

GenTAI: generador de tests adaptativos informatizados

Javier López-Cuadrado, Joaquín M^a Sánchez Fernández

Dpto. Lenguajes y Sistemas Informáticos, Universidad del País Vasco (UPV-EHU)
Aptd. 649, San Sebastián 20080, España.
javilo@si.ehu.es

Resumen: Este artículo presenta una herramienta que permite la generación automática de Tests Adaptativos Informatizados experimentales o a pequeña escala a partir de un banco de ítems que esté calibrado según el modelo logístico de 3 parámetros de la Teoría de Respuesta al Ítem. GenTAI es un software amigable, de fácil uso, intuitivo, portable, escalable, altamente configurable y fuertemente documentado, que al mismo tiempo da respuesta a los objetivos de un test adaptativo. Otros sistemas similares pecan en la calidad de sus interfaces y en la forma de interactuar con el usuario. Como novedad, GenTAI incluye un módulo que permite al usuario elaborar simulaciones para evaluar o comparar los errores de medida obtenidos tras la aplicación de los tests.

Palabras clave: tests adaptativos informatizados, bancos de ítems, herramientas de evaluación.

Abstract: This paper presents a tool that allows the automatic generation of experimental or minor Computerized Adaptive Tests from an item bank that is calibrated by means of the 3-parameter logistic Item Response Theory model. GenTAI is a friendly, easy-to-use, intuitive, portable, scalable, highly configurable and strongly documented software, that fulfils the goals of an adaptive testing at the same time. Some similar systems do not offer enough quality in their interfaces and have a lack of proper ways to interact with the user. GenTAI includes a new feature that allows the user to perform simulations to assess or compare the measurement errors obtained after test administrations.

Key words: computerized adaptive tests, item banks, assessment tools.

1. Introducción

El test es sin duda el principal método utilizado en psicometría para medir las facultades intelectuales, tanto educativas como psicológicas. La aplicación de un test consiste en administrar una serie de preguntas o actividades denominadas *ítems*, que habitualmente tienen un formato de elección múltiple, con el fin de cuantificar para cada sujeto una variable de rasgo o habilidad, como la memoria, la capacidad de aplicar el conocimiento adquirido o la comprensión de textos.

Las teorías de tests proporcionan métodos para la construcción de tests y proveen de modelos matemáticos que facilitan la interpretación y validación de los resultados obtenidos.

Tradicionalmente, el ámbito profesional de la medición psicológica y educativa ha estado dominado por la *Teoría Clásica de los Tests* (TCT). Esta teoría surgió a partir de los trabajos de Spearman (1904; 1907; 1913), si bien no fue hasta mediados de los años sesenta cuando recibió su forma axiomática (Novick, 1966). La TCT se fundamenta en un modelo que establece una relación lineal entre la habilidad del examinado y la puntuación del test. Concretamente, se considera que la puntuación empírica obtenida en la prueba (X) es igual a la suma de dos componentes hipotéticos y desconocidos a priori: la puntuación verdadera o habilidad del evaluando (V) y un cierto error de medida (e). Los inconvenientes más destacables de la TCT son, por una parte, que los supuestos que plantea no se pueden contrastar empíricamente, por otra, que las mediciones de habilidad obtenidas dependen de la

naturaleza del test utilizado, y que, a la inversa, propiedades como la dificultad de una prueba están supeditadas a las características de los sujetos a quienes se les aplique.

La *Teoría de Respuesta al Ítem* (TRI) dará solución a estos (y otros) problemas de la TCT. Se trata de una teoría relativamente joven y en continua evolución (Lord, 1952; Rasch, 1960; Lord y Novick, 1968; Lord, 1980; Hambleton, Swaminathan y Rogers, 1991; van der Linden y Hambleton, 1997) cuyos modelos establecen una relación no lineal entre la habilidad del examinado y una serie de propiedades o parámetros de los ítems que componen el test. Concretamente, estos modelos reflejan la probabilidad de que los sujetos acierten los ítems que se les administran durante la aplicación de una prueba de evaluación.

El modelo unidimensional más utilizado de los que ofrece la TRI es el logístico de tres parámetros (3PL), cuya expresión da la Ecuación 1 (Birnbaum, 1968):

$$P(\theta) = c + \frac{1-c}{1+e^{-aD(\theta-b)}}$$

Ecuación 1 – Modelo de tres parámetros.

Como puede observarse, según este modelo la probabilidad de que un examinado con habilidad θ acierte un ítem (esto es, $P(\theta)$) sólo depende de tres características que son propias del ítem (el poder discriminativo a , la dificultad b , y el pseudoacierto o probabilidad de acierto al azar c) y de un factor de escalado D que, una vez establecido, es constante.

El parámetro de dificultad se corresponde con el valor de la habilidad donde la curva $P(\theta)$ presenta su punto de inflexión. En este punto la probabilidad de una respuesta correcta se encuentra a medio camino entre el pseudoacierto y 1.0. La dificultad se establece utilizando la misma escala que se usa para medir la habilidad de los examinados, por lo general la que está definida en el intervalo (-4, 4) del eje horizontal, siendo el 0 su punto medio. Cuanto mayor es el valor del parámetro b , más difícil es el ítem, de modo que los ítems fáciles aparecerán en la parte izquierda del eje θ , y los difíciles a la derecha.

Por su parte, el parámetro de poder discriminativo del ítem corresponde a la máxima pendiente de $P(\theta)$, la cual se alcanza precisamente en su punto de inflexión. Cuanto mayor es su valor, que difícilmente rebasa el intervalo (0.2, 2), más significativa es el ítem a la hora de evaluar al examinado.

Finalmente, el pseudoacierto se define como la probabilidad de que una persona con poca habilidad (región izquierda del eje horizontal) responda al ítem correctamente.

La sección siguiente expone el concepto de test adaptativo fundamentado en la TRI; la sección 3, que está dedicada al estado del arte en el uso de este tipo de pruebas de evaluación, dará pie a la presentación de GenTAI, herramienta en la que se centrará la sección 4; finalmente, la sección 5 cerrará con las conclusiones y posibles líneas de trabajo futuro.

2. Tests Adaptativos Informatizados

Por estar orientada al ítem, la TRI permite obtener estimaciones de la habilidad de los evaluandos que son independientes del conjunto específico de ítems que se haya administrado. Además, esta teoría ofrece un marco en el que, incluso aunque dos examinados hayan respondido series distintas de preguntas, sus puntuaciones se pueden comparar directamente. Esta circunstancia, impensable en el contexto de la TCT, es la que ha impulsado el desarrollo cada vez más acentuado de tests adaptativos basados en la TRI. Se habla de adaptación en cuanto a que el test se genera dinámicamente dependiendo de las respuestas previas del examinado. Así, si el sujeto falla una pregunta (esto es, el ítem ha resultado ser difícil) la siguiente cuestión será considerablemente más fácil porque no parece razonable seguir insistiendo con preguntas de dificultad alta; y viceversa, como consecuencia de un acierto se administrará un ítem más difícil. La idea subyacente es la de que cuanto más se ajusten los ítems seleccionados a la habilidad real del alumno, tanto más útil será la información que aporten sus respuestas.

Los *Tests Adaptativos Informatizados* (TAI) son la implementación de esta idea en un programa informático que automáticamente presenta los ítems, y recoge y evalúa las respuestas (Wainer, 2000b). Los

elementos que forman un TAI son un algoritmo de selección de ítems, un criterio de finalización del test, un método para estimar la habilidad en base a las respuestas recopiladas hasta el momento, y un banco o base de datos que almacene los ítems calibrados según el modelo de la TRI que se vaya a utilizar.

El algoritmo que define el funcionamiento de un TAI (Muñiz, 1997) es un procedimiento iterativo (Figura 1) que, partiendo de una estimación inicial de la habilidad del examinado, primero evalúa todos los ítems que no han sido utilizados aún en el test con el fin de seleccionar el mejor con respecto a la estimación de la habilidad actual; después administra el ítem elegido y recoge la respuesta; acto seguido computa una nueva estimación de habilidad considerando las respuestas a todos los ítems utilizados hasta el momento; y sigue repitiendo los pasos anteriores hasta satisfacer el criterio de parada establecido. En general, un TAI puede finalizar porque se ha alcanzado un nivel concreto de precisión en la estimación de habilidad, porque se ha utilizado un número de ítems determinado, o porque se ha superado el tiempo límite definido.

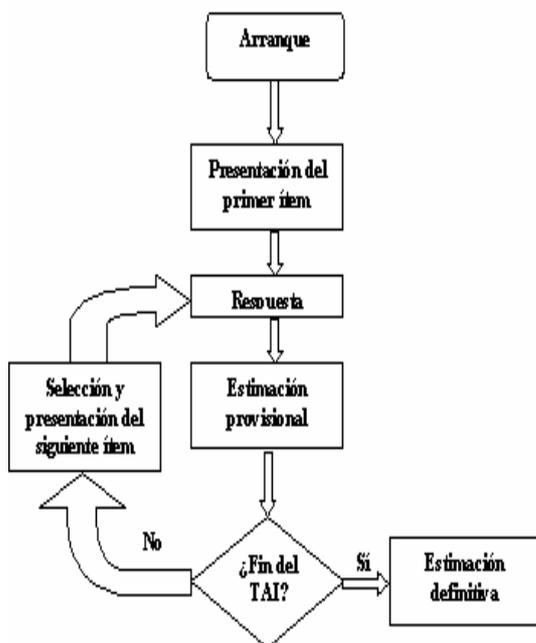


Figura 1 – Algoritmo de administración del TAI.

La eficiencia de un TAI depende de dos procesos complementarios y estrechamente vinculados: el método estadístico utilizado para estimar la habilidad y el criterio de selección de ítems. Es fundamental que las estimaciones de la habilidad del examinado que realice el TAI sean precisas, pues en ellas se basa el correcto funcionamiento del test. Los métodos más utilizados son los de máxima verosimilitud (Lord, 1980) y los bayesianos (Owen, 1975). Por su parte, la selección del siguiente ítem, que se lleva a cabo bajo la suposición de que la estimación de la habilidad del evaluando es precisa, es la tarea responsable de que el TAI sea adaptativo.

En este contexto cabe destacar el criterio de máxima información (Lord, 1980), que consiste en seleccionar, de entre los que aún no han sido administrados, el ítem que proporciona más información para el nivel de habilidad estimado en cada momento. La información aportada por un ítem $I(\theta)$ sólo depende de las propiedades del modelo, concretamente de la curva característica del ítem $P(\theta)$ y de su primera derivada con respecto a θ , y se define como (Birnbbaum, 1968):

$$I(\theta) = \frac{P'(\theta)^2}{P(\theta)(1 - P(\theta))}$$

Ecuación 2 – Función de información del ítem.

La idea subyacente al criterio de máxima información es que cuanto más informativo es un ítem para un valor θ concreto, tanto más precisa hace, después de haber sido administrado, la nueva estimación de la habilidad del examinado; o dicho de otro modo, menos error produce en dicha estimación. Con el fin de evitar que los ítems más informativos se apliquen sistemáticamente en todas las evaluaciones, y en consecuencia los menos valiosos no se administren nunca, todo método de selección de ítems suele implementar algún procedimiento para controlar la sobreexposición de los ítems (Olea y Ponsoda, 2003).

Cuando finaliza la aplicación del TAI se dispone de una aproximación del nivel de habilidad del examinado. No obstante, para que dicha estimación sea precisa y fiable, es necesario que el banco de ítems que se haya utilizado esté calibrado según alguno de los modelos de la TRI. Esto significa que

las características que definen sus ítems (en el caso del modelo 3PL, dificultad, discriminación y pseudoacierto) deben ser conocidas, por ejemplo, por haber sido estimadas mediante algún proceso estadístico (Hambleton y Swaminathan, 1985).

3. Estado del Arte

La idea original de los TAI fundamentados en la TRI data de comienzos de los años setenta (Lord, 1970), y fue implementada por vez primera durante los ochenta a partir del sistema *Computerized Adaptive Screening Test* de la armada estadounidense, precursor de lo que a día de hoy se conoce como CAT-ASVAB (Segall y Moreno, 1999).

En 1990 sólo se administraron unos pocos cientos de TAI, pero esta cifra se incrementó en 1999 hasta superar el millón de aplicaciones, por lo que el crecimiento en la utilización de tests adaptativos se sigue estimando exponencial (Wainer, 2000a). De hecho, en la actualidad existen multitud de instituciones y programas estandarizados de evaluación que incluyen TAI basados en TRI, como el *National Assessment of Educational Progress* o el *Educational Testing Service*, en cuyo marco se desarrollan varios de los más importantes programas de evaluación adaptativa (Rojas Tejada, 2001), como el *Test Of English as Foreign Language*, que se administra en casi 90 países.

Semejante aumento en la aplicación de TAI se debe principalmente a las ventajas que ofrecen sobre los tests tradicionales de lápiz y papel basados en la TCT: entre otras, requieren menos tiempo de administración, realizan estimaciones de habilidad más precisas, y dotan al test de mayor seguridad (Olea y Ponsoda, 2003). No es de extrañar, por tanto, que el catálogo de ámbitos en los que a día de hoy se utilizan los TAI sea tan extenso, y que abarque campos como la medición de aptitudes intelectuales, la selección de personal y las pruebas de admisión, certificación o evaluación educativa.

Prácticamente todas las implementaciones de TAI que se han desarrollado se utilizan a gran escala, y, en la mayor parte de los casos, con fines comerciales. De hecho, apenas se conocen aplicaciones gratuitas para realizar tests adaptativos experimentales o a pequeña

escala, entre las que cabe mencionar DEMOTAC (Renom, 1993) y ADTEST (Ponsoda, Olea y Revuelta, 1994) como ejemplos representativos.

Muchas veces son los propios psicómetras quienes elaboran herramientas generadoras de TAI para uso propio, para que su alumnado pueda crear pequeños tests durante las prácticas en clase, o con algún otro fin experimental. En estos casos, pese a ser muy eficientes en los cálculos y estimaciones de habilidad, los productos finales suelen ofrecer interfaces de usuario muy pobres. Por lo general se trata de software desarrollado en lenguajes de programación imperativa tradicionales (como C, Fortran o Pascal) que limitan la configurabilidad del programa, los formatos de entrada y la interfaz de comunicación. En concreto, no siempre se ofrece la posibilidad de utilizar el ratón como dispositivo de entrada, las pantallas se suelen basar íntegramente en interfaces de matriz de texto al estilo de MS-DOS, y el usuario se ve por lo general obligado a utilizar ficheros de texto ASCII con un formato concreto (establecido de antemano y que no siempre resulta intuitivo) para almacenar su banco de ítems calibrado.

Este texto presenta GenTAI, un sistema de fácil manejo que aporta fundamentalmente mejoras en la interfaz de usuario, utilidades de ayuda y el incremento de posibilidades de configuración, no sólo a nivel de aplicación informática sino también como generador de TAI. Se trata de una herramienta muy orientada al usuario, escalable, portable y de fácil manejo que genera TAI a partir de un banco de ítems calibrado según el modelo de tres parámetros. Un sencillo patrón de interfaz gráfica e interactiva dirige las principales operaciones, lo que, junto con una alta configurabilidad en las opciones y preferencias del usuario, facilita su utilización. Además, y como novedad, GenTAI incluye un módulo de simulación que permite evaluar los errores de medida obtenidos después de la aplicación de los tests.

4. La Herramienta GenTAI

GenTAI se ha desarrollado en Java, un lenguaje de programación orientada a objetos que facilita la construcción de aplicaciones que interactúan con cualquier tipo de base de datos, simplificando la creación de interfaces de usuario basadas en

ventanas, y permitiendo el uso del ratón sobre elementos gráficos dotados de funcionalidad (como botones, listas desplegables o menús). Además, GenTAI es portable, lo que significa que puede ejecutarse en cualquier plataforma, con independencia de cuál sea el sistema operativo usado.

Se trata de un programa muy sencillo de utilizar, pues lo único que se necesita es que el evaluador especifique la ubicación de la tabla con los ítems calibrados según el modelo 3PL, para que el sistema genere de modo sencillo e intuitivo tests adaptativos. Dicha tabla puede estar almacenada, por ejemplo, en una base de datos de tipo Microsoft Access (la más utilizada en entornos particulares, domésticos y educativos a pequeña escala), aunque el software GenTAI está preparado para utilizar cualquier otro gestor de bases de datos. Para ello, lo único que hace falta es modificar, en el código de la aplicación, la directiva de conexión a la base de datos, que por defecto está implementada mediante el establecimiento de un puente *jdbc-odbc* para realizar la conexión con una base de datos Microsoft Access.

Desde el punto de vista programático, la herramienta es fácilmente escalable, pues se ha diseñado con estructuras abiertas y poco solapadas, de manera que la inclusión en el futuro de nuevas funcionalidades y opciones de configuración resulte sencilla.

GenTAI incorpora un novedoso sistema de simulación que automáticamente calibra su funcionamiento y ofrece resultados para un posterior análisis. Asimismo, permite la generación de archivos de texto con históricos de sesión y la presentación de resultados en diversos formatos que pueden complementarse.

El sistema GenTAI proporciona siete módulos o pantallas de presentación. El paso de una sección a otra puede darse mediante los botones establecidos en cada pantalla para ello, aunque también puede efectuarse a través de la barra de menú. Esta barra, común para toda la aplicación, facilita la navegación por las distintas opciones y brinda acceso a las funcionalidades más importantes.

A continuación se presentarán en detalle las características y funcionalidades principales de cada uno de los siete módulos que componen el sistema GenTAI (Figura 2), y que son: pantalla de presentación, preferencias, simulación, configuración del TAI, aplicación o administración del test, presentación de resultados y ayuda.

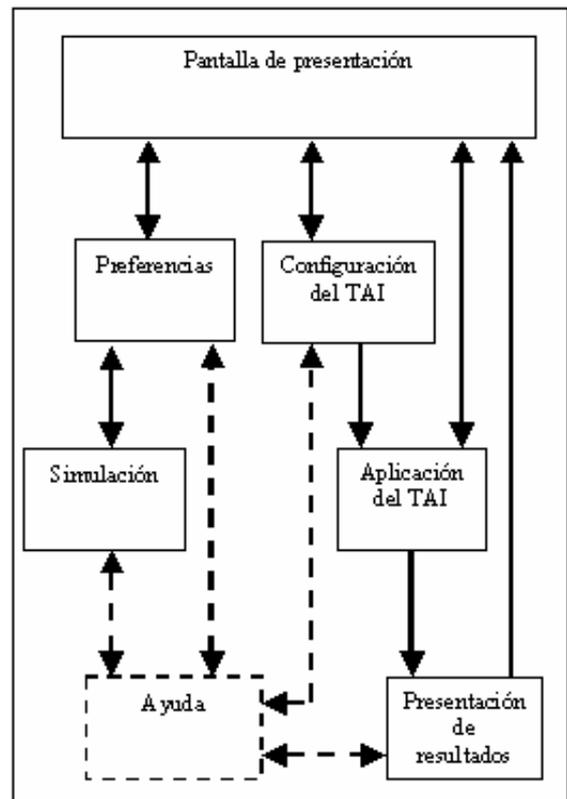


Figura 2 – Pantallas de la herramienta GenTAI.

4.1. La Pantalla de Presentación

La pantalla de presentación (Figura 3) es la vía de entrada a la aplicación, y permite acceder al resto de elementos y funcionalidades que ofrece GenTAI. Opcionalmente puede especificarse en ella un nombre de usuario con el que posteriormente se personalicen los informes y resultados de los tests.

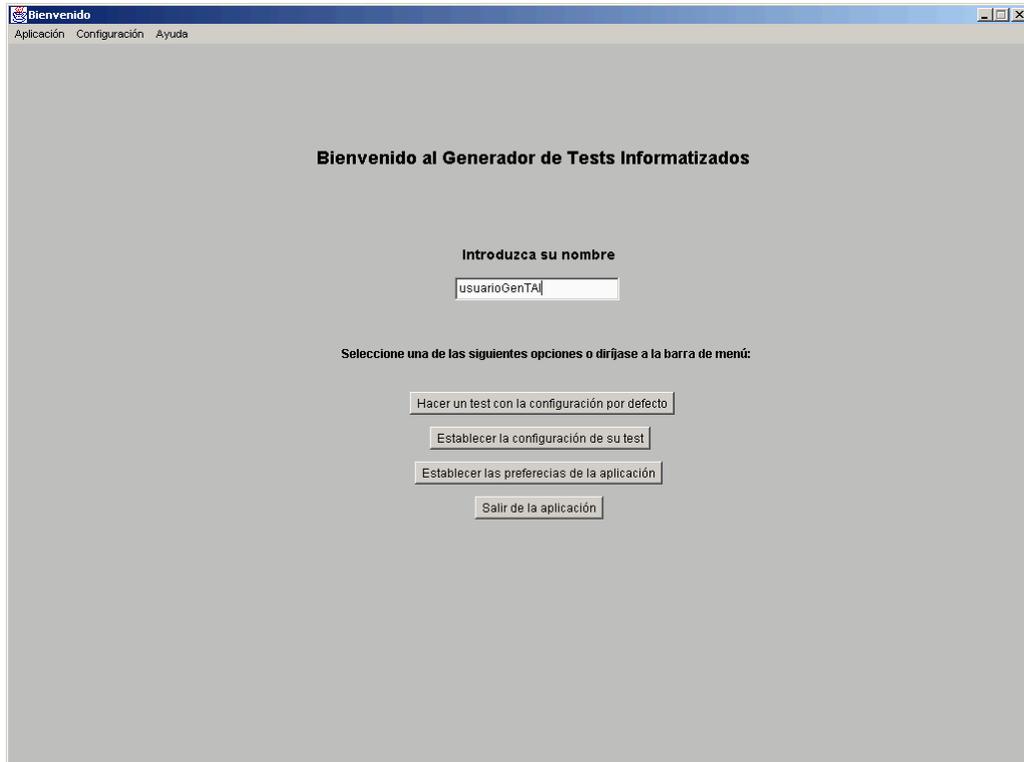


Figura 3 – Pantalla de presentación de GenTAI.

4.2. Preferencias

En esta sección (Figura 4) el usuario puede configurar las distintas opciones que ofrece la herramienta. Con el fin de evitar problemas en el funcionamiento de GenTAI, la pantalla de preferencias establece un estricto control sobre los datos que proporciona el usuario, en particular, como se verá enseguida, sobre la ubicación del banco de ítems. Todas las posibilidades que pueden establecerse en este punto del programa son:

- El tipo de ítems a utilizar: de momento sólo se incluye la posibilidad de utilizar preguntas de elección múltiple con tres o cuatro posibilidades de respuesta, que por otra parte son los más utilizados en las pruebas de evaluación mediante tests.
- La ubicación de los datos: especificación de la base de datos en la que se encuentran los ítems del banco y sus parámetros a , b y c . En el caso particular de que la base de datos sea de tipo Microsoft Access, GenTAI incorpora un sistema de comprobación del formato de las tablas, que

muestra un aviso ante cualquier anomalía, lo que facilita la gestión y mantenimiento de las mismas.

- Restricciones de accesibilidad: GenTAI permite restringir el acceso a las opciones de configuración de los tests, así como al menú de preferencias, por ejemplo para poder realizar numerosas pruebas con un mismo patrón de configuración y evitar que los examinados modifiquen los valores establecidos para la aplicación de los TAI. Para ello, se establece una contraseña que se guardará en un formato cifrado con el fin de impedir su decodificación a partir del fichero que la almacena.
- Establecimiento de los valores por defecto: se facilita la configuración de los TAI que se vayan a realizar directamente desde la pantalla de presentación. También se permite especificar el nombre que GenTAI dará a los archivos de históricos de resultados.
- Opciones de simulación: proporciona el acceso al módulo de simulación y a sus opciones de configuración, elementos de los que se hablará a continuación.

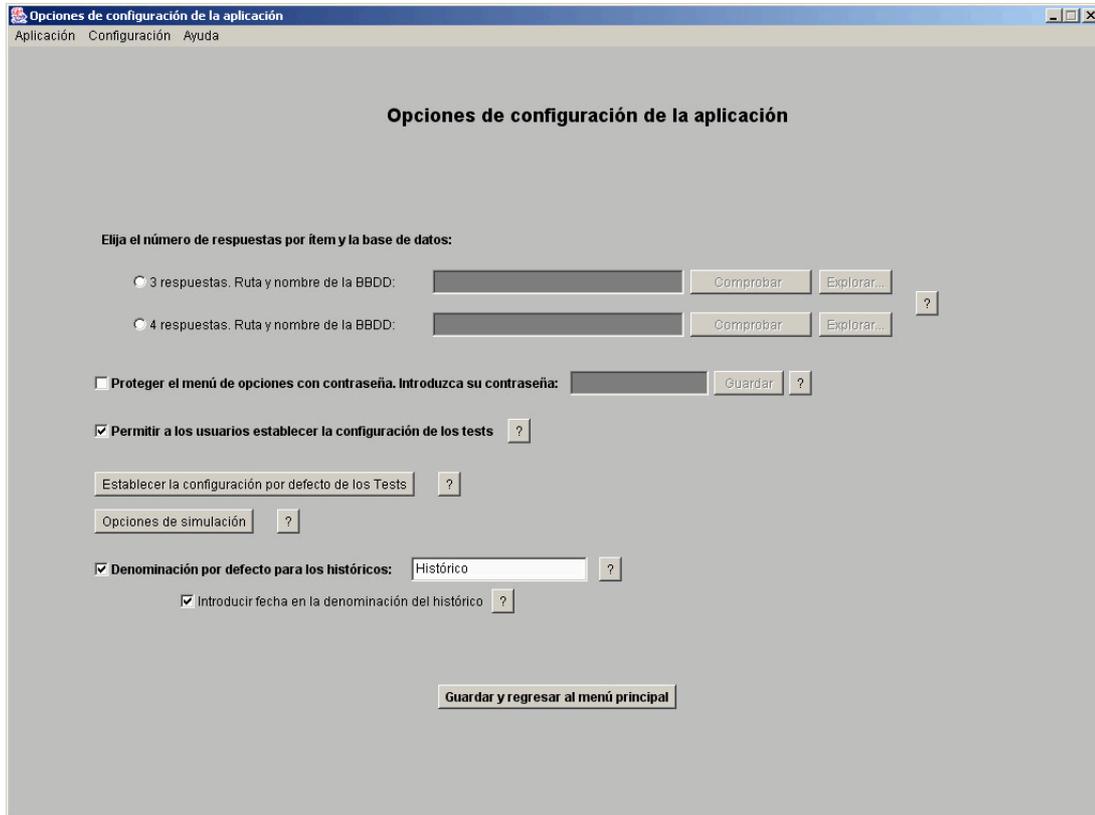


Figura 4 – Pantalla de preferencias de GenTAI.

4.3. Módulo de Simulación

Esta pantalla, que, como se acaba de ver, es accesible desde la ventana de preferencias, sirve para testar el funcionamiento de la herramienta en términos de precisión o error en las estimaciones. Este subproceso es el encargado tanto de recoger las opciones de la propia simulación como de la ejecución de la misma.

El funcionamiento de las simulaciones es muy sencillo: el sistema genera automáticamente una muestra de sujetos cuyas habilidades establece a priori; después simula la administración de un TAI a cada uno de estos sujetos virtuales, por lo que como resultado obtiene una estimación para cada una de las habilidades que ha determinado previamente; por último, una comparación entre las habilidades establecidas en el inicio y las estimaciones recién obtenidas dará al usuario una idea de lo preciso que es el TAI que se ha aplicado a los sujetos simulados.

Los aspectos que pueden configurarse en lo concerniente a la simulación son el tipo de test a administrar, el método de selección del primer ítem, el método de selección de los ítems sucesivos, y el criterio de parada. Estos parámetros de configuración se describirán con detalle en el siguiente epígrafe, pues son exactamente los mismos que se han definido para la administración de los TAI.

Es en lo relativo al formato de los resultados donde surgen las diferencias, dado que el módulo de simulación elabora automáticamente un tipo de informe de salida más específico que los que se pueden obtener tras la aplicación de un TAI. En efecto, una vez finalizada una sesión de simulación, los resultados de las administraciones efectuadas se almacenan, de modo transparente al usuario, en un archivo tabulado de texto plano que facilita su futura importación por parte de programas de cálculo

numérico o análisis estadístico como Microsoft Excel o el paquete SPSS. Entre los resultados computados se encuentran la raíz del error cuadrático medio (RSME), el sesgo (BIAS) y el error estándar (SE) de las estimaciones de habilidad obtenidas durante el proceso de simulación, no sólo para todo el intervalo de habilidades considerado $[-4, 4]$, sino también para diferentes subintervalos de tamaño 0.2 (concretamente, $[-4, -3.8]$, $[-3.8, -3.6]$, etcétera). La Ecuación 3 muestra las fórmulas utilizadas para calcular los índices de precisión del algoritmo TAI utilizado durante la simulación.

$$RSME = \sqrt{\frac{\sum_{k=1}^N (\theta_k^* - \theta_k)^2}{N}}$$

$$BIAS = \frac{\sum_{k=1}^N (\theta_k^* - \theta_k)}{N}$$

$$SE = \frac{\sum_{k=1}^N (\theta_k^* - \bar{\theta}^*)}{N} \quad \text{donde} \quad \bar{\theta}^* = \frac{1}{N} \sum_{k=1}^N \theta_k^*$$

Ecuación 3 – Medidas de precisión en la simulación.

En todas estas fórmulas, N es el número de sujetos simulados del intervalo o subintervalo considerado a los que se les ha aplicado un TAI, θ_k la habilidad (establecida de antemano por el propio módulo de simulación) del sujeto k -ésimo, y θ_k^* la estimación que el TAI ha computado para la habilidad del sujeto k -ésimo. En el caso ideal en el que el algoritmo de aplicación de la prueba resultase absolutamente preciso, estos dos últimos valores serían idénticos.

La administración simulada de TAI se efectúa utilizando varios cientos de sujetos con niveles de habilidad distribuidos uniformemente a lo largo del intervalo $[-4, 4]$, a quienes se les aplican los tests siguiendo la configuración especificada, y tomando ítems de un banco que puede ser generado íntegramente por GenTAI o extraído de una base de datos externa. Para el primer caso, la propia herramienta crea un banco de 3000 ítems cuyos parámetros siguen una distribución uniforme (por defecto, $U(0.4, 1.4)$ el de discriminación, $U(-4, 4)$ el de dificultad y $U(0, 0.2)$ el de pseudoacierto).

4.4. Configuración del TAI

Si bien es posible comenzar la administración de un TAI directamente desde la pantalla de presentación (en cuyo caso se utilizarán los valores que estén establecidos por defecto), lo más habitual será hacerlo después de haber visitado la pantalla de configuración del TAI (Figura 5). Este módulo, en el que se han invertido grandes esfuerzos para favorecer la escalabilidad de la aplicación (esto es, la inclusión futura de nuevas funcionalidades), permite establecer los métodos y criterios a seguir durante la posterior administración de los TAI. Al igual que ocurre en el módulo de preferencias (y en general, en cualquier pantalla en la que el usuario ha de introducir algún dato), a la hora de capturar esta información GenTAI realiza un control exhaustivo de los datos de entrada. El objetivo es evitar la inserción de valores no deseados o inválidos, por lo que, en el supuesto en que se dé uno de estos casos, el sistema muestra por pantalla un aviso individualizado y automáticamente restaura el valor erróneo a uno aceptable (por ejemplo, al asignado por defecto).

En el módulo de configuración del TAI puede especificarse:

- El tipo de prueba a desarrollar: en este momento se ofrecen tres tipos de test, a saber, TAI fácil (en el que los ítems que se presentan tienen menor dificultad que la óptima con el fin de reducir la ansiedad del examinado), TAI clásico (en el que se sigue estrictamente el método de máxima información) y TAI difícil (en el que los ítems son ligeramente más difíciles de lo esperable).
- El método de selección del primer ítem: con el propósito de que el algoritmo de aplicación del TAI disponga de una estimación inicial de la habilidad del examinado, el usuario debe establecer un valor numérico exacto, o en su defecto especificar un rango de valores para que sea el sistema quien elija aleatoriamente la habilidad de partida dentro de dicho intervalo.
- El método de selección de los siguientes ítems: se trata del criterio de máxima información, que puede combinarse con uno de los métodos de control de sobreexposición de los ítems que se han implementado, y que son el 5-4-3-2-1, el *randomesque* y el progresivo (Olea y Ponsoda, 2003).

- El criterio de parada: puede especificarse si el algoritmo de aplicación del TAI debe finalizar tras la administración de un número específico de ítems (criterio de longitud fija), cuando se ha alcanzado una precisión determinada (criterio de longitud variable), o como resultado de una combinación de ambos (criterio mixto). De modo complementario, es posible definir un tiempo máximo para la conclusión de la prueba.
- El o los formatos de presentación de los resultados: existen cinco formas distintas de mostrar los resultados, ninguna de ellas excluyente del resto. Se trata de la habilidad pura (θ estimada), una transformación a la escala [0,10], un gráfico de evolución, un histórico y el porcentaje de aciertos. Aunque es en la pantalla de configuración donde se especifica cuál o cuáles de los diferentes formatos se desea para visualizar los resultados, tras la finalización de cada prueba se le da al usuario nuevamente la

posibilidad de elegir entre los cinco modos de presentación.

4.5. Administración del TAI

Tanto si se ha establecido una configuración específica como si se van a utilizar los valores por defecto, la aplicación del TAI tiene lugar en una serie de pantallas en las que, a partir de la configuración determinada para el test actual, se muestran los ítems apropiados para la evaluación y se recogen las respuestas del examinado. Mientras que el usuario se limita a seleccionar la respuesta que considera correcta para cada uno de los ítems que GenTAI le presenta, el sistema procesa las operaciones necesarias para la ejecución del algoritmo de administración del TAI. Las tareas concretas que realiza este módulo son las que se enumeran a continuación:

Configuración
Aplicación Configuración Ayuda

Establezca la configuración de su Test

Elija el tipo de TAI: TAI Fácil ?

Elija el criterio de selección del primer ítem: Al azar -> Rango de selección entre [] y [] ? Establecido por usted -> Su habilidad inicial es: 0.0 ?

Elija el criterio de parada: Fijo -> Número de ítems: 10 ? Variable -> Error máx.: [] ? Mixto -> Número de ítems: [] y error máx.: [] ? Tiempo -> Duración: 10 minutos ?

Elija el criterio de selección de ítems: Máxima Información ? Sobreexposición: 5-4-3-2-1 ?

¿Cómo quiere que el programa muestre los resultados? Habilidad pura ? Transformación a escala [0, 10] ? Gráfico de evolución ? Histórico ? Porcentaje de aciertos ?

[Empezar el Test!] [Volver al menú principal] [Cargar la configuración por defecto]

Figura 5 – Pantalla de configuración del TAI.

- Seleccionar el ítem más adecuado: atendiendo a los criterios de selección de ítems establecidos en la configuración del test, se selecciona del banco de ítems el siguiente que se va a mostrar. Esta tarea, que depende de la estimación actual de la habilidad del examinado, de los ítems que ya han sido administrados y de si se requiere o no la utilización de métodos de control de sobreexposición, se anticipa a la respuesta del usuario, ya que efectúa los cálculos durante los segundos que tarda el examinado en responder a cada ítem. De esta manera se consigue que el proceso no afecte a la interacción con el sistema (en concreto, ralentizando la presentación de los ítems), y mucho antes de que el evaluando responda un ítem GenTAI ya tiene decidido cuál va a ser el siguiente que le va a presentar si acierta, y cuál si elige una respuesta incorrecta.
- Recalcular la habilidad estimada del sujeto: dado que la selección del siguiente ítem a administrar requiere conocer de antemano la estimación de habilidad actual del sujeto, esta tarea, que sólo depende de las características de los ítems administrados hasta el momento y de sus respuestas, se incluye en la anterior. Por lo tanto, el sistema puede conocer con antelación no sólo qué ítem deberá administrar a continuación tanto si el sujeto acierta el actual como si no, sino que además puede disponer del nuevo valor de habilidad estimado para cada caso. Para obtenerlo, GenTAI utiliza el método de máxima verosimilitud con la optimización propuesta por Herrando (1989), evitando así divergencias en el proceso de cálculo al suponer de antemano que el sujeto ha respondido correctamente un ítem considerablemente fácil y ha fallado uno de dificultad extrema.
- Comprobar los criterios de parada: antes de realizar los cálculos precisos para la selección de ítems y actualización de la habilidad estimada para el examinado, el sistema debe comprobar si es necesario o no administrar un nuevo ítem, o lo que es lo mismo, detectar si hay que dar por concluido el TAI. Esta tarea se intercala en el proceso de selección de ítems, mediante la inclusión de rutinas que permiten controlar si se cumple alguno de los criterios de finalización establecidos en la fase de configuración. Si se identifica el final del TAI, se invoca al módulo de presentación de resultados (del que se hablará en el siguiente epígrafe) para que muestre al usuario su evolución y/ o puntuación.
- Informar al usuario: además del enunciado y las posibles respuestas de los ítems, y dependiendo de qué criterios de parada se hayan habilitado en la configuración de la prueba, el módulo de aplicación del TAI también presenta al examinado información relativa al estado de cumplimiento de los criterios de finalización. Concretamente, GenTAI puede mostrar, en la parte superior derecha de la pantalla, el número de ítems respondidos sobre el total (si el TAI es de longitud fija), un mensaje avisando de que el test está a punto de finalizar (cuando el error de la estimación actual de habilidad está próximo al máximo permitido en los TAI de longitud variable) y, en la esquina inferior derecha, los minutos restantes para la finalización del test (en el caso de que se haya establecido un tiempo límite). Estos mensajes pueden ir acompañados de un color: verde si aún quedan bastantes ítems por administrar; amarillo si presumiblemente faltan unos pocos; y rojo si el final de la prueba es inminente.

4.6. Presentación de Resultados

Una vez finalizada la administración del TAI por el cumplimiento de alguna de las condiciones de parada, se da paso a la interfaz de presentación de resultados de la prueba. Antes de proceder a la visualización de los mismos se da la oportunidad al examinado de modificar los formatos de presentación de resultados que haya seleccionados en la configuración por defecto de los tests, y que son:

- Habilidad pura: se trata de la estimación final obtenida por el sistema, y que viene representada como un número real dentro de la escala $[-4, 4]$.
- Transformación al intervalo $[0,10]$: consiste en obtener un escalado de la habilidad pura con el fin de ofrecer una visión más familiar del resultado del TAI.
- Gráfico de evolución: este formato de presentación utiliza un modelo matemático a partir del que se obtiene un diagrama (Figura 6) en el que se muestra la evolución del valor

estimado de la habilidad del examinado a lo largo de la administración del test.

- Porcentaje de aciertos: también se ofrece la posibilidad de visualizar este índice, pese a no ser el más apropiado para evaluar la habilidad en los TAI, ya que, independientemente de la destreza de los examinados, siempre rondará el 50-60%.

- Históricos de evolución: GenTAI permite elaborar dos formatos de documento (tabulado y comentado) para recoger los ítems administrados durante la aplicación del TAI, así como las sucesivas (re)estimaciones de habilidad y los errores de medida asociados. El histórico tabulado (Figura 7) utiliza un algoritmo que encasilla toda la información ajustando la longitud del texto a los diferentes anchos de columna, mientras que el histórico comentado (Figura 8) reproduce la administración del TAI en una secuencia de acciones.



Figura 6 – Gráfico de evolución.

Ambos tipos de histórico de aplicación del test pueden, además de visualizarse por pantalla, ser almacenados en disco en un archivo con formato de texto plano, cuyo nombre (*.txt) será dependiente de lo que se haya especificado en las preferencias de GenTAI.

HISTÓRICO TABULADO DE LA SESIÓN							
Núm. ítem	Cód.	a	b	c	A/F	Habilidad	Se
1	68	1.139449412	-0.193731498	8.42311E-4	F	-0.2943156567440752	0.9692057844087496
2	107	1.326244844	-0.586077456	0.061085849	A	-0.38067104697983273	0.7568732819271601
3	83	1.55142209	-1.074617756	0.016003906	A	-0.2472624992323744	0.5999530229143838
4	151	1.196070573	-0.61098056	0.07253029	A	-0.05772695375696265	0.5849412035255286
5	63	1.17349862	-0.520157476	0.072182379	F	-0.3993700429704905	0.47778722946121477
6	84	1.360938657	-1.088778344	0.132389294	A	-0.33484189212636745	0.44068044033464704
7	123	1.259568651	-0.791650136	0.191836299	A	-0.25491916467997255	0.427271001874242
8	97	1.309548795	-0.50819422	0.271489608	F	-0.46006869895674946	0.3819800549265836
9	72	1.12179953	-0.804101688	0.111331523	A	-0.3897103346000403	0.3752780503628862
10	121	1.463209062	-1.342448195	0.096362194	F	-0.6547387879553315	0.3409031265558883
11	139	1.708841981	-1.399334696	0.136756493	F	-0.924179069158023	0.32902953856454253
12	54	1.3619175	-1.620532853	0.067586291	A	-0.8895135965609576	0.31038139991615105
13	57	1.18942842	-1.521408734	0.017285684	A	-0.850080695907979	0.3035781633550496
14	118	1.407655989	-1.632740257	0.146305734	F	-1.0291517185638488	0.2933407183167267
15	31	1.415363957	-1.943784906	0.013357952	F	-1.2139478399959864	0.2851328313702739
16	71	1.471211479	-1.895199438	0.106094546	A	-1.186043528648218	0.276684886733061
17	92	1.207082849	-1.543626209	0.167995239	A	-1.146166829721634	0.2697903707984915
18	69	1.034678749	-1.401287881	0.024234748	F	-1.225625818582217	0.2637129225781939
19	125	1.285260171	-1.635914182	0.270216987	F	-1.335976403553654	0.25954823848473724
20	59	1.21865219	-1.765800958	0.268486587	A	-1.3082773884905585	0.25282132629589055

Figura 7 – Histórico tabulado del TAI.

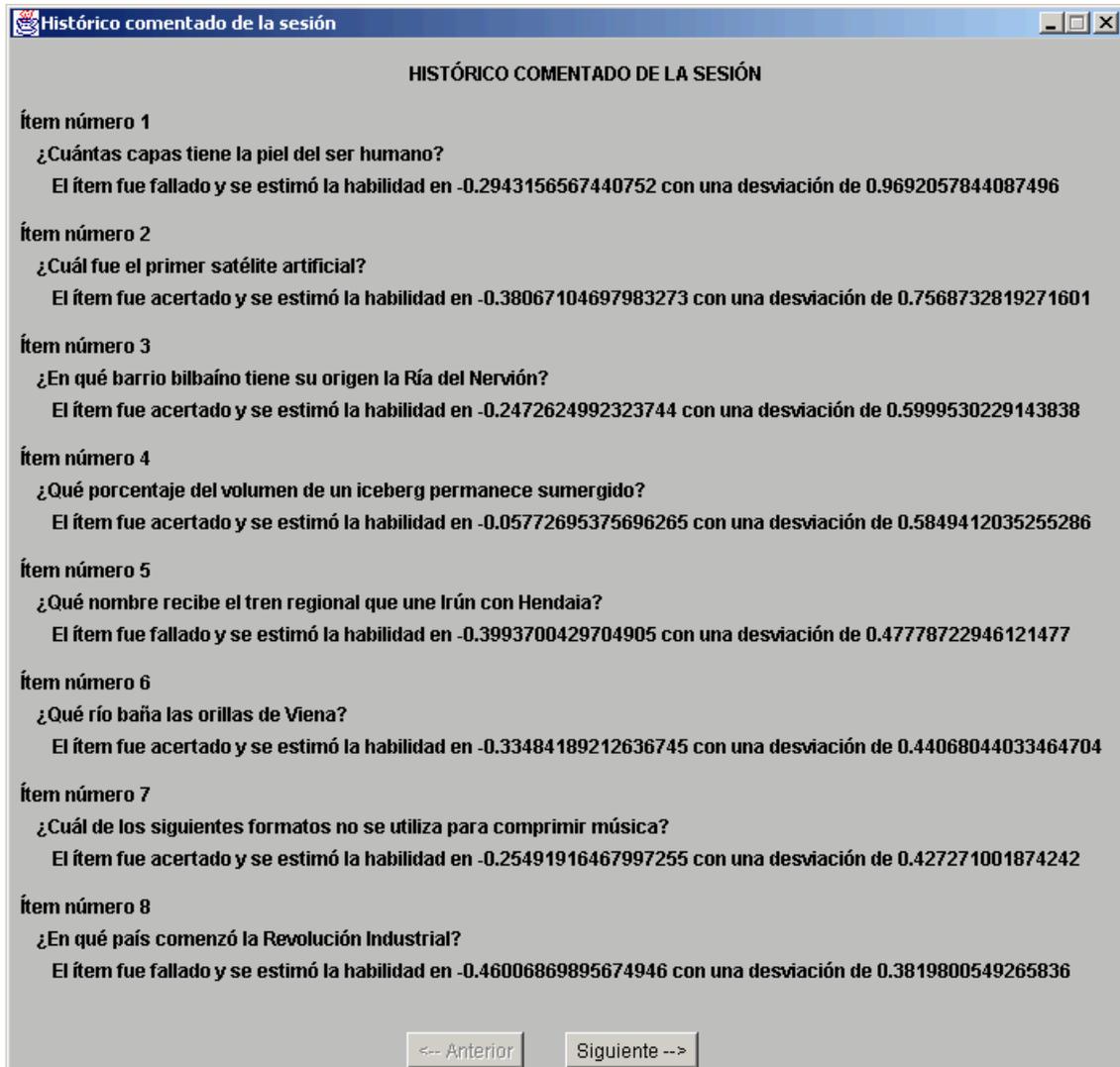


Figura 8 – Histórico comentado del TAI.

4.7. Sistema de Ayuda

La aplicación está dotada de un sistema de ayuda que tiene como finalidad solventar las dudas que el usuario pueda tener en algún punto de la interacción con el sistema.

Por una parte, la ayuda puede mostrarse de modo contextual a través de una serie de botones que hay habilitados en determinados puntos de la pantalla

para tal efecto. Cuando el usuario pulsa uno de ellos, una ventana le muestra información concreta acerca del tema relacionado con el elemento junto al que se encuentra dicho botón (Figura 9). La ayuda contextual se ha definido básicamente para cada una de las opciones de configuración, para los aspectos técnicos relacionados con la naturaleza del TAI, y para los formatos de presentación de resultados que ofrece el sistema.

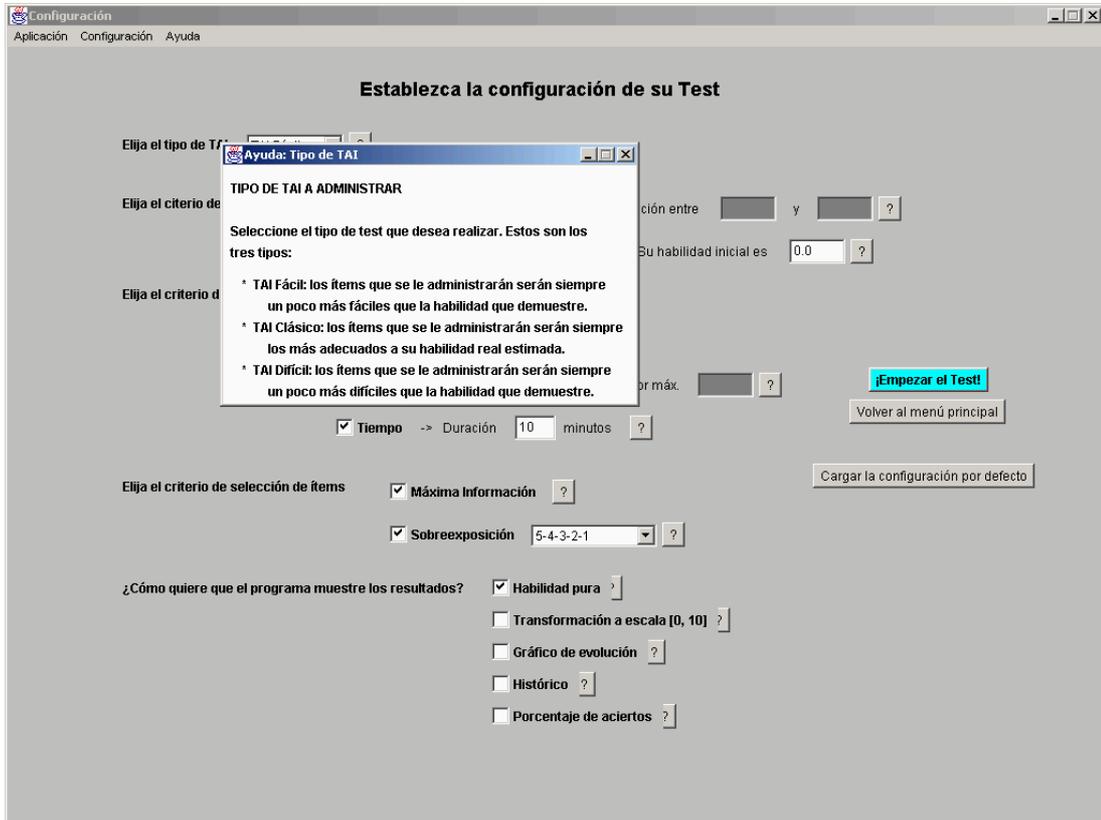


Figura 9 – Ayuda contextual de GenTAI.

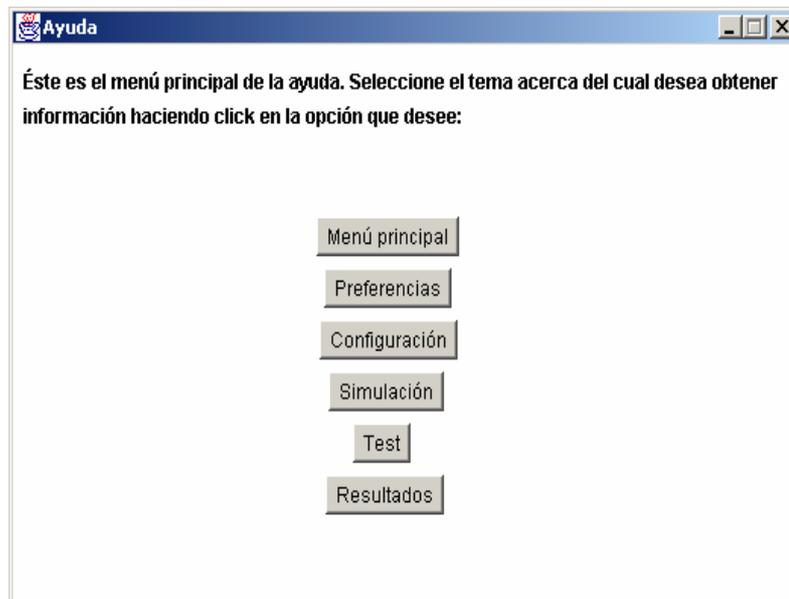


Figura 10 – Módulo de ayuda temática.

Asimismo, también se incluye un módulo de ayuda temática (Figura 10), habilitada en todo momento durante la interacción con el sistema, que se encuentra accesible en la barra de menú de la aplicación.

Este componente ofrece al usuario la posibilidad de navegar por los contenidos de ayuda, que se encuentran organizados por temas según las diferentes funcionalidades de GenTAI (preferencias, configuración, simulación, test y resultados). Este tipo de ayuda está dirigida al usuario inexperto, con lo que se espera posibilitar el uso de esta herramienta por un amplio colectivo de usuarios.

5. Conclusiones

Este artículo ha presentado GenTAI, un generador de tests adaptativos informatizados para utilizar a pequeña escala o en entornos experimentales por un usuario, si bien pueden existir varias instancias del programa, cada una de ellas operada por una persona.

El único requisito para poder utilizar la herramienta es disponer de una Máquina Virtual Java (JVM – *Java Virtual Machine*) instalada en el ordenador. La JVM ofrece independencia con respecto al sistema operativo, es gratuita y puede descargarse directamente de la página web de Sun (2005), aunque GenTAI ofrece la posibilidad de instalar el software JSDK (*Java Software Development Kit*) 1.4, también gratuito, que la incluye.

GenTAI es una herramienta muy amigable, gráfica, de fácil e intuitivo manejo, portable a cualquier plataforma y altamente configurable, y que como novedad incluye un módulo de simulación que permite al usuario evaluar los errores de medida producidos tras la aplicación con diferentes configuraciones de los tests.

GenTAI es una aplicación especialmente diseñada para facilitar la inclusión de nuevas funcionalidades y opciones en el futuro.

- En primer lugar, se prevé ampliar el catálogo de formatos de ítems, para que los tests no se limiten a utilizar preguntas de elección múltiple de tres o

cuatro posibilidades de respuesta.

- Por otra parte, se considera la posibilidad de guardar automáticamente información del usuario entre sesiones.
- Al mismo tiempo, está prevista la inclusión de otros tipos de tests adaptativos, como los Tests Autoadaptados Informatizados (Rocklin y O'Donnell, 1987), en los que se le ofrece al examinado la oportunidad de decidir si el siguiente ítem a administrar será más fácil o más difícil.
- Otra de las líneas de desarrollo futuras es la de añadir más métodos de control de la sobreexposición de los ítems, en particular el método progresivo generalizado (Eggen, 2001), el de Sympson y Hetter (1985) y alguna de sus variantes.
- Por último, también se ha contemplado la posibilidad de orientar la herramienta hacia un uso basado en web. De hecho, el lenguaje de programación utilizado para implementar GenTAI es muy adecuado para dicha labor. Aunque podría tratarse de una conversión tan simple como la inclusión de un *applet*, por motivos de seguridad en las pruebas y de control en los accesos al banco de ítems, está prevista la creación de una aplicación web implementada mediante páginas *jsp* y *servlets*.

Por el momento GenTAI sólo se ha utilizado de manera experimental con un banco de ítems piloto, habiéndose producido resultados satisfactorios. No obstante, en cuanto concluya la calibración mediante el modelo 3PL de un banco de ítems para la evaluación de conocimientos de la lengua vasca, y en consecuencia se disponga de estimaciones precisas de los parámetros de sus ítems (López-Cuadrado, Armendariz y Pérez, 2005), se pretende utilizar la aplicación con usuarios reales.

Referencias

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. Statistical theories of mental test scores. F.