

# RECURSOS TECNOLÓGICOS PARA LA EVALUACIÓN PSICOEDUCATIVA

Consuelo Belloch

Unidad de Tecnología Educativa (UTE). Universidad de Valencia

Las innovaciones tecnológicas y los nuevos desarrollos teóricos en el campo de la evaluación psicoeducativa, han propiciado diferentes modos de integrar el ordenador en el proceso. De este modo, aunque los test más utilizados actualmente en nuestro contexto siguen siendo los de lápiz y papel, se producen nuevos avances e investigaciones que nos alejan del concepto de test formado por un conjunto fijo de ítems, presentados en un orden predeterminado y en donde únicamente se valoran los aciertos o errores cometidos, acercándonos más al concepto de test flexible y adaptado al contexto o a las características de los evaluados, en donde es posible no solo atender a los aciertos, sino también a determinadas variables ligadas a la respuesta al ítem (ej. tiempo de reacción, acciones realizadas, etc.).

## 1. Evaluación Informatizada

En la evaluación informatizada orientada al diagnóstico (Computer-Based Assessment) se incluyen las técnicas evaluativas que utilizan el ordenador en la estimación de niveles de habilidad o destrezas en contextos diagnósticos y de aprendizaje.

Al igual que ha ocurrido con otras aplicaciones de las herramientas informáticas, la evaluación informatizada ha sufrido transformaciones importantes y rápidas debido a los recientes avances tecnológicos en el campo de la informática, pero esto no podría producirse sin los desarrollos e innovaciones que se han producido en la Medición Educativa, sobre todo los modelos teóricos de la Teoría de Respuesta al Ítem, y la Psicología Cognitiva. Bunderson, Inouye y Olsen (1989) distinguen cuatro etapas o generaciones en la evaluación asistida por ordenador:

- **1ª Generación: *Evaluación Informatizada*.** Se desarrollan los Tests Informatizados, administrándose los exámenes tradicionales compuestos por preguntas de opción múltiple mediante el ordenador.
- **2ª Generación: *Evaluación Adaptativa*.** Se desarrollan los Tests Adaptativos Informatizados, en los que se presentan las preguntas, una a una, adaptando las preguntas siguientes a las respuestas previas del sujeto.
- **3ª Generación: *Evaluación Continua*.** Se desarrollan los Sistemas Tutoriales que integran la estimación de los cambios en la trayectoria del aprendizaje de los sujetos.
- **4ª Generación: *Evaluación Inteligente*.** Se desarrollan los Sistemas Tutoriales Inteligentes, basados en la Inteligencia Artificial. Mediante estos sistemas se generan perfiles de los resultados de los estudiantes con base en los conocimientos y procedimientos de inferencia.

Tanto para una evaluación de corte más tradicional como en los nuevos modelos evaluativos, el uso del ordenador como medio para facilitar el proceso de evaluación ha sido utilizado, atendiendo a las potencialidades que brinda este instrumento, facilitando:

- **La edición, administración y proceso de los datos en pruebas de evaluación.** Mediante el uso de programas informáticos se editan y diseñan, en múltiples ocasiones, los instrumentos de recogida de la información (cuestionarios, tests, escalas de observación, etc.). Además, en un primer momento, la capacidad del ordenador para procesar grandes cantidades de información de forma rápida, permitió que los ordenadores se utilizaran principalmente en procesos de corrección y emisión de informes mecanizados, en los que se interpretan las puntuaciones obtenidas en las pruebas.
- **El diseño y presentación de ítems.** Es posible, mediante el ordenador, la presentación de ítems en el que se muestren contenidos a partir de diferentes modalidades de información (textual, gráfica, sonora), así como la presentación de los componentes que conforman el ítem dinámicamente. Además, se dispone de un mayor control sobre la presentación del test. Estas nuevas potencialidades que brindan los ordenadores en cuanto al contenido y presentación de los ítems, amplía sustancialmente los rasgos, conocimientos y destrezas que pueden ser evaluados y son, además, la base de muchos estudios evaluativos sobre los procesos y estrategias que utilizan los sujetos en la resolución de problemas.
- **Sistemas de evaluación dinámica y adaptada a los sujetos.** Los tutoriales inteligentes y los tests adaptativos informatizados, permiten adaptar los procesos de evaluación al desempeño de los sujetos. Mediante el uso del ordenador, es posible analizar no solo los resultados obtenidos, sino el proceso seguido, de modo que se establece una interacción dinámica entre el usuario y el ordenador, adaptándose este último a las características del primero. La utilización de bancos de ítems, con sus correspondientes valores métricos (índices de dificultad, validez, discriminación, etc.) permiten establecer pruebas evaluativas personalizadas, con la posibilidad de la comparación entre diferentes pruebas.
- **Evaluación de productos y procesos.** Una de las posibilidades más interesantes, en cuanto al uso del ordenador en los procesos evaluativos, consiste en analizar no tanto el producto, esto es los resultados obtenidos en las pruebas evaluativas, sino el proceso que sigue el sujeto para realizar las distintas tareas, es decir, las estrategias y los procesos cognitivos que se utilizan ante la tarea o prueba. Este tipo de análisis de los procesos cognitivos superiores, sería muy difícil de llevar a cabo sin el uso del ordenador, por no decir, imposible, ya que la posibilidad de este medio para almacenar cualquier tipo de información y realizar un seguimiento puntual de las diversas acciones realizadas por el estudiante, permite obtener información sobre estos procesos y estrategias, con fines tanto diagnósticos como formativos.
- **Evaluación y diagnóstico de personas con discapacidad.** Diversos autores (García L., 1990; Toledo, 1985; Negre, 1995) señalan las posibilidades de las herramientas informáticas para la realización del diagnóstico psicopedagógico en niños que tienen problemas motores y de comunicación graves.

*“Es muy difícil realizar una valoración psicopedagógica del niño con PCI, ya que las dos áreas más afectadas son la de comunicación verbal y la manipulativa, y son precisamente éstas, las que componen la mayoría de tests y pruebas psicopedagógicas; consecuentemente los resultados obtenidos no siempre se corresponden con el estado real del alumno. Por este motivo los ordenadores, ofreciendo un medio de comunicación y*

*reduciendo al mínimo las actuaciones motoras, pueden ser de gran utilidad como fuente de determinadas destrezas". (Negre, 1995).*

Las graves dificultades de comunicación, que presentan personas con diferentes patologías, impedía conocer sus verdaderas capacidades intelectuales y de aprendizaje. La evaluación se basaba principalmente en la observación de sus acciones, pero, en estos casos, es posible encontrar grandes capacidades que no se traducen en acciones, dados los problemas físicos que padecen. La incapacidad de los métodos de evaluación de lápiz y papel para conocer el estado real de estas personas, ha podido llevar en el pasado a realizar diagnósticos equivocados, en donde sujetos inteligentes han sido considerados poco inteligentes. En este sentido, el ordenador, y las nuevas tecnologías de ayuda a la comunicación, brindan nuevas posibilidades en la evaluación de estas personas, ya que es posible tanto llevar a cabo pruebas psicopedagógicas, como acceder a los procesos mentales y cognitivos están implicadas, en la realización de las diferentes tareas.

En conclusión, podríamos decir que el uso del ordenador ha permitido nuevos enfoques, estrategias y objetivos en los procesos de evaluación psicoeducativa.

## **2. Breve Reseña Histórica**

Para realizar esta breve reseña histórica hemos partido de revisiones que otros autores han realizado sobre el tema (Renom, 1993; Mills y Stocking, 1996; Olea y Hontangas, 1999). Las primeras aportaciones de la informática en la evaluación mediante tests, se produjeron en la corrección e interpretación de los tests convencionales (Computer Based Tests Interpretations -CBTI-), pretendiéndose agilizar el proceso y proporcionar informes sin errores o sesgos de los profesionales.

Ya en los años 30 se utilizó el ordenador para la corrección automática del Strong Vocational Insters Blanck (SVIB) utilizando para ello tarjetas perforadas. En la década de los 40 se realiza la corrección automática del Minnesota Multiphasic Personality Inventory (MMPI). Una década después, en los 50, comienzan a utilizarse las hojas de respuesta electrónica y las lectoras ópticas por el Measurement Research Center de Iowa, lo que facilita enormemente la introducción de los datos.

Durante la década de los 60 y 70 se produce una proliferación en Estados Unidos en el desarrollo de sistemas automatizados de interpretación de tests como: 16PF, Rorschach y WAIS. Asimismo se realiza la primera aplicación informatizada de tests de personalidad e intereses: MMPI - Minnesota Multiphasic Personality Inventory, EPI -Eysenck Personality Inventory-, 16PF y Strong-Campbell Interest Inventory. Al mismo tiempo, durante los años 70, se comienzan a informatizar los tests de aptitudes e inteligencia: Wais, el Test de las Matrices Progresivas de Raven, el Test de Figuras Enmascaradas y el Peabody Picture Vocabulary Test. En 1977 Cliff y colaboradores realizan un primer estudio sobre una adaptación informatizada, basada en la Teoría de Respuesta al Item, de la escala Stanford-Binet. Mejorada en 1980 por Cudek y colaboradores al conseguir reducir el 50% de los ítems empleados en la primera versión.

La década de los 80 puede ser conocida en este ámbito como la década en que se produce el desarrollo de tests informatizados siguiendo los criterios de los tests convencionales y se inicia la

evaluación adaptativa informatizada. Así:

- El Educational Testing Service desarrolló sistemas multimedia para pruebas de certificación y nivel (Bridgeman, Bennet y Swinton, 1986).
- El Waterford Testing Center realizó el diseño de un conjunto de tests informatizados para evaluar el conocimientos de alumnos de primaria en Lectura, Lenguaje y Matemáticas (Slawson, 1986).
- McDermott (1980) desarrolló un conjunto de programas para evaluar los posibles problemas intelectuales, motivacionales o conductuales que pueden estar en la base de determinados problemas de aprendizaje.
- Kent y Albanese realizan en 1987 la primera evaluación con tests adaptativos informatizados de estudiantes de medicina de segundo curso.
- En 1988 se comienza la introducción paulatina de exámenes mediante TAI en institutos de Estados Unidos.
- El laboratorio Armstrong de las fuerzas armadas norteamericanas desarrolla la batería de tests informatizados Cognitive Abilities Measurement (CAM), con la que se intentan medir habilidades de procesamiento de la información que anteriormente se estudiaban en los laboratorios experimentales (Kyllonen y Cristal, 1989).
- En gran Bretaña se desarrolla la batería MICROPAT, para la evaluación de las destrezas que intervienen en el trabajo de los pilotos aéreos, incluye tests adaptativos informatizados y la generación automática de tests por ordenador (Bartram, 1995).

En 1993 La Universidad Autónoma de Baja California (UABC) desarrolló un Sistema Computarizado de Exámenes (SICODEX) para la administración informatizada, a través de la red, de un test de lápiz y papel, el Examen de Habilidades y Conocimientos Básicos (EXHCOBA), aplicándose desde 1993 a 1995 a distintas universidades de Méjico, Estados Unidos y Canadá (Backhoff, et al., 1999). El EXHCOBA se diseñó con el propósito de disponer de una prueba normativa que permitiera seleccionar a los mejores alumnos que desean ingresar en la UABC. En el examen todas las preguntas son de opción múltiple con cuatro posibles respuestas, pero cada evaluando responde a un tests con preguntas distintas, lo que aumenta la seguridad de la información. Los resultados obtenidos en la aplicación de la versión informatizada del EXHCOBA, como prueba de admisión, diagnóstico y como instrumento de investigación, indican que:

- Se han observado muy buenos niveles de eficiencia, fiabilidad y facilidad de uso.
- Los alumnos evaluados muestran consistentemente su preferencia por la versión informatizada.
- Las versiones informatizada y de lápiz y papel, se mostraron equivalentes.

En 1994 se crea el programa NCLEX/CAT con el que The National Council of State Boards of Nursing reemplaza sus certificaciones basadas en exámenes tradicionales por tests adaptativos informatizados.

En nuestro país, la evolución ha sido similar a la observada en otros contextos socioculturales, si bien ha estado marcada por un cierto retraso tecnológico –que se ha ido superando en los últimos años, desde el advenimiento de la democracia y la consiguiente apertura exterior-. De este modo, las aplicaciones más usuales han estado orientadas inicialmente a la corrección de pruebas y a la edición de informes psicopedagógicos. Empresas como COSPA, TEA, EOS o MEPSA han desarrollado diversas aportaciones que se sitúan en el marco de la oferta de servicios a centros escolares y psicopedagógicos.

En la UVEG se ha desarrollado un sistema de generación de programas para la aplicación de

pruebas a discapacitados SEA (Ordoñez, 2000). Las ventajas del sistema radican en que no es un programa cerrado, sino un generador de aplicaciones que, además, incorpora diversas herramientas para la individualización atendiendo a las necesidades de los usuarios discapacitados –diferenciados por tipos y grados de discapacidad-.

En conjunto, las soluciones que se han ofrecido en nuestro país han producido y aportado novedades que enriquecen el panorama de las aplicaciones informáticas en la evaluación de los individuos (Hontangas, 1999).

### **3. Tests Informatizados**

Las innovaciones tecnológicas y los nuevos desarrollos teóricos en el campo de la evaluación psicoeducativa, han propiciado diferentes modos de integrar el ordenador en el proceso. De este modo, aunque los test más utilizados actualmente en nuestro contexto siguen siendo los de lápiz y papel, se producen nuevos avances e investigaciones que nos alejan del concepto de test formado por un conjunto fijo de ítems, presentados en un orden predeterminado y en donde únicamente se valoran los aciertos o errores cometidos, acercándonos más al concepto de test flexible y adaptado al contexto o a las características de los evaluados, en donde es posible no solo atender a los aciertos, sino también a determinadas variables ligadas a la respuesta al ítem (ej. tiempo de reacción, acciones realizadas, etc.).

Algunos autores (Olea, Ponsoda y Prieto, 1999) reservan el término Tests Informatizados a los instrumentos de evaluación que cumplen dos requisitos:

1. Se conocen sus propiedades métricas, obtenidas por un modelo matemático y
2. Se administran y corrigen mediante el ordenador.

Las primeras aportaciones de los tests informatizados, consistían en aplicar, corregir y emitir resultados de pruebas, siguiendo los criterios métricos de la Teoría Clásica de los Tests (TCT) en su aplicación y corrección, por medio del ordenador, aunque como indican Julio Olea y Vicente Ponsoda "*algunos se fundamentan en la calibración que se realiza a partir de los modelos de la Teoría de Respuesta al Ítem (TRI)*". (1998b: 163)

Básicamente suponen la traducción a soporte informático de tests de lápiz y papel. La administración de las pruebas puede realizarse de forma escrita (pruebas de lápiz y papel) o directamente sobre el soporte informático. Al mismo tiempo, se van creando y diseñando nuevos tests para ser administrados únicamente desde el ordenador.

En la primera fase de informatización, los tests eran administrados del modo convencional, pero su corrección se realizaba de forma automática mediante el ordenador. Es necesario, por tanto, después del pase de la prueba introducir los datos en el ordenador bien manualmente o por medio de lectoras ópticas, posteriormente, el ordenador se encargará de realizar la corrección y emisión de informes. Esta modalidad de uso del ordenador en la evaluación ha sido, y es actualmente, muy utilizada en la aplicación de pruebas colectivas en el ámbito psicoeducativo.

Diversas entidades ofrecen servicios de informatización en la aplicación de los tests (TEA, CEPE, COSPA, EOS, etc.), algunas de ellas, disponen de baterías estándares de pruebas para ser aplicadas a los diferentes niveles educativos, compuestas por pruebas de aptitudes, de personalidad, sociométrico, etc., sobre las que se emiten resultados de los análisis estadísticos realizados tanto a nivel individual, como de grupo. Asimismo, las editoriales de tests vienen ofreciendo junto a los

mismos, programas de corrección automatizada.

Es necesario tener presente que la corrección automatizada de tests ha sido utilizada en algunos casos incorrectamente. Las directrices de la American Psychological Association orientan sobre la utilización adecuada de los informes generados por el ordenador indicando que estos "*deben utilizarse únicamente en conjunción con el juicio profesional*" (1986: 12). Es decir, el evaluador no puede, ni debe, emitir los resultados del proceso de evaluación únicamente atendiendo a los resultados obtenidos en el test, sino que estos le sirven como ficha técnica, sobre la que puede integrar o confrontar otras informaciones disponibles sobre la persona a evaluar (historial, características, rendimiento, etc.). Si bien este no es un tema específico de los tests de corrección automática, puesto que puede procederse igualmente en tests corregidos manualmente, dadas las características del medio informático se puede producir más habitualmente de lo deseado.

No obstante, no debemos minimizar las ventajas, fundamentalmente de tipo instrumental, que tiene el uso del ordenador en la corrección automatizada de tests (Roid, 1991). Entre ellas cabe destacar las siguientes:

1. Evitar errores en la corrección y baremación de las pruebas.
2. La rapidez de procesamiento de los datos.
3. Facilita la combinación de resultados en tests compuestos por subtests.
4. Facilita el proceso estadístico y el análisis de perfiles, que manualmente sería difícil de realizar. Como el análisis comparativo entre diferentes grupos, perfil medio del grupo, etc. que pueden resultar interesantes para la dinámica del profesor en el aula.

Con el fin de comprobar la equivalencia entre las dos modalidades en la presentación de los ítems (test de lápiz y papel y test informatizado) se han analizado las propiedades métricas de los tests, comparando los resultados obtenidos en las dos modalidades de presentación de un mismo test. Los estudios analizados presentan en algunos casos conclusiones contrapuestas en las que puede pesar el tipo de test y software utilizado. Sin embargo, en general parecen indicar que no existen diferencias significativas entre ambas modalidades de presentación, excepto en determinados rasgos aptitudinales y actitudinales.

Así, en diversos estudios no se hallaron diferencias significativas entre ambas versiones -salvo en la rapidez de ejecución, que presenta puntuaciones más elevadas en la modalidad de administración con ordenador- (Mazzeo y Harvey, 1988; López Pina y cols. 1990; Mead y Drasgow, 1993). La revisión realizada por Rolls y Feltham (1993) con tests de personalidad concluye indicando que las diferencias encontradas son mínimas o nulas.

Sin embargo, Wise y Plake (1989) señalan que en la mayoría de los estudios de equivalencia sobre tests de rendimiento se obtienen puntuaciones medias superiores en la modalidad de lápiz y papel. En esta misma línea, investigaciones realizadas con pruebas aptitudinales (Bartram, 1989; Kyllonen, 1991) señalan que alrededor del 10% de la varianza puede atribuirse al medio de presentación de la prueba.

Diversos análisis del tema (Muñiz y Hambleton, 1999; Sykes e Ito, 1997) señalan la importancia de que la administración del test por ordenador no lleve a la eliminación de algunos ventajas que presentan los tests convencionales de lápiz y papel. Por ello es importante que los tests informatizados permitan la revisión y modificación de respuestas y que la distribución del tiempo de cada ítem sea la deseada por el sujeto. En los tests informatizados se ha producido un cambio de orientación respecto a los tests convencionales, pasando de estar centrados en el resultado o producto, a atender principalmente a los procesos cognitivos o estrategias de resolución.

| TESTS CONVENCIONALES INFORMATIZADOS   |   |
|---|---|
| Ventajas  | Limitaciones  |
| <ul style="list-style-type: none"> <li>• Permite evaluar aptitudes no considerados en los test tradicionales, como son la aptitud espacial dinámica, la habilidad para coordinar múltiples fuentes de información o la capacidad para estimar el tiempo.</li> <li>• Permite la presentación aleatorizada de los ítems.</li> <li>• La generación de ítems con una dificultad previsible.</li> <li>• Mayor control sobre el objetivo de la prueba, tanto en la rapidez de ejecución como en la precisión de la respuesta, o bien en la combinación de ambas.</li> <li>• El feedback producido a partir de un determinado resultado es inmediato.</li> </ul> | <ul style="list-style-type: none"> <li>• Tests largos que requieren muchos ítems para alcanzar el grado de precisión (fiabilidad) necesaria en la estimación de la medida.</li> </ul> |

La administración de nuevos tests por ordenador, suponen un salto cualitativo en la aportación de la informática al proceso evaluativo, puesto que en este caso, el ordenador, además de ser un instrumento en el que se utiliza su potencialidad de almacenamiento de datos y rapidez de cálculo, permite medir determinados procesos que serían difíciles de analizar con los test de lápiz y papel ya que trabaja con diferentes tipos de información (visual, auditiva, imágenes en movimiento, etc.) que enriquecen profundamente los contenidos a evaluar. Además, es posible tener un mayor control tanto en el tiempo de exposición de las preguntas o ítems como en el almacenamiento de los procesos y tiempos de la respuesta.

En este sentido, *"el ordenador va a permitir utilizar ítems más complejos y cercanos a la realidad, que incluso se puede simular, lo que conllevaría posibles mejoras en la validez predictiva, al acercarse más el test a la realidad criterial que se pretende predecir"*. (Muñiz y Hambleton, 1999).

Estas nuevas posibilidades, junto con los nuevos planteamientos e innovaciones en la Psicología Cognitiva, han permitido el desarrollo de los TIs. Estas pruebas no pretenden únicamente obtener una puntuación del rasgo medido, sino analizar los procesos cognitivos involucrados en la solución de las tareas o ítems, partiendo de una idea fundamental: los procesos cognitivos son la base del nivel de rendimiento, por lo tanto su análisis nos permitirá conocer el proceso o procesos responsables de un mal rendimiento. Por tanto estas pruebas tienen una clara orientación hacia la evaluación formativa y diagnóstica. Estos acercamientos enfatizan la importancia de las estrategias y estilos de resolución de problemas, haciendo hincapié en las condiciones que deben incorporarse en los ítems para suscitar determinados procesos cognitivos (Prieto y Delgado, 1996).

Para la gestión del Banco de Items, únicamente se especifica el número de ítems que deben conformar el test no incluyéndose algoritmos de selección de ítems, puesto que en los TIs se presentan a todos los sujetos los mismos ítems al igual que sucede con los tests convencionales.

#### **4. Teoría de Respuesta al Ítem (TRI)**

La base del desarrollo de los Sistemas Integrados de Evaluación Informatizada (SIEI) se produce con la integración de las aportaciones de los modelos basados en la Teoría de Respuesta al Ítem (TRI) y los avances informáticos, permitiéndose la creación de tests informatizados basados en esta teoría.

La TRI tiene como objetivo obtener mediciones que no varíen en función del instrumento utilizado, disponer de instrumentos de medida que no dependen de los objetos medidos, es decir, que sean invariantes respecto a los sujetos evaluados y avances técnicos como funciones de información de los ítems y del test, errores típicos de medida diferentes para cada nivel de la variable medida y el establecimiento de bancos de ítems con parámetros estrictamente definidos. La TRI se centra en las propiedades particulares de cada ítem, mientras que la TCT se dirige a las propiedades de la puntuación global. Los supuestos de la TRI son:

1. La **unidimensionalidad del rasgo latente**. Es decir que los ítems que constituyen un test deben medir sólo una aptitud o rasgo.
2. La **independencia**. Es decir que las respuestas de un examinado a cualquier par de ítems son independientes y no existe relación entre las respuestas de un examinado a diferentes ítems.

Los modelos basados en la TRI relacionan a sujetos e ítems de modo interactivo lo que permite localizar al mismo tiempo en un continuo psicológico que representa a la variable a sujetos e ítems, el proceso de medición se puede representar como la localización de personas e ítems en una línea recta (Wright y Stone, 1979; Wright y Master, 1982). Así, la posición de las personas en la línea dependerá de sus respuestas a los ítems del test, del mismo modo los ítems tendrán distintas localizaciones dependiendo de su nivel de dificultad.

Ya que para llevar a la práctica el modelo de la Teoría de Respuesta al Ítem, se hace indispensable el uso del ordenador (Olea, Ponsoda y Prieto, 1999), puesto que se requiere ejecutar algoritmos matemáticos muy complejos para:

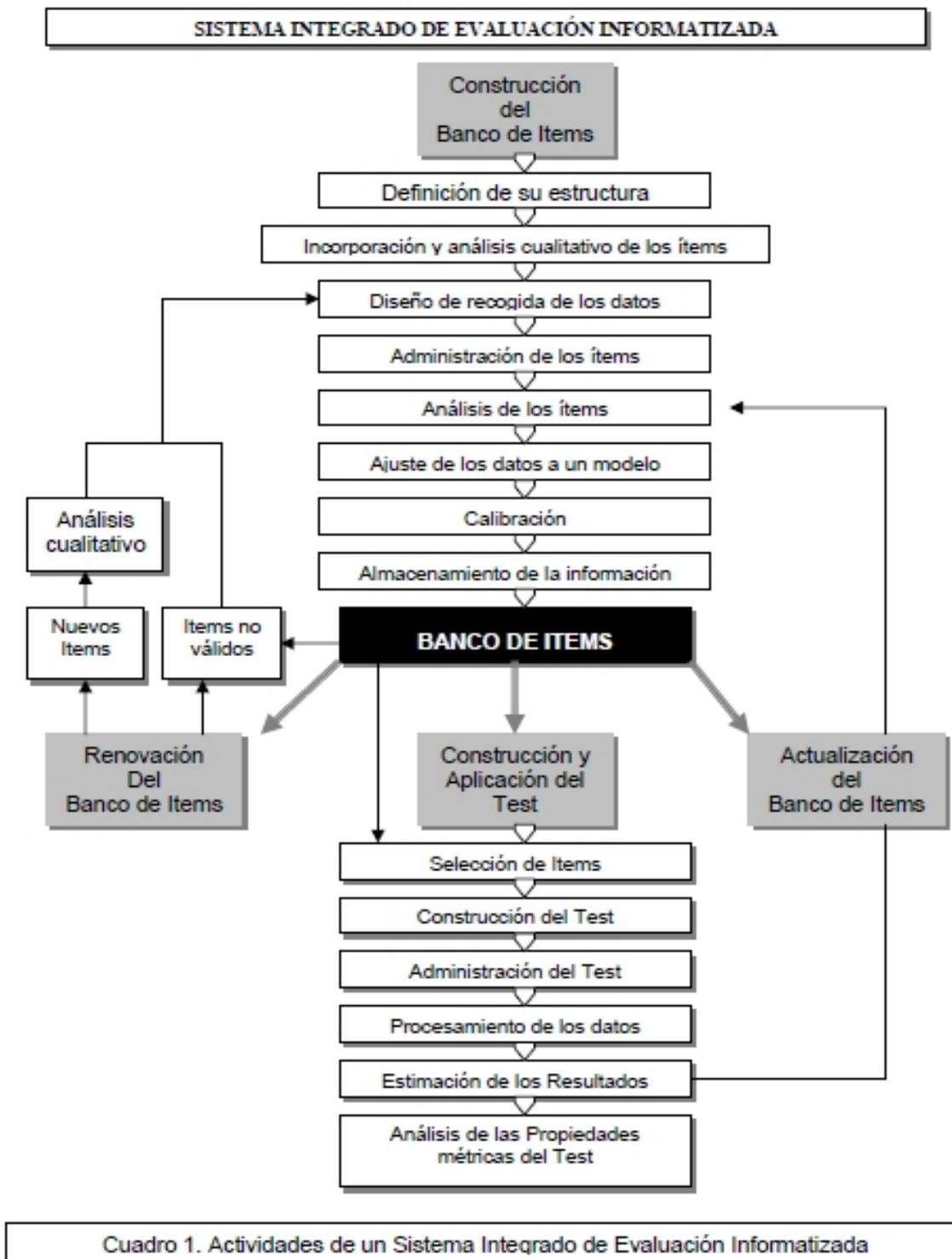
- los procedimientos de estimación de parámetros,
- comprobar en cada caso los supuestos de los diferentes modelos y el ajuste del modelo a los datos empíricos.
- Analizar la precisión de las estimaciones y la contribución de cada ítem a dicha precisión.

Los Sistemas Integrados de Evaluación Informatizada, son sistemas que pretenden realizar mediante el ordenador, las distintas tareas relacionadas con un sistema de evaluación, desde la creación del banco de ítems a partir del cual se realizará la construcción del test hasta la emisión de resultados y estimación de parámetros.

Estos sistemas parten del uso de bancos de ítems a partir de los cuales se extraen los ítems que conformarán la prueba. El banco de ítems es entendido dentro de un concepto dinámico, de modo que puede ser actualizado o renovado según las necesidades y resultados obtenidos en su aplicación. Como señalan Hambleton, Swaminathan y Rogers (1991) la invarianza de los parámetros de los ítems en los modelos TRI permite incorporar nuevos ítems sin necesidad de utilizar la misma muestra de sujetos que se ha utilizado previamente para los que ya se encuentran en el banco.

El banco de ítems es la base sobre la que se sustentan todas las demás tareas propias de los SIEI. Un banco de ítems es un conjunto amplio de ítems almacenados junto con sus características (métricas y de contenido), para que posteriormente puedan seleccionarse aquellos que se adapten a

las condiciones deseadas para un determinado test.



Cuadro 1. Actividades de un Sistema Integrado de Evaluación Informatizada

La información métrica que deberá almacenarse va a depender del marco teórico seguido en la construcción del banco de ítems (TCT o TRI). Los desarrollos de la TRI han producido un verdadero auge de la utilización de bancos de ítems.

Barbero realiza un análisis de las diferentes descripciones de un banco de ítems, entresacando de las mismas tres características que lo definen:

"a) Consiste en una colección relativamente grande y estructurada de ítems que miden un rasgo o dominio de conocimiento bien definido.

b) Los ítems se encuentran almacenados en un ordenador junto con sus características tanto de contenido como métricas, entre las que se incluyen las estimaciones de sus parámetros obtenidas utilizando el mismo modelo de respuesta, por lo general un modelo de la TRI. No obstante, también se pueden construir desde la perspectiva de la teoría clásica de los tests (TCT).

c) Las estimaciones de los parámetros han sido transformadas a una escala común mediante un proceso de equiparación." (1999: 66)

Para Barbero, la construcción de un banco de ítems presenta una serie de fases o pasos a realizar:

### 1. Definición de la estructura.

- Para determinarla es necesario tener presente:
  - el rasgo o característica a medir,
  - el tipo de población sobre la que vamos a realizar la medición,
  - las formas de categorización necesarias para poder posteriormente seleccionar los ítems (áreas de contenido, procesos cognitivos, etc.).
  - la información métrica a almacenar en función del marco teórico que se va seguir (TCT o TRI).

### 2. Desarrollo y análisis cualitativo de los ítems.

- La creación de los ítems puede realizarse tanto manualmente como utilizando el ordenador para una generación automática de ítems. Para el desarrollo de ítems es conveniente tener presente una serie de reglas básicas: (Gronlund, 1985)
  - La relevancia del ítem viene marcada por su ajuste al objetivo o capacidad que pretende medir y por su nivel de representatividad del dominio a evaluar.
  - Es necesario evitar la ambigüedad en las formulaciones de los ítems.
  - Los enunciados de los ítems no deben ser excesivamente largos.
  - Se debe ajustar el vocabulario utilizado en la formulación de los ítems al nivel de los estudiantes.
  - No utilizar estructuras gramaticales excesivamente complejas.
  - Las instrucciones deben ser claras.
  - Evitar elementos o ilustraciones que puedan confundir al estudiante.
  - No realizar preguntas con sesgo racial, étnico o sexual.
- Posteriormente, se debe llevar a cabo un análisis cualitativo de los ítems con el fin de detectar posibles problemas en la forma o contenido de los mismos (ej. Juicio de expertos).

### 3. Elección del diseño de recogida de datos.

- Indicación del número de muestras necesarias en la administración y los ítems a aplicar en cada una de ellas
- Determinación del sistema de anclaje a utilizar:
  - Anclaje de ítems (los sujetos de todas las muestras responden a un conjunto de ítems).
  - Anclaje de sujetos (todos los ítems se aplican a un grupo de sujetos).

4. Administración de los ítems.
  - Administración de los ítems, realizada generalmente sin el ordenador, ya que resulta muy costoso por el número de sujetos e ítems a responder.
  - Procesamiento de las respuestas.
5. Ajuste de los datos a un modelo. Se puede utilizar tanto el marco teórico de la TCT como la TRI, aunque este último es el más adecuado puesto que sus modelos aportan estimaciones de los parámetros de los ítems independientes de la muestra de sujetos sobre la que se realiza la calibración.
6. Calibración. Mediante este proceso se obtendrá la información que se introducirá en el banco de ítems. Para ello es necesario realizar:
  - La estimación de los parámetros de los ítems de acuerdo con el modelo ajustado.
  - El análisis de la fiabilidad de las estimaciones.
  - El proceso de equiparación para situar en una escala común las estimaciones de los parámetros.
7. Almacenamiento de la información de forma estructurada en un banco de ítems. Esta información puede ser de tres tipos:
  - Información general. Se trata de incluir la categorización de todas aquellas informaciones que nos permitan posteriormente seleccionar el ítem (contenido, proceso cognitivo, etc), y también las informaciones que permitan su presentación (formato, etc.) y corrección (puntuación, respuesta correcta, etc.).
  - Contenido. No solo se deberá almacenar el contenido del ítem, sino también otras informaciones adicionales a presentar e información sobre el feedback que se desea presentar.
  - Estimaciones métricas del ítem. Como el índice de dificultad, discriminación, función de información del ítem, etc.

Como indican Molina, Sanmartin y Pareja (1999) el desarrollo, utilización y mantenimiento de bancos de ítems se facilita enormemente mediante el soporte informático que permite el almacenamiento, organización, proceso, recuperación y selección de su información de una forma rápida y eficaz. Al mismo tiempo, estos sistemas posibilitan un uso integrado de esta información, con otras tareas propias del proceso evaluativo. (Baker 1989).

De este modo, estos bancos de ítems serán el punto a partir del cual se construyan los test, que pueden ser en el SIEI aplicados y corregidos mediante el ordenador. Los bancos de ítems, además aseguran, si el proceso se realiza correctamente, que las pruebas evaluativas cumplan las propiedades métricas clásicas que debe cumplir todo test: unidimensionalidad de los ítems utilizados en cada estimación, fiabilidad o precisión de la medida, y validez en cuanto a que se mide lo que pretende medir.

#### 4.1. TAIs

Un Test Adaptativo Informatizado (TAI) *"permite la aplicación de ítems diferentes a cada evaluando: aquellos que resultan más efectivos para estimar de forma precisa su nivel de habilidad, es decir, los que reducen un mayor grado de incertidumbre sobre el nivel de rasgo del evaluando"* (Olea y Ponsoda, 1998:167). Este tipo de evaluación informatizada y adaptativa parte de los supuestos métricos que aporta la Teoría de Respuesta al Ítem, permitiendo calibrar el banco

de ítems y obtener sus parámetros de dificultad, discriminación y pseudoazar.

Los TAIs permiten, además, la comparación entre diferentes tests, pues el resultados de sus mediciones se expresan en la misma métrica.

Para poder aplicar un TAI es necesario disponer de un banco de ítems, con sus correspondientes características métricas. De este banco se irán seleccionando los ítems a aplicar, de forma individual -uno a uno-, que mejor se ajustan al nivel mostrado por el sujeto en las respuestas previas, de modo que cada respuesta del sujeto a un ítem implica una nueva estimación de su nivel de habilidad aumentando progresivamente la precisión de la estimación. La finalización del test se produce cuando se cumple un determinado criterio de parada, generalmente vinculado a la obtención de un nivel de precisión adecuado. Como indican Muñiz y Hambleton "*la gran ventaja de esta estrategia de adaptar la prueba al nivel de la persona examinada es que se mejora la precisión de la medida y se ahorra tiempo de aplicación al acortarse los tests*". (1999: 27)

Diversos autores como (Millman, 1984; Renom 1997; Renom y Doval 1999; Muñiz y Hambleton, 1999) señalan ventajas e inconvenientes en el uso de los TAIs.

| TESTS ADAPTATIVOS INFORMATIZADOS   |  |
|--|--|
| Ventajas   | Limitaciones   |
| <ul style="list-style-type: none"> <li>• Los TAIs son las pruebas que aportan mayor precisión en la estimación de la medida.</li> <li>• Se reduce el número de ítems en la aplicación y en consecuencia el tiempo de aplicación.</li> <li>• Permiten la incorporación de nuevos contenidos utilizando formatos de ítem multimedia</li> <li>• Se reduce el error típico de medida a lo largo del continuo de medida.</li> </ul> | <ul style="list-style-type: none"> <li>• No es posible revisar ni modificar las respuestas dadas.</li> <li>• Se precisa disponer de grandes bancos de ítems para mantener la seguridad de los TAIs.</li> <li>• Es difícil mantener la seguridad de los tests puesto que los ítems que más discriminan son aplicados con mayor frecuencia que los otros.</li> <li>• La selección de los ítems se realiza a partir de sus propiedades estadísticas, siendo difícil realizarla atendiendo a la validez de contenido.</li> <li>• Exigen un mayor esfuerzo en la construcción y aplicación que los tests convencionales.</li> </ul> |

En relación a la aceptación social de estos sistemas de evaluación es posible vislumbrar determinados problemas. Así, por ejemplo si se opta por la utilización de Test Adaptativos Informatizados en pruebas de Selectividad, ¿cómo se explicaría que es posible comparar los resultados obtenidos por dos personas que han sido evaluadas con pruebas diferentes?, ¿cómo se podría entender socialmente que una prueba con más aciertos, tuviera menor o igual puntuación a una prueba con menos aciertos?. Sin lugar a duda, este proceso de evaluación que choca las visiones más arraigadas en los procesos evaluativos, por ello, el profesor como agente evaluador debe poder responderse a estas preguntas a partir de los conocimientos de estos sistemas de evaluación y de las teorías y procedimientos métricos sobre los que se sustentan.

A pesar de las ventajas que presentan los TAIs su implantación en los procesos evaluativos educativos ha sido reducida, para algunos autores esto se debe principalmente a la complejidad que supone el desarrollo de Bancos de Items adecuados.

## 4.2. Test Óptimos

Los tests óptimos son aquellos que contienen los ítems más adecuados para un determinado objetivo. Su planteamiento teórico se basa en la TRI, y en la capacidad de esta teoría para obtener la Función de Información del ítem y el test en cada habilidad. Los tests óptimos parten de un banco de ítems calibrado que permite conocer la efectividad de cada ítem en la precisión de la medida para cada nivel de habilidad, pudiéndose de este modo seleccionar los más efectivos, logrando tests con niveles altos de precisión en la estimación de la medida, y al mismo tiempo con un número reducido de ítems.

Partiendo de los Bancos de Items es posible, además, seleccionar ítems que cumplan una serie de condiciones (Hambleton y cols, 1996) relativas al formato, contenidos, propiedades métricas, etc. en función de las necesidades y objetivos de la evaluación. Existe también la posibilidad de diseñar diferentes formas de un test que tengan funciones de información equivalentes.

## 4.3. TADIs

Los Tests Autoadaptados Informatizados (TADIs) son equivalentes a los TAIs con una diferencia importante entre ellos, a los sujetos se les permite elegir el nivel de dificultad de los ítems. En estos tests propuestos por Rocklin y O'Donnell (1987) el banco de ítems se agrupa en distintas categorías en función de su nivel de dificultad, de este modo en la aplicación del test, ítem a ítem, el sujeto puede elegir el nivel de dificultad de cada ítem, hasta responder un número concreto de ítems o alcanzar una determinada precisión en la estimación de la medida.

Como indica Wise (1999) se debería emplear el mayor número de categorías posible, con el máximo número de ítems que hicieran posible que un sujeto pudiera elegir todos los ítems de una misma categoría o nivel de dificultad.

En los TADIs al igual que ocurre con los TAIs, la propiedad de invarianza de la TRI permite la comparación de los resultados obtenidos por distintos sujetos, aunque los ítems aplicados a los mismos tengan diferente nivel de dificultad.

Diversas investigaciones (Rocklin et al., 1995; Vispoel et al., 1996; Roos et al., 1997; Wise et al., 1998) apuntan algunas ventajas y limitaciones del uso de TADIs con respecto a los TAIs.

| Tests Autoadaptados Informatizados  |   |
|---|---|
| Ventajas  | Limitaciones  |
| El uso de un Test Autoadaptado Informatizado (TADI) reduce la influencia de la ansiedad sobre el rendimiento, siendo más adecuado su empleo en sujetos con alta ansiedad y bajo autoconcepto que en sujetos que no desean un control de su rendimiento. | * Son menos precisos que los TAIs en la estimación del rendimiento, lo cual depende del ajuste entre el nivel de dificultad de las ítems seleccionados por el sujeto y su nivel de conocimiento.<br>* Su tiempo de aplicación es mayor. |

## 4.4. Tests MultiEtápicas

Tal como indica su nombre los tests multietápicas realizan el proceso de medición en varias etapas. La primera de ellas tiene por objeto realizar una estimación inicial del nivel del sujeto, aplicándose para ello un conjunto de ítems. A partir del nivel inicial estimado los sujetos realizarán en la segunda etapa el test que más se ajuste a él (fácil, medio o difícil).

Como podemos ver en la tabla siguiente, los tests multietápicos se encuentran situados entre los TIs y los TAIs en cuanto a la precisión en la estimación de la puntuación. Asimismo, con ellos es posible evitar algunas de las limitaciones que presentan los TAIs y que han sido analizadas en el apartado anterior.

| TESTS MULTIETÁPICOS  |   |
|--|---|
| Ventajas   | Limitaciones  |
| <ul style="list-style-type: none"> <li>* Permite que los sujetos revisen y cambien las respuestas a los ítems en cada etapa.</li> <li>* Es posible comparar las puntuaciones de los sujetos con una norma de grupo o con un estándar de ejecución, puesto que los ítems de las dos etapas están calibrados mediante los modelos TRI.</li> <li>* Los constructores de los tests pueden revisar la validez de contenido de los tests utilizados.</li> <li>* Es posible utilizar formas paralelas de los tests lo que incrementaría la seguridad de los mismos</li> <li>* No precisan disponer de bancos de ítems tan amplios como los TAIs.</li> </ul> | <ul style="list-style-type: none"> <li>* No proporcionan una precisión tan alta como los TAIs en la estimación de la medida.</li> </ul> |

Si bien las investigaciones han estado más centradas en los TAIs que en los tests multietápicos, actualmente se está produciendo un interés por conocer las características de las etapas, subtests e ítems para conseguir optimizar su utilización.

## 5. Software Utilizado en la Evaluación y Diagnóstico Psicoeducativo

- 
- **Sosia.** (Gordon, L.V., ECPA y TEA Ediciones). Test informatizado que engloba todas las características subyacentes de la persona, relacionadas con una actuación de éxito en un puesto de trabajo.
- **TEPO.** Test de Entrenamiento para Opositores. La finalidad de este test on-line es el entrenamiento y la familiarización con este tipo de pruebas y ayudar a los opositores a evitar el miedo ante el examen.
- **BAMI.** Bateria de aptitudes mentales (N. Seisdedos y J. Carrasco). BAMI es una herramienta informatizada, para la aplicación y corrección de una completa batería de aptitudes intelectuales mediante ordenador: Aptitud numérica, Razonamiento verbal, Aptitud espacial, Memoria, Atención/percepción, Flexibilidad cognitiva y Comprensión mecánica.
- **TRASI.** Test informatizado para la evaluación del razonamiento secuencial y la instrucción. TRASI es el primer test adaptativo informatizado comercializado en Europa.
- **EXHCOPA.** Examen de habilidades y conocimientos básicos. Desarrollado por el Instituto de Investigación y Desarrollo Educativo (IIDE) de la Universidad Autónoma de Baja California (UABC), para ser utilizado en las pruebas de admisión a la universidad

## **5. Tests específicos para logopedia**

Mientras que cada vez encontramos más tests informatizados en los ámbitos de la psicología y la educación, en logopedia se plantean dificultades especificar para el desarrollo de estos tests. A la dificultad intrínseca al desarrollo de instrumentos de diagnóstico y evaluación adecuados, se añaden una serie de dificultades cuando trabajamos en los ámbitos de la comunicación, el lenguaje y el habla.

A los problemas con los que se encuentran los TAIs para su desarrollo derivados del alto coste económico y de recursos que conlleva la creación de grandes bancos de ítems, hay que añadir:

- La dificultad de elaborar instrumentos informáticos que permitan evaluar técnicas de recogida de información observacionales y de interacción con el sujeto.
- Estos desarrollos no resultan rentables económicamente, dado que tienen un coste importante y van dirigidos a un ámbito muy específico. Este coste hace que difícilmente sean utilizados en lugar de los de lápiz y papel.
- El tratamiento informatizado del sonido aún presenta importantes dificultades. Al problema básico derivado de la gran cantidad de espacio necesario para almacenar archivos de sonido en los dispositivos de memoria, se añade una dificultad más relevante, hoy por hoy la tecnología aún no ha conseguido resultados totalmente satisfactorios en los sistemas de reconocimiento de voz y tratamiento del habla.

A pesar de estas dificultades también es necesario resaltar que el uso del ordenador como instrumento de presentación y recogida de información para pruebas diagnósticas, presenta algunas ventajas:

- Resulta adecuado para almacenar y corregir los ítems, automatizando el proceso y permitiendo de forma rápida y eficaz la emisión de informes.
- En las pruebas que contienen ítems en los que se utilizan imágenes (estáticas o dinámicas) y sonidos es posible controlar de forma sencilla muchas variables de presentación de estos estímulos (modalidad, tiempo, etc.).
- Mediante el ordenador es posible recoger información sobre las respuestas y el proceso que realiza el sujeto para realizar la prueba.

A pesar de estas ventajas, actualmente contamos con muy pocos tests informatizados para su uso en logopedia. Una excepción interesante es la batería **BLOC** de Miguel Puyuelo, Jordi Renom y Antoni Solanas, mediante ella podemos evaluar cuatro componentes del lenguaje:

- morfología,
- sintaxis,
- semántica y
- pragmática.

Las dificultades son muchas, y solo se observará interés por pruebas evaluativos de calidad que aporten beneficios importantes sobre el sistema tradicional en soporte papel, más fácil, cómodo y económico. Únicamente el desarrollo de aplicaciones integradas y flexibles, que acompañadas de futuros avances tecnológicos, posibiliten el análisis y evaluación del lenguaje (discriminar errores fonológicos, evaluar la estructura morfosintáctica, etc...)

## 5.2. Software de Evaluación

Actualmente existen algunos programas que nos permiten completar fichas para la evaluación de los sujetos con trastornos de comunicación y/o lenguaje. Entre ellas, cabe destacar por su calidad la **Guía EVELPIR: Guía para la evaluación y elaboración de programas de intervención rehabilitadora** de ONDAEDUCA, con la que podemos realizar un completo registro y evaluación del lenguaje. Para ello, además de un módulo que nos permite realizar la anamnesis del paciente, dispone de siete módulos de evaluación:

- Prerrequisitos.
- Prosodia
- Fonética
- Léxico
- Semántica
- Sintáxis
- Pragmática

Los objetivos de EVELPIR son:

- Evaluar los diferentes aspectos del lenguaje en los niños con dificultades lingüísticas de diferente etiología.
- Elaborar programas de desarrollo y rehabilitación del lenguaje individualizados.

EVELPIR ofrece también un selector de ejercicios, mediante el cual podemos realizar ejercicios con el ordenador para la evaluación del lenguaje.

## BIBLIOGRAFÍA

- BARBERO, M<sup>a</sup> I. (1999). Gestión informatizada de bancos de ítems. En J. Olea, V. Ponsoda, y G. Prieto (ed.): Tests Informatizados: Fundamentos y Aplicaciones. Madrid: Pirámide.
- BURIN, D.; DELGADO, A.R. y PRIETO, G. (1998) "Indicadores de estrategias de solución en una tarea informatizada de visualización espacial", *Relieve*, v4, n2, 5. [http://www2.uca.es/RELIEVE/V4N2\\_5.htm](http://www2.uca.es/RELIEVE/V4N2_5.htm)
- CORTADA DE KOHAN, N. (2004) "Teoría de Respuesta al ítem", *Evaluar*, nº 4. [http://www.aristidesvara.com/metodologia/psicometria/teoria\\_respuesta/trisup\\_aristidesvara.pdf](http://www.aristidesvara.com/metodologia/psicometria/teoria_respuesta/trisup_aristidesvara.pdf)
- EOEP (Gobierno del Principado de Asturias). Recursos informáticos para la atención educativa a los alumnos con discapacidad auditiva. <http://web.educastur.princast.es/eoep/eeaoeved/recursos/recursosprofesores/recinforal.pdf>
- HONTANGAS, P. (1999). Software para la construcción y administración de tests informatizados. En J. Olea, V. Ponsoda, y G. Prieto (ed.), Tests Informatizados: Fundamentos y Aplicaciones. Madrid: Pirámide.
- HONTANGAS, P.; OLEA, J. y PONSODA, V. (1998) "Elección de la dificultad en los tests adaptativos informatizados: un estudio piloto", *Relieve*, v4, n2, 3. [http://www2.uca.es/RELIEVE/V4N2\\_3.HTM](http://www2.uca.es/RELIEVE/V4N2_3.HTM)
- HONTANGAS, P.; PONSODA, V.; OLEA, J. y WISE, S.L. (2000) "The choice of item difficulty in self-adapted testing". *European Journal of Psychological Assessment*, 16, 1, 3-12.
- OLEA, J. PONSODA, V. y PRIETO, G. (1999) Tests Informatizados: Fundamentos y Aplicaciones. Madrid: Pirámide.