiento de personal en las fuerzas armadas, y es obligatoria para todo estudiante de último año de bachillerato.

Los resultados obtenidos con los TAI's han sido satisfactorios por tal motivo hoy en día, gracias a las condiciones tecnológicas actuales y el desarrollo teórico de los últimos 60 años, ha prosperado el desarrollo y uso de los TAI's en muchos campos de estudio.

Por ejemplo, un tema fundamental en el mundo actual es certificar la capacidad de manejar idiomas diferentes a la lengua de nacimiento, y para ello existen herramientas como J-CAT [6], el cual es un sistema de evaluación adaptativa que fue desarrollado ante la necesidad de evaluar y certificar el manejo del idioma japonés de estudiantes de intercambio que van a prepararse en Japón. La certificación de conocimientos no se limita al lenguaje, también aplica para el conocimiento general, tal es el caso de las pruebas GRE [7] y GMAT [8], el primero es un examen obligatorio para la mayoría de los programas de postgrado de Estados Unidos, y en menor medida en Europa y Canadá y el segundo es el más usado en el mundo para medir las aptitudes de los recién graduados en escuelas de Negocios y Finanzas. Sirve como criterio de selección para las admisiones a los MBA's principalmente en los Estados Unidos y también en otros lugares del mundo. Ambas evaluaciones se administran por medio de TAI's. También se realizaron aplicaciones en la salud, con proyectos como CAT-Health [9], un test adaptativo informatizado desarrollado en España para la medición de la calidad de vida relacionada con la salud de un paciente. Con éste se facilitó la medición de los resultados reportados

por los pacientes en la práctica clínica, y un trabajo desarrollado por Ruo et. al [10] en el cual los autores desarrollaron un CAT que les permitiera conocer datos sobre la disnea como la severidad y frecuencia del síntoma. El uso de CAT en este escenario es vital para minimizar la carga del paciente.

Con este contexto en mente, este trabajo busca desarrollar un sistema de información para la aplicación de tests adaptativos informatizados que permita realizar evaluaciones capaces de adaptarse al nivel de conocimiento de cada estudiante, el cual es estimado de acuerdo con las respuestas que ofrece en cada pregunta, siguiendo los fundamentos establecidos en la Teoría de Respuesta al Ítem y enfocado en el área de fundamentos de programación

2. ESTADO DEL ARTE

2.1. Teoría de Respuesta al Ítem

La teoría de respuesta al ítem (TRI) se originó a finales de 1960 y tiene sus raíces en la teoría de la medición psicológica. Según Lord, el término latent trait (característica latente), se refiere originalmente a la deducción de hipótesis psicológicas que se suponían que debían enmarcar el comportamiento observado de un individuo. En el contexto de las evaluaciones, la definición de éste término se convierte en un atributo que se identifica por la consistencia de las respuestas del test [12]. Cada pregunta o ítem en un test se le asigna una Curva Característica del Ítem (CCI) la cual es una función que representa la probabilidad de que dada una respuesta correcta a esa pregunta dado un cierto nivel de conocimiento del estudiante $\theta \in (-\infty, +\infty)$, definida como un número real. Representamos esto por la expresión: $P(U_i = 1|\theta)$ o simplemente P_i . Lógicamente, la probabilidad de fallar la pregunta es $P(U_i = 0|\theta) = 1 - P(U_i =$

$1|\theta)$, o simplemente Q_i . Si el test está compuesto por n preguntas, sabiendo los CCIs, y suponiendo independencia local de los ítems, es posible construir una función probabilística:

$$L(u_1, u_2 \dots u_n|\theta) = \prod p_i^{u_i} Q_i^{1-u_i}, i = 1 \dots n$$

El valor máximo de esta función proporciona una estimación del valor más probable de θ . Es posible obtener una distribución de la probabilidad de θ aplicando n veces la regla de Bayes [12].

2.2. Curva característica del Ítem (CCI)

Uno de los principales problemas de la IRT es hallar las CCIs. Han sido propuestos muchos modelos. Los más populares son aquellos que suponen que los CCIs pertenecen a una familia de funciones que dependen de una, dos o tres parámetros. Estas funciones se construyen basándose en la función de distribución normal o la función de distribución logística.

Por ejemplo, basándose en la función logística, para el modelo de 3-parámetros, el CCI puede describirse como:

$$P_i(\theta) = c_i + (1 - c_i) \frac{1}{(1 + e^{(-1.7a_i(\theta - b_i)))}}$$

donde b_i es conocida como la dificultad de la pregunta; c_i es el factor de adivinación, y a_i el factor de discriminación [13].

El factor de adivinación es la probabilidad de que un estudiante sin conocimiento alguno resuelva la pregunta. La dificultad representa el nivel de conocimiento en el cual el estudiante tiene la misma probabilidad de responder la pregunta bien o mal excluyendo el factor de adivinación. El factor de discriminación es proporcional a la inclinación de la curva. Si el factor de discriminación es alto, es probable que los estudiantes con nivel menor que b fallen y los estudiantes con nivel mayor que b con mayor seguridad contestarán la pregunta correctamente.

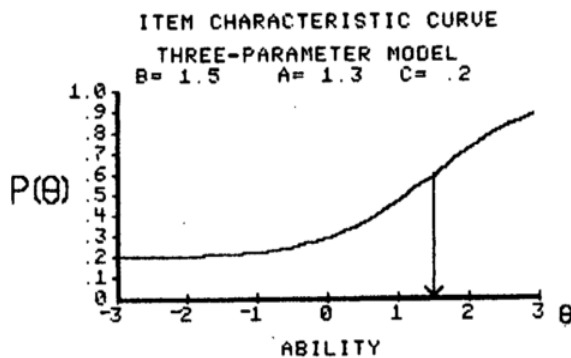


Figura 2. CCI para un modelo de 3-parámetros (Fuente: Baker).

Asumiendo que las formas de CCI pertenecen a esta familia, el problema es formulado como la estimación de los parámetros que se ajusta mejor. Para estimar estos parámetros, es sabido que el modelo de 1-parámetro es relativamente simple de usar y ofrece estimaciones confiables con 100 sujetos mínimo. El modelo de 2-parámetros toma en cuenta las posibles diferencias en la discriminación de los ítems y, aunque los resultados poseen mayor precisión, también significa que la mayor complejidad del modelo requerirá un mayor número de participantes, 200 por lo menos. El modelo de 3-parámetros tiene en cuenta el factor de adivinación, pero el precio a pagar incluye una muestra de mínimo 1000 individuos [13]. La situación más común se encuentra cuando tanto la habilidad como los parámetros de dificultad de los ítems son desconocidos. En este caso se necesita una matriz formada por las respuestas dadas por N estudiantes para un conjunto de n ítems.

Hay muchos métodos de estimación de tales parámetros, sin embargo los más usados son el método de máxima verosimilitud conjunta y el método de máxima verosimilitud marginal.

2.3. Supuestos de la Teoría de Respuesta al Ítem

Los modelos matemáticos de la TRI incluyen unos supuestos sobre los datos a los cuales se le está aplicando el modelo. Éstos supuestos hacen viable algunas operaciones clave de la TRI como el cálculo de la probabilidad de responder un ítem correctamente dependiendo de la habilidad del evaluado y de su nivel de conocimiento.

Unidimensionalidad: La unidimensionalidad se refiere al hecho de que los ítems de un test adaptativo solo miden una sola habilidad. Esto no se puede garantizar completamente en un test ya que existen multitud de factores que influyen en el resultado final de la evaluación. Por lo tanto los ítems del test deben contar con un factor dominante que influya en el rendimiento del test para ser considerados unidimensionales [13].

Independencia local: La independencia local ocurre entre las respuestas de dos ítems de un test e indica que la respuesta de un ítem no tiene ninguna incidencia en la respuesta de otro ítem, es decir, son independientes una de otra. Esto permite calcular la probabilidad de responder un conjunto de preguntas por medio del producto de la probabilidad de respuesta de cada una de las preguntas que conforman el conjunto [13].

2.4. Estimación de parámetros por máxima verosimilitud conjunta

Cuando se considera una muestra de N estudiantes respondiendo a n preguntas, la función de máxima verosimilitud conjunta es:

$$L(u_1, u_2, \dots, u_N | \theta, a, b, c) = \prod_i \prod_a P_{a_i}^{(u_i)} Q_{a_i}^{(1-u_i)}, \quad a=1 \dots N, i=1 \dots n$$

donde se supone la existencia de independencia local y u_a es el patrón de respuesta de un estudiante correspondiente a los n ítems considerados, θ es un vector que sus componentes son N parámetros de la habilidad de la persona (una por cada estudiante); a, b, c también son vectores formados por los parámetros de dificultad

del ítem [14]. En el modelo de 3-parámetros hay $3n$ parámetros de ítem, en el modelo de 2-parámetros $2n$ y en el modelo de 1-parámetro n parámetros. Entonces en el modelo de 3-parámetros tendrán que ser estimados un total de $3n + N$. Los valores de los parámetros se consiguen al maximizar la función de verosimilitud o su logaritmo dado por:

$$\ln L = \sum \sum [u_{ai} \ln P_{ai} + (1 - u_{ai}) \ln Q_{ai}], a = i = 1 \dots N$$

La dificultad del ítem y la habilidad de la persona no son determinadas unívocamente por éste método. Para eliminar esta solución indeterminada se elige una escala arbitraria para θ y b , usualmente la escala estándar con promedio 0 y desviación estándar 1. Luego se seleccionan los valores iniciales para el parámetro de la habilidad de la persona, normalmente $\ln(\# \text{ éxitos} / \# \text{ fracasos})$; para cada uno [14].

Ahora los parámetros del ítem se estiman de la misma forma que se describe en el paso anterior. En una segunda fase, los parámetros del ítem se consideran conocidos y se estiman los parámetros de la habilidad. Este procedimiento se repite hasta que se alcancen las diferencias mínimas en los estimadores de los parámetros entre dos etapas sucesivas.

3. MATERIALES Y MÉTODOS

3.1. Análisis y diseño

En esta etapa se llevó a cabo un análisis del problema para identificar los requerimientos funcionales y no funcionales del sistema. Luego se modeló la información que el sistema maneja y el flujo de información que tenía lugar en el sistema de acuerdo a eventos, estados y condiciones particulares. En esta etapa se logró describir de forma detallada las características del sistema que permitirían su posterior implementación.

3.2. Desarrollo

En esta etapa se eligieron las herramientas de desarrollo, los patrones de diseño, y se producirá el sistema siguiendo los parámetros de diseño establecidos en la etapa de diseño. Se utilizó PHP como lenguaje de programación y una arquitectura cliente-servidor siguiendo el patrón MVC. Se contó con un sistema gestor de bases de datos relacionales MySQL y tecnologías web como Html5 y Javascript para las vistas.

3.3. Conformación del banco de ítems

Los tests adaptativos informatizados requieren una cantidad amplia de ítems para poder ser ejecutados. Por tal razón se hizo una búsqueda en internet del mayor número posible de preguntas relacionadas con el área de programación y luego se seleccionó un subconjunto de ítems que fueran adecuados a los propósitos del proyecto. Estos ítems fueron ingresados en el sistema y posteriormente fueron calibrados por medio de simulaciones con el programa WinGen [15] en el cual se establecen valores para los parámetros de los ítems de acuerdo con una distribución normal con promedio 0 y desviación estándar 1.

3.4. Pruebas

En esta etapa se realizaron pruebas unitarias de cada componente del sistema y posteriormente se simula una situación real usando dos estudiantes y un docente quienes participarán en el proceso de evaluación usando el sistema.

4. RESULTADOS Y DISCUSIÓN

4.1. Sistema ADAPTIKE

El sistema se desarrolló usando un modelo dividido en capas como lo establece el patrón Modelo-Vista-Controlador (Figura 2), en el cual se tiene una capa de presentación para que los

usuarios puedan acceder a través de internet a la vista de la aplicación por medio de un navegador web. La interfaz de usuario se implementa usando tecnologías orientadas a la web como Html, CSS, Javascript y PHP. La aplicación se encuentra alojada en un servidor web apache, el cual representa la capa de negocio del sistema. Allí se encuentra el núcleo del sistema, con sus respectivos módulos y funcionalidades:

- Administración del test adaptativo: Es un módulo que permite ejecutar un test adaptativo y procesar los resultados de dicho test, seleccionar las preguntas del test y calcular los niveles de conocimiento de los individuos evaluados.
- Gestión de información: Agrupa todos los módulos de gestión de datos que

permiten administrar la información del sistema. Tienen en común que todos realizan operaciones CRUD básicos entre otras funcionalidades.

- Control de acceso: Es un módulo que permite controlar los accesos a las funcionalidades del sistema e identificar a los usuarios que acceden al sistema para se adapte al rol del usuario.

Para la persistencia de la información que se maneja, se tiene una capa de datos en la cual se encuentra un servidor de bases de datos Mysql y una base de datos que contiene las tablas utilizadas por el sistema. Para la conexión con la aplicación se cuenta con una interfaz de conexión que es utilizada desde los controladores de la aplicación para poder comunicarse con la base de datos e intercambiar información.

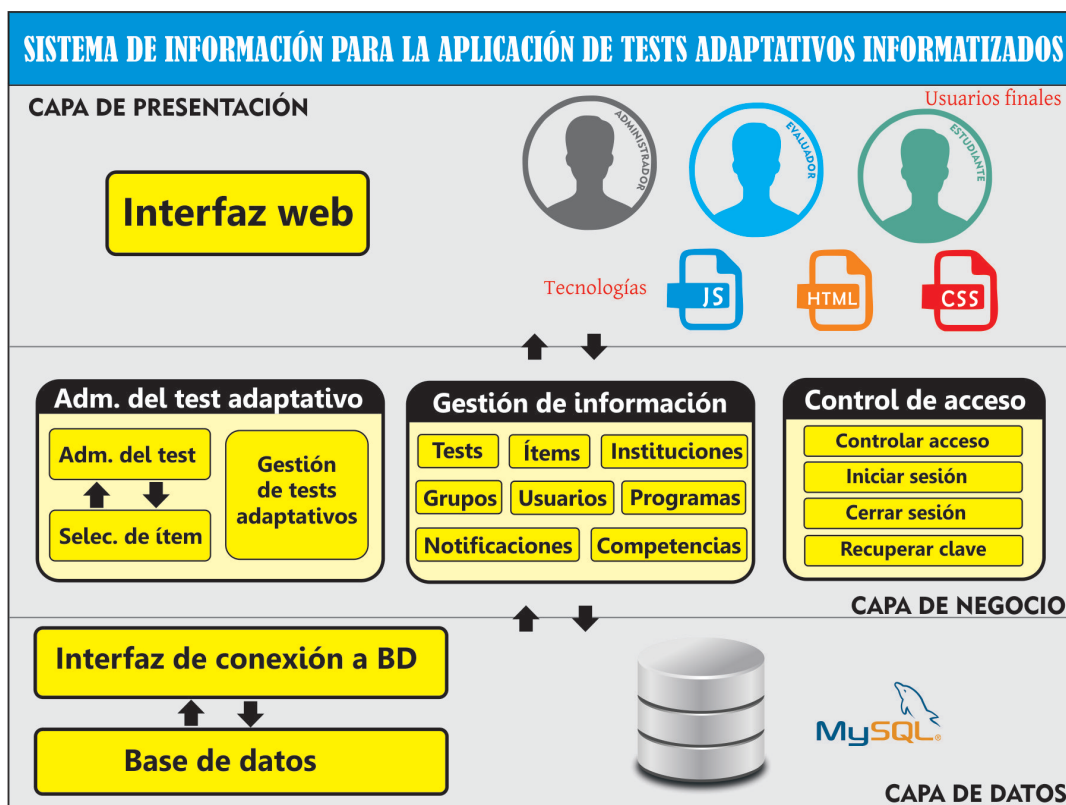


Figura 3. Arquitectura del sistema Adaptike

4.2. Pruebas y discusión

Para las pruebas del sistema se accede al sistema como evaluador quien cuenta con la opción de probar el test adaptativo, administrar el banco de ítems, competencias, entre otras funciones. Se accede a la opción Tests y luego crear test. Para la creación del test se define la categoría a evaluar, la cual corresponde a la competencia que se quiera poner a prueba, en este caso seleccionamos la de Programación, la cual está asociada a un conjunto de ítems previamente ingresados. Luego se selecciona el tipo de criterio de parada del test. Se cuenta con el criterio de parada de número de ítems en el cual el test para cuando cumple con el número de ítems administrados definidos en esta opción. También se cuenta con el criterio de nivel de conocimiento, en el cual se

ingresa un valor numérico de -2 a 2 siendo -2 el menor valor y 2 el mayor valor. Con esto se logrará que el test termine cuando el individuo alcance el nivel de conocimiento configurado. Una vez creado el test se procede a probarlo por medio de la opción administrar. El test muestra la primera pregunta y se diseñó una opción para verificar el proceso de selección de ítems y demostrar la aplicación de la Teoría de Respuesta al Ítem en el proceso.

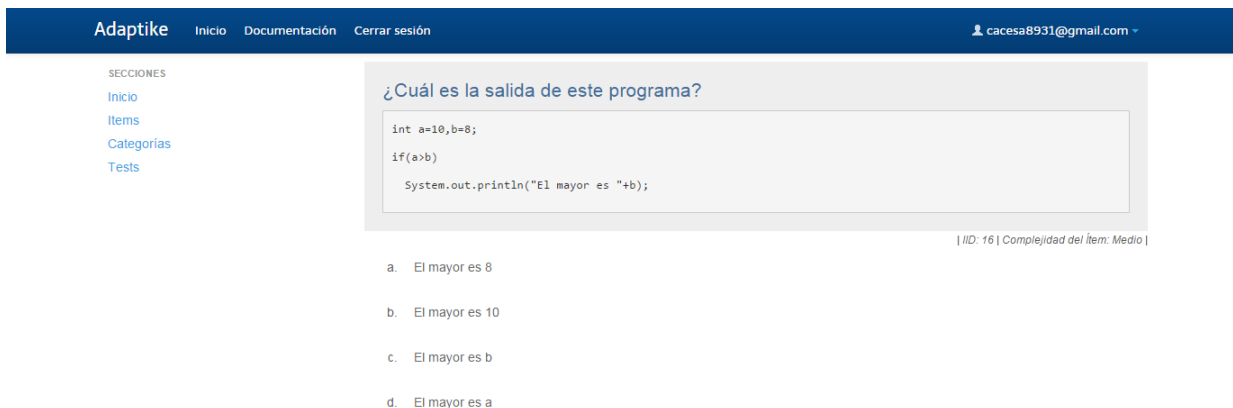


Figura 4. Ejecución del test adaptativo

Esta opción está disponible haciendo click en la tecla D. Al hacer esto se despliega una ventana emergente en la cual se muestra el nivel de conocimiento

actual del individuo, la respuesta correcta y una tabla que muestra los datos de los ítems disponibles para administrar y que participan en la selección del ítem actual.

13	-0.382	-0.077	0.147	-0.0045769422243637
14	-1.102	-0.553	0.021	-0.65270253519873
15	0.847	0.408	0.118	-0.065610583741498
16	-1.205	-0.88	0.082	0.22054313837238
17	-1.422	0.086	0.063	-0.042769432192239
18	-0.901	1.162	0.121	-0.0010950864185105
19	-0.269	0.76	0.129	-0.0011710790135694
20	0.086	-0.36	0.029	-0.00020771256874442
21	-0.76	-1.135	0.122	0.16418983720058
22	-0.966	1.195	0.05	-0.00090659821746518
23	0.178	-1.627	0.149	-0.00039903890550979
24	2.216	-0.58	0.044	-0.0031190219915327

Figura 5. Debug del proceso de selección

Aparece el id del ítem, los parámetros a, b, c y el indicador de Fisher, el cual presenta la probabilidad de que el ítem sea contestado correctamente por el individuo teniendo en cuenta su nivel de conocimiento actual. Se elige el ítem que tenga un máximo en esta probabilidad. Gracias a esto se observa que el id del ítem resaltado coincide con el id del ítem administrado en el test. El comportamiento del test continúa de la misma manera hasta que el usuario alcanza el nivel de conocimiento configurado como criterio de parada del test.

5. CONCLUSIÓN

Se consiguió implementar los algoritmos de selección de ítems y de cálculo del nivel de conocimiento por medio de la estimación por máxima verosimilitud conjunta.

Se requiere la validación de expertos en el área de la psicometría y de la educación para que el sistema propuesto pueda ser aplicado como una herramienta efectiva de evaluación.

REFERENCIAS

- [1]. W. J. van der Linden and C. A. W. Glas, Computerized Adaptive Testing: Theory and Practice. 2000.
- [2]. H. Wainer, N. J. Dorans, R. Flaugher, B. F. Green, and R. J. Mislevy, Computerized Adaptive Testing: A Primer. 2000.
- [3]. A. Binet and T. Simon, "Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux," *Annee. Psychol.*, vol. 11, no. 1, pp. 191–244, 1904.
- [4]. W. Patience, "MicroCAT Testing System Version 3.0.," *J. Educ. Meas.*, vol. 27, no. 1, pp. 82–88, 1990.
- [5]. M. Pommerich, "The nine lives of CAT-ASVAB: Innovations and revelations," *Proc. 2009 ...*, 2009.
- [6]. S. Imai, S. Ito, Y. Nakamura, K. Kikuchi, Y. Akagi, H. Nakasono, A. Honda, and T. Hiramura, "Features of

- J-CAT (Japanese Computerized Adaptive Test).” D. J. Weiss (Ed.), Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing., 2009.
- [7]. C. N. Mills and M. Steffen, “The GRE Computer Adaptive Test: Operational Issues,” in *Computerized Adaptive Testing: Theory and Practice*, Springer Netherlands, 2000, pp. 75–99.
- [8]. Lawrence M. Rudner, “Implementing the Graduate Management Admission Test Computerized Adaptive Test,” in *Elements of Adaptive Testing*, Wim J. van der Linden and C. A. W. Glas, Eds. Springer New York, 2010, p. pp 151–165.
- [9]. P. Rebollo, E. García Cueto, P. C. Zardáin, J. Cuervo, I. Martínez, J. Alonso, M. Ferrer, and J. Muñiz, “[Development of the CAT-Health, the first adaptive computerized test (CAT) for the evaluation health-related quality of life in Spain].,” *Med. Clin. (Barc.)*, vol. 133, no. 7, pp. 241–51, Jul. 2009.
- [10]. B. Ruo, S. W. Choi, D. W. Baker, K. L. Grady, and D. Cella, “Development and validation of a computer adaptive test for measuring dyspnea in heart failure.,” *J. Card. Fail.*, vol. 16, no. 8, pp. 659–68, Aug. 2010.
- [11]. F. Lord, “Individualized Testing and Item Characteristic Curve Theory.,” 1972.
- [12]. J. M. Fernández, *Introducción a la teoría de respuesta a los ítems*. Pirámide, 1997.
- [13]. R. K. Hambleton, H. Swaminathan, and H. J. Rogers, *Fundamentals of Item Response Theory*. 1991.
- [14]. F. Baker, *The basics of item response theory*. 2001.
- [15]. K. T. Han, “WinGen: Windows software that Generates IRT parameters and item responses,” *Applied Psychological Measurement*, 2007. [Online]. Available: <http://www.umass.edu/remf/software/simcata/wingen/homeR.html>. [Accessed: 08-Apr-2015].