

García Jiménez, E., Gil, J. & Rodríguez Gómez, G. (1998). La evaluación de tests adaptativos informatizados. *RELIEVE*, 4 (2), art. 6. [http://www.uv.es/RELIEVE/v4n2/RELIEVEv4n2\\_6.htm](http://www.uv.es/RELIEVE/v4n2/RELIEVEv4n2_6.htm)

**e-Journal of Educational  
Research, Assessment and  
Evaluation**



**Revista ELectrónica de  
Investigación y EValuación  
Educativa**

## LA EVALUACIÓN DE TESTS ADAPTATIVOS INFORMATIZADOS

by/por

[Article record](#)

[About authors](#)

[HTML format](#)

**García Jiménez, Eduardo** ([egarji@cica.es](mailto:egarji@cica.es))

**Gil Flores, Javier** ([jflores@cica.es](mailto:jflores@cica.es))

**Rodríguez Gómez, Gregorio** ([gregorio.rodriguez@uca.es](mailto:gregorio.rodriguez@uca.es))

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

### Abstract

We present the state of art on computerized adaptive testing. Its possibilities in the improvement of teaching and learning are discussed

### Keywords

Evaluation; computerized adaptive tests

### Resumen

Presentamos una revisión del conocimiento actual sobre los tests adaptativos informatizados. Se analizan sus posibilidades en la mejora de los procesos de enseñanza y aprendizaje.

### Descriptores

Evaluación; tests adaptativos informatizados

Cuando todavía no puede hablarse de una verdadera tradición en la construcción y uso de tests adaptativos informatizados (en adelante, TAIs), y en nuestro país ni siquiera de una popularización de su uso, puede parecer pretencioso abordar ya su evaluación.

Este artículo no pretende revisar los recursos existentes, sino más bien orientar a aquellos que o bien han participado en la elaboración de este tipo de pruebas o que las han utilizado en alguna ocasión, en el análisis de sus posibilidades y limitaciones. Concretamente, vamos a presentar algunas aportaciones relacionadas con el modo en que puede reconocerse la valía de los TAIs en el análisis de habilidades cognitivas, así como sus posibilidades de aplicación en la mejora de los procesos de enseñanza y aprendizaje. No obstante, como instrumentos de medida que son, resulta igualmente importante revisar aquí algunas normas que permitan analizar sus características psicométricas. Por ellas, precisamente, vamos a comenzar este artículo.

### 1. Normas de evaluación

En los últimos cuarenta años la APA, la AERA y el NCME han constituido una media docena de comités diferentes (1954, 1966, 1974, 1977, 1985 y 1986) para dictar normas técnicas que orientasen el desarrollo y uso de los tests. En el comité de 1986, se abordó específicamente la preparación de un conjunto de recomendaciones que deberían ser aplicadas a los tests adaptativos informatizados. Si bien, con anterioridad, otros autores (Green y col., 1983) también se habían ocupado de formular un conjunto de guías para la evaluación de los TAIs. Y, más recientemente, Green (1988) o Wainer y Kiely (1987) se han ocupado de analizar las dificultades asociadas al desarrollo de este tipo de tests. Una de estas aportaciones (Green y col, 1983), que aparece recogida en el Cuadro 1, aborda nueve dimensiones diferentes en el análisis de los TAIs. Consideraciones relativas al contenido, dimensionalidad, fiabilidad, validez, estimación de parámetros, anclaje, características de los bancos de items, selección de

items y puntuación de los tests, factores humanos.

La preocupación de comités e investigadores individuales por la construcción y uso de los TAIs deriva del hecho de que, a pesar de sus muchos puntos en común con los tests de lápiz y papel, constituyen un instrumento esencialmente diferente a los tests tradicionales. La idea de Lord a finales de los sesenta<sup>[1]</sup>, de realizar una administración adaptativa que rompiese con la ineficacia de los tests de longitud fija para evaluar a los sujetos situados en los extremos de una distribución, ha generado un conjunto de posibilidades pero también ha planteado no pocos interrogantes e inconvenientes. Más aún si consideramos los cambios introducidos gracias a la informática en la selección, presentación y puntuación de los items.

Dos son las cuestiones fundamentales que conciernen a la evaluación técnica de los TAIs: (1) la relación entre las puntuaciones (y las decisiones a ella asociadas) y la administración convencional o adaptativa informatizada de los tests; y, (2), la viabilidad de los modelos de la TRI para proporcionar un sistema de medición técnicamente adecuado (Hambleton, Zaal y Pieters, 1990). En relación con estas cuestiones, las investigaciones de Green (1988) y de Wainer y Kiely (1987), han mostrado algunos de los problemas asociados al desarrollo de TAIs. Estos problemas afectan tanto a la equiparación de puntuaciones como a la selección de los items.

En lo que concierne a la equiparación de puntuaciones, los resultados de las investigaciones realizadas al respecto parecen demostrar la no equivalencia de las propiedades de medida de un test de lápiz y papel y un test adaptativo.

Mientras dos tests, uno de lápiz y papel y otro adaptativo, pueden ser igualados o calibrados en la misma escala utilizando un modelo de la TRI, en general, los dos tests no proporcionan el mismo grado de precisión en todos los puntos del continuo de habilidad (es decir, en general, las funciones de información de los tests difieren entre sí). (Hambleton, Zaal y Pieters, 1990: 360).

En lo relativo a la selección de items, los hallazgos de investigación vienen a subrayar la idea de que el contexto en el que se administra un item no es ajeno al resultado que se obtiene en él. De lo que se deriva que la respuesta que un sujeto da un item está influenciada por los items que se le hayan administrado con anterioridad. Esto, evidentemente, puede hacer muy difícil la comparación, por ejemplo, entre un conjunto de alumnos que contestan a los items de un test.

Estas afirmaciones ponen de relieve la importancia de considerar la valoración técnica como un requisito en su construcción y posterior aplicación de los TAIs. Aunque no debemos olvidar que el ajuste a unas normas psicométricas es una condición necesaria para cualquier test, pero no una condición suficiente. La posibilidad de aplicar este tipo de tests en la evaluación de habilidades complejas, de introducirlos como instrumentos asociados a los procesos de enseñanza y aprendizaje o de utilizarlos como herramientas en los procesos diagnósticos no resulta, desde luego, un asunto baladí. Precisamente, la elaboración de criterios para determinar el valor de los TAIs en la evaluación de habilidades cognitivas y en la mejora del proceso de enseñanza y aprendizaje, es el tema que nos ocupará en las páginas que siguen.

Cuadro 1. Orientaciones para la evaluación de Tests Adaptativos Informatizados (TAIs), de Green y col. (1983), tomadas de Hambleton, Zaal y Pieters (1990)

<b>DIMENSIONES</b>	<b>A. ORIENTACIONES</b>
Consideraciones relativas al contenido	<ol style="list-style-type: none"> <li>1. Deberían ser las mismas en los TAIs que en los tests de lápiz y papel</li> <li>2. El contenido de los ítems que forman un banco debería estar relacionado con las especificaciones que se han hecho de un determinado contenido</li> <li>3. Los ítems deberían diseñarse teniendo en cuenta las posibilidades del tipo de ordenador que va a utilizarse</li> </ol>
Dimensionalidad	<ol style="list-style-type: none"> <li>4. Hay que comprobar el ajuste al modelo de la TRI</li> <li>5. Deben seleccionarse ítems que sean altamente discriminativos</li> <li>6. Es necesario hacer un análisis factorial de las correlaciones tetracóricas entre los ítems</li> <li>7. Hay que examinar el supuesto de independencia local</li> <li>8. Es mejor formar subtests cuando no resulta posible asegurar la unidimensionalidad de una prueba</li> <li>9. Un test debería estar balanceado para reflejar la heterogeneidad del dominio de contenido y de los formatos de ítems</li> </ol>
Fiabilidad	<ol style="list-style-type: none"> <li>10. El error estándar de medida de cada test debería interpretarse como una función de la puntuación del test, en la métrica de la puntuación</li> <li>11. El error estándar de medida de cada test también debería situarse en la métrica de la habilidad</li> </ol>
Validez	<ol style="list-style-type: none"> <li>12. Debería valorarse el grado de similitud entre las matrices de varianza-covarianza obtenidas para los TAIs y las pruebas de lápiz y papel</li> <li>13. También deberían compararse las estructuras de covarianza de los dos tipos de pruebas</li> <li>14. Las versiones TAI y lápiz y papel de un test deberían validarse en función de un criterio externo</li> <li>15. Es necesario valorar la amplitud del sesgo en las predicciones con subpoblaciones importantes</li> </ol>
Estimación de parámetros	<ol style="list-style-type: none"> <li>16. El tamaño de las muestras utilizadas en la calibración de los ítems debe ser adecuado, generalmente en torno a los mil casos</li> <li>17. La muestra de calibración debe ser seleccionada de modo que sea posible contar con un número suficiente de examinados en el rango de habilidad necesaria para estimar la asíntota inferior y el punto de inflexión de la CCI</li> <li>18. El procedimiento para estimar los parámetros del ítem debe ser “empíricamente consistente” (las muestras grandes deberían generar buenas estimaciones)</li> <li>19. El procedimiento para estimar los parámetros del ítem debería ser insesgado, o en su defecto tendría que especificarse la naturaleza del sesgo</li> <li>20. Las CCI deberían ajustarse a los datos observados</li> <li>21. Es necesario comparar la dificultad de los ítems administrados tanto en la versión TAI como en la versión lápiz y papel</li> </ol>
Anclaje	<ol style="list-style-type: none"> <li>22. Hay que describir con detalle el procedimiento utilizado para situar los parámetros de los ítems en una métrica común</li> <li>23. Si se utilizan procedimientos de anclaje basados en la equivalencia de los grupos, es preciso demostrar esa equivalencia</li> </ol>
Características del banco de ítems	<ol style="list-style-type: none"> <li>24. Es necesario presentar tanto la distribución de los parámetros de los ítems como la distribución de los estadísticos descriptivos para las estimaciones</li> <li>25. Debe ofrecerse información sobre el total de los ítems del banco</li> </ol>
Selección de ítems y puntuación del test	<ol style="list-style-type: none"> <li>26. El procedimiento para la estimación de la habilidad y la selección de ítems debe ser documentado de forma explícita y detallada</li> <li>27. El procedimiento debería incluir un método de variación de los ítems seleccionados para evitar usar exclusivamente unos pocos ítems.</li> <li>28. El algoritmo utilizado por el ordenador debería administrar los ítems diseñados y registrar las respuestas poder como operaciones independientes, sin interferir con el proceso adaptativo</li> <li>29. El ordenador debería ser capaz de elegir el primer ítem apoyándose en información previa</li> </ol>
Factores humanos	<ol style="list-style-type: none"> <li>30. El ambiente del aula de informática en el que se administre el test debería ser tranquilo, confortable y libre de distracciones</li> <li>31. La disposición de la pantalla debería evitar reflejos o deslumbramientos</li> <li>32. Es necesario valorar empíricamente la legibilidad de la presentación en pantalla de los ítems</li> <li>33. La presentación debería poder incluir representaciones gráficas en las que se aprecien los detalles</li> </ol>

## 2. Análisis e interpretación de los TAIs

En una de las escasas ocasiones en las que el autor de la teoría triádica de la inteligencia se detuvo a analizar los aspectos psicométricos de la psicología cognitiva, lo hizo para hacernos reflexionar sobre la importancia relativa de los avances en el campo de la medición, cuando éstos se sitúan ante las exigencias planteadas por los cambios introducidos en las teorías cognitivas a lo largo de las dos últimas décadas. Robert Sternberg concluía su capítulo en la obra editada por Hambleton y Zaal (1990) diciendo:

Los progresos de la última década en la teoría psicométrica también han sido impresionantes. Pero cambiando las técnicas de análisis estadísticos no cambiaremos los tests, para que el progreso futuro en la medida de la inteligencia sea lo más favorable posible son necesarios tanto una mayor sofisticación en los tests como en los modos de analizarlos (pp. 388-389).

La necesidad de conciliar los avances teóricos con los procedimientos de medición no sólo es sentida en el campo de la inteligencia sino que también cabe extenderla a las áreas implicadas

en el estudio de las opiniones y actitudes o el rendimiento educativo. En este último campo se han producido importantes modificaciones en la definición y el análisis de los contenidos educativos (v.g. la aplicación de los mapas conceptuales, el análisis estructural de capacidades, etc.) que reclaman un tratamiento diferenciado desde la construcción y al uso de los tests.

La elaboración e interpretación de los TAIs debería prestar atención a tales novedades, avanzando no sólo en nuevos desarrollos psicométricos sino también en las direcciones marcadas desde las propias áreas de conocimiento que son objeto de medición. En este sentido, la evaluación de los TAIs debería adoptar criterios que vayan más allá de los estrictamente psicométricos y que se sean útiles para analizar el modo en que tales pruebas se

adaptan a las exigencias del contenido que tratan de medir.

En el trabajo, ya mencionado, de Sternberg se recoge algunos aspectos que pueden marcar pautas de trabajo para la elaboración de tales criterios, no sólo en el campo de la medida de la inteligencia sino en otros en los que también se aplican los TAIs. Una de las ideas subrayadas por Sternberg es que “los tests de inteligencia deberían basarse en la teoría cognitiva”, *a lo que cabría añadir que cualquier TAI (inteligencia, rendimiento, personalidad, etc.) debería apoyarse en alguna teoría o formulación suficientemente fundamentada y contrastada que asegurase no sólo la validez de la medida sino también el valor de la misma*. Una segunda idea que propone este autor es que “debería ampliarse la concepción de la inteligencia en la que se apoyan los tests”, especialmente habilidades de procesamiento de la información o habilidades de competencia social. En relación con estas dos propuestas, cabría señalar que muchos de los TAIs construidos en la actualidad son adaptaciones hechas de tests de lápiz y papel basados en teorías de la inteligencia o de las aptitudes intelectuales que encuentra cada vez menos crédito entre investigadores y prácticos.

También señala Sternberg que “los tests de inteligencia deberían estar más estrechamente relacionados con el entrenamiento”, es decir, con los programas de entrenamiento que favorecen el desarrollo de las habilidades cognitivas. Una implicación que puede extraerse de aquí es que a la hora de evaluar un TAI es necesario que tengamos en cuenta *si 2 tan sólo nos aporta un índice del dominio de habilidad o si más bien nos indica el punto de partida desde el que se puede iniciar la modificación de las habilidades de un sujeto*.

Una nueva propuesta de Sternberg hace hincapié en que “los tests deberían proporcionar información diagnóstica en relación con las habilidades del procesamiento de la información”. Haciendo una lectura más amplia que la que se desprende de esta afirmación, de por sí válida, podría decirse que un criterio para enjuiciar la valía de un TAI es conocer *si ofrece al investigador o al educador la posibilidad de*

*determinar las fortalezas y debilidades que presentan los sujetos en relación con las habilidades o contenidos objeto de medición.*

“Las puntuaciones de los tests deberían reflejar una diversidad de variables dependientes”, esta afirmación de Sternberg pretende destacar el hecho de que la mayoría de los tests, entre los que podría incluirse a los TAIs, basan sus puntuaciones relativas a la capacidad medida en los aciertos y errores de los sujetos o en la posibilidad de que éstos hayan respondido al azar. Otras variables dependientes, tales como el tiempo de reacción de los sujetos, las ideas previas, las estrategias de respuesta, etc. no son habitualmente incluidas. Un nuevo criterio de evaluación de los TAIs podría formularse entonces en términos del número de variables dependientes que son consideradas en el test.

La calidad de la información que pueden ofrecer los tests, es también comentado por Sternberg en el sentido de que “las puntuaciones de los tests deberían reflejar tanto los aspectos cuantitativos como cualitativos del procesamiento de la información”. Junto a la información cuantitativa que pueden ofrecer los TAIs, una prueba de este tipo resultará tanto más útil al usuario en *la medida en que ofrezca descripciones sobre el comportamiento del sujeto en el test* (v.g. tipos de contenidos en los que es/no es competente, clases de estrategias utilizadas en la solución de problemas, categorías de errores conceptuales cometidos, etc.). Conseguir que tales informaciones aparezcan como parte de los resultados de la administración de un TAI supone sólo un pequeño esfuerzo complementario en su diseño y, sin embargo, constituye un gran paso hacia la definitiva incorporación de este tipo de instrumentos en el trabajo cotidiano de terapeutas, orientadores o otros profesionales de la educación.

En definitiva, lo que se desprende de las consideraciones anteriores es que la valoración un TAI no puede hacerse al margen de la teoría de la inteligencia, la personalidad o el rendimiento que sustenta su contenido y, desde luego, sin considerar la valía que dicho instrumento tiene para sus usuarios. El síndrome del CI, considerar que los tests sólo proporcionan un valor

numérico de carácter general escasamente vinculado con la práctica profesional, puede convertirse en el síndrome q si no se presta más atención al modo en que se presentan los resultados de los TAIs.

### **3. Los TAIs en el proceso de enseñanza y aprendizaje**

¡ De qué nos sirve que sean adaptativos, si no podemos aprovechar su posibilidad de adaptación! Situados en el marco del proceso de enseñanza y aprendizaje, esta afirmación es clave para entender no sólo la naturaleza de los TAIs sino para atisbar sus posibilidades de aplicación. En ese marco de trabajo, la verdadera aportación que pueden suponer los TAIs no está en su posibilidad de adecuarse a la situación particular de cada examinado, sino en su virtualidad para conocer (gracias a esa adecuación) qué progresión realiza un aprendiz en su descripción, comprensión, análisis, valoración ... de determinada realidad física o intelectual.

Sin embargo, los TAIs que hemos construido (García, Gil y Rodríguez, 1996; Gil y otros, 1998; García y otros, 1998), así como aquellos otros que se han difundido y comercializado, no permiten realizar una aproximación al momento del proceso de aprendizaje en que se encuentra el alumno. Y este hecho, nosotros entendemos que es extensible tanto a situaciones tutoriales como a aprendizajes reales.

La posibilidad de que un TAI permita obtener información sobre los progresos en el aprendizaje es, entonces, un indicador clave de su verdadera naturaleza adaptativa que va más allá de su, incluso, demostrada valía psicométrica.

#### **3.1. Un punto de partida: Experiencias de evaluación**

Para ilustrar algunas de las posibilidades, pero también de sus limitaciones e inconvenientes, de la evaluación de TAIs vamos a considerar una experiencia en la que se llevó a cabo la valoración de un test construido para medir la capacidad de comprensión lectora entre alumnos de ESO (García, Gil y Rodríguez, 1996).

La prueba, denominada COMTEX (COMprensión de TEXtos), se genera en sus diversas adaptaciones a partir de un banco calibrado de

437 items. Su administración adaptativa e informatizada se construyó a partir del programa MICROCAT y en ella los items se seleccionan según el procedimiento de máxima información. Para estimar la capacidad de cada examinado, éste debe contestar a un mínimo de 10 items y a un máximo de 35, deteniéndose la administración en el momento en que el error de estimación para la capacidad del sujeto se sitúa por debajo del valor 0,30.

### 3.1.1. Finalidad de la evaluación

Con la evaluación de la prueba COMTEX(TAC) pretendíamos responder al menos a dos finalidades. De una parte, tratábamos de apreciar el valor del instrumento construido, recogiendo para ello las opiniones de quienes han tenido experiencia práctica con el mismo, y contrastando empíricamente la validez de los resultados que arrojará. De otra parte, la evaluación perseguía detectar las posibles deficiencias o puntos débiles de la prueba, que exigirían una revisión de la misma de cara a su mejora.

Es decir, se trataba de comprobar si el resultado de todo el proceso seguido para la construcción de un test adaptativo administrado por ordenador había satisfecho los objetivos de partida, respondiendo a las expectativas albergadas en relación con el mismo, y en caso de no ser así, de proponer las posibles medidas correctoras que optimicen el instrumento.

### 3.1.2. Diseño de evaluación

De acuerdo con la finalidad propuesta, la evaluación desarrollada respondía a un diseño basado en la descripción y valoración de diferentes aspectos del TAI, tomando como referencia la recogida de información de diferentes fuentes. En el ámbito operativo, se trataba de definir una serie de estándares o valores esperados en el objeto de evaluación y recoger información para decidir en qué grado se satisfacían dichos estándares. Ello implicaba la comparación de lo observado y lo suscitado en los participantes durante la administración del TAI con aquello que se pretendía conseguir.

El objeto de evaluación lo constituía la prueba construida y su utilización para la evaluación de las capacidades de los alumnos de Educa-

ción Secundaria Obligatoria en cuanto a la comprensión de información escrita en el área de Geografía, Historia y Ciencias Sociales. Concretamente, los aspectos que interesa evaluar en el TAI se referían a:

- Formato de la prueba
- Contenido de la prueba
- Aceptación por parte de los alumnos
- Validez de las puntuaciones otorgadas a los examinados
- Utilidad para la docencia en el área de Geografía, Historia y Ciencias Sociales

Los términos de comparación representaban, pues, los referentes respecto a los cuales equipararíamos el objeto evaluado. En este caso, el TAI y su utilización en centros de Educación Secundaria Obligatoria se compararon con lo que, de acuerdo con la literatura sobre tests adaptativos computerizados, se espera de este tipo de pruebas. Es decir, tomamos como modelo de referencia la adaptación de las preguntas formuladas a la capacidad del individuo examinado; una prueba "Aideal" en la que es posible la reducción en la duración y longitud de la prueba respecto a la administración tradicional no adaptativa, motivar a los alumnos y ofrecer una mayor facilidad para responder.

De otra parte, la presentación de la prueba sería valorada teniendo como referencia la definición de un formato de pantalla adecuado o la exposición clara de las instrucciones de respuesta y de los resultados alcanzados. Por último, nos interesaba valorar tanto el contenido del TAI como las puntuaciones asignadas a los sujetos. Así, de una parte, compararíamos los conocimientos y habilidades requeridas por esta prueba con los que se contemplan como contenidos en el curriculum oficial de la ESO para el Área de Geografía, Historia y Ciencias Sociales y, de otra, las puntuaciones otorgadas a los sujetos con las obtenidas en la última evaluación correspondiente al Área de Geografía, Historia y Ciencias Sociales, realizada por el profesor.

### 3.1.3. La administración del test COMTEX

Como ya hemos anticipado, una vez elaborado el TAI, se llevó a cabo su aplicación a alum-

nos de Educación Secundaria Obligatoria, con el fin de evaluar la utilidad y adecuación de la prueba de cara estimar el nivel de capacidad de los sujetos en la habilidad medida. En esta administración participaron un total de 5 Institutos de Enseñanza Secundaria, que fueron seleccionados siguiendo estrategias de muestreo teórico o deliberado, de acuerdo con los siguientes criterios:

a) Los centros seleccionados deberían poseer aula de informática con la suficiente dotación de recursos para permitir la administración de la prueba. Un escaso número de ordenadores supondría la imposibilidad de que todos los alumnos de un grupo-clase pudieran contestar simultáneamente al ins-

trumento y haría costosa en tiempo la administración.

b) Era necesario contar con cierta dispersión geográfica de los centros, de modo que no fueran elegidos todos los centros en un mismo contexto sociocultural. Por este motivo, se incluyeron en el estudio centros de las provincias de Cádiz, Huelva y Sevilla.

c) Sería conveniente localizar centros en los que el profesorado del área de Geografía, Historia y Ciencias Sociales estuviera pre-dispuesto e interesado en realizar la experiencia.

Como resultado de la combinación de estos criterios, fueron seleccionados los centros que aparecen en el Cuadro 2 en los cuales se llevó a cabo la administración del TAI.

Cuadro 2. Centros en los que se administró el TAI

<i>Centro</i>	<i>Localidad</i>
I.E.S. APaterna del Campo@	Paterna del Campo (Huelva)
I.E.S. ADoñana@	Almonte (Huelva)
I.E.S. "Drago"	Cádiz
I.E.S. "El Viso del Alcor"	El Viso del Alcor (Sevilla)
I.E.S. "Pablo Picasso"	Sevilla

La fase de aplicación estuvo precedida de reuniones con el profesorado que impartía en los centros las materias del área de conocimientos de Geografía, Historia y Ciencias Sociales. Además de exponer la finalidad del trabajo y explicar brevemente los pasos seguidos hasta el momento, se les indicaba la necesidad de someter a prueba el test, aplicándolo a los alumnos. Con objeto de que los profesores se familiarizaran con el TAI, se les entregó una copia en disquete, indicándoles normas básicas para su puesta en marcha y sugiriéndoles que trataran de responder ellos mismos a la prueba y la aplicaran a algunos de sus alumnos, pues nos interesaría posteriormente recoger sus opiniones sobre la misma. Un total de 6 profesores participaron en esta experiencia de aplicación del TAI.

Estas reuniones previas con el profesorado de cada centro sirvieron además para planificar la

aplicación a los alumnos, seleccionando algunos grupos en función de la disponibilidad de horarios y recursos informáticos. En la aplicación del TAI, en aulas de informática de los centros, participaron tanto los profesores como personal del equipo de investigación. La prueba COMTEX se administró a un total de 438 alumnos, obteniéndose para ellos una estimación de su capacidad y una traducción de la misma a una escala de 0 a 10, asumiendo la distribución normal de las capacidades. El método seguido para ello fue el método de Hastings, que permite la traducción de puntuaciones estandarizadas a percentiles mediante una relativamente simple expresión de cálculo.

Junto a la estimación de la capacidad de los examinados, podía obtenerse también una descripción de los errores cometidos, gracias a la clasificación de los elementos del banco en 11 categorías: conocimiento extratextual, conoci-

miento lingüístico léxico, captación de detalles, identificación de ideas principales, etc.

De los 438 alumnos que respondieron al TAI, 183 (41.8%) cursaban 31 de Educación Secundaria Obligatoria y 123 (28.1%) seguían estudios en 41 curso del mismo nivel educativo. Además de 27 alumnos (6.2%) de 11 de BUP,

91 alumnos (20.8%) de 21 de BUP y 14 alumnos (3.2%) de 31 de BUP. Un 71% de estos alumnos declararon poseer experiencia previa con ordenadores. La distribución de los alumnos por centros queda recogida en la Tabla 1

Tabla 1: Distribución de los alumnos examinados por centros

Centro	Nº alumnos
I.E.S. El Viso del Alcor	183
I.E.S. Pablo Picasso	68
I.E.S. Paterna del Campo	81
I.E.S. Almonte	65
I.E.S. Drago	41

La administración se llevó a cabo en el tercer trimestre del curso académico, teniendo lugar durante el mes de mayo y las primeras semanas de junio de 1996.

#### 3.1.4. La recogida de datos: elaboración y administración de instrumentos

De cara a la recogida de datos, se utilizaron dos instrumentos o técnicas: el cuestionario y la entrevista; y se recogieron datos a través de la revisión de documentos académicos oficiales, concretamente las actas correspondientes a la segunda evaluación realizada en el área de Geografía, Historia y Ciencias Sociales durante el curso 1995/96.

Con el cuestionario, tratamos de recoger las opiniones de los alumnos inmediatamente después de que eran examinados por el TAI. El instrumento constaba de un total de 21 items, consistentes en declaraciones acerca de distintos aspectos de la prueba, ante las cuales los alumnos debían emitir una valoración en una escala de cinco puntos que iban desde *Nada* hasta *Mucho*, según su grado de acuerdo con dicha declaración. Junto a esta serie de items, únicamente se solicitaba de los sujetos el nombre del centro, el curso y si poseían experiencia previa en el trabajo con ordenadores. El cuestionario tenía por tanto un carácter anónimo, que pretendía favorecer la sinceridad en las

respuestas. Un ejemplar de este instrumento aparece recogido como Anexo 1.

Las declaraciones que han servido de base para la valoración del TAI por parte de los alumnos examinados, aludían a aspectos tales como el atractivo de la prueba, el formato de presentación en pantalla, la facilidad para

trabajar con la prueba, la claridad de las instrucciones, la facilidad de comprensión de los resultados o la longitud y duración de la prueba. Además, los alumnos se pronunciaban acerca del carácter adaptativo del TAI, es decir, acerca del hecho de que cada alumno hiciera una prueba diferente según su capacidad, y también acerca de los beneficios formativos de la prueba, en cuanto que la experiencia de trabajo con la misma pudiera proporcionar ciertos aprendizajes.

Las entrevistas tenían por finalidad la recogida de opiniones y valoraciones de los profesores que más directamente habían tenido contacto con la utilización del TAI. Estos profesores, además de participar en la administración de la prueba a los alumnos, habían contado con tiempo suficiente para examinar y practicar con el instrumento, de forma que en el momento de celebrar las entrevistas, una vez finalizada la administración del TAI, poseían un conocimiento suficiente del instrumento evaluado.



Las entrevistas se desarrollaron en los propios centros de trabajo, y fueron registradas en audio para su posterior análisis. El desarrollo de las mismas respondió a un modelo de entrevista semiestructurada, en el que contando con un guión previo de cuestiones, no se descartaba incluir nuevas preguntas en función de las respuestas que pudiéramos ir obteniendo de los entrevistados.

De acuerdo con los objetivos de la evaluación, y teniendo en cuenta los términos de

comparación que iban a ser usados, el guión de entrevista recogía aspectos relativos al impacto sobre el alumnado, el contenido de la prueba, su adecuación al curriculum oficial de la ESO, la utilidad del instrumento y de la información proporcionada por el mismo de cara a la práctica educativa, o las ventajas e inconvenientes percibidos, junto con las medidas paliativas que pudieran contribuir a superar las limitaciones encontradas en el uso del TAI. El guión de entrevista usado aparece recogido el Cuadro 3.

Cuadro 3. Guión de entrevista a los profesores

- De acuerdo con la reacción de los alumnos o los comentarios que han hecho ¿cuál cree que es la opinión del alumnado respecto al TAI?
  - ¿Qué tipo de aprendizaje permitiría evaluar el TAI?
  - ¿Se corresponde el contenido de la prueba con el curriculum de la ESO?
- ¿En qué medida coinciden los resultados obtenidos en el TAC con las calificaciones que normalmente obtienen los alumnos?
  - ¿Qué utilidad tiene para usted la información obtenida a través del TAI?
- ¿Qué posibilidades habría de utilizar este instrumento en el marco de su trabajo habitual? ¿De qué modo podría incluirse? ¿En qué momento?
- ¿Qué elementos o mejoras introduciría para que un instrumento como éste resulte aún más útil para el profesorado de Geografía e Historia de la ESO?
  - ¿Podría sintetizar las ventajas e inconvenientes encontrados en el TAI?

Finalmente, se recogió información a partir de la revisión de las actas correspondientes a la segunda evaluación realizada durante el curso. Los propios profesores proporcionaron estos documentos, a partir de los cuales se vaciaron las calificaciones obtenidas por los alumnos que respondieron al TAI.

### 3.1.5. Análisis de datos: estrategias utilizadas

Una vez recogida la información de las distintas fuentes consideradas (alumnos, profesores y documentos oficiales) se procedió al análisis de la misma. Dado que los datos presentan distinta naturaleza, fue necesario recurrir a diferentes estrategias de análisis.

Los datos numéricos que se obtuvieron a partir de los cuestionarios respondidos por los alumnos, una vez codificadas las modalidades de respuesta de 1 a 5, generaron una matriz numérica cuyo tratamiento estadístico suponía una descripción de las respuestas, considerando

cada ítem como una variable. La distribución de frecuencias (expresadas en porcentajes) y la representación gráfica mediante diagramas de barras fueron la base para obtener una descripción de los alumnos sobre el TAI.

En el caso de las entrevistas, su transcripción generó un corpus de datos cualitativos textuales, en cuyo análisis hemos seguido los procedimientos habitualmente usados para el caso de este tipo de datos. En primer lugar, se llevó a cabo tareas de reducción de datos. Se realizó una segmentación del corpus en unidades de contenido, asignando las mismas a determinadas categorías de significado. Esta operación conceptual consistió en examinar cada fragmento, cada declaración, tratando de identificar el tema sobre el que se trata, planteándonos cuestiones tales como: ¿sobre qué habla este fragmento?, ¿qué aspecto valora en relación al TAI?.

El sistema de categorías, principal herramienta en el análisis de la información, fue generado a partir del guión de entrevista. Tomando como punto de partida los grandes temas allí apuntados, se procedió a completar el sistema de modo inductivo, a medida que se iban analizando los textos. Cada fragmento fue analizado para determinar su inclusión en alguna categoría, o la conveniencia de incluir alguna categoría nueva. Se trataba de un proceso de asimilación de los nuevos fragmentos y acomodación del sistema de categorías para albergarlos. Durante el proceso de categorización determinadas categorías tuvieron que ser subdivididas, dada la

heterogeneidad de temas que incluían; otras en cambio llegaron a ser fusionadas por recoger temas afines, modificadas, renombradas o simplemente suprimidas. El resultado fue un sistema emergente de categorías que representaba en sí mismo un primer resultado del análisis, por cuanto constituía un verdadero mapa conceptual de las opiniones de los profesores acerca del TAI.

Las posibles relaciones que pueden establecerse entre las categorías nos permitieron agrupar éstas en dimensiones o metacategorías. De esta manera, el sistema de categorías final es el que recogemos en el Cuadro 4.

Cuadro 4. Sistema de categorías

CUESTIONES	CATEGORÍAS
1	Reacciones positivas Reacciones negativas
2	Procedimientos Comprensión lectora Reacciones negativas
3	Opiniones
4	Hay correspondencia Según los casos
5	Carácter computerizado Carácter adaptativo Otras Medios necesarios Contenido Otros
6	Exploración inicial Individualización de la enseñanza Evaluación sumativa Evaluación formativa
7	Posibilidades limitadas Diagnóstico inicial Instrumento de evaluación Material didáctico Elemento motivador
8	Administración Capacidad medida Adecuación del contenido Formato de los items Resultados

### 3.2. ¿En qué medida pueden los TAIs ayudar a mejorar la enseñanza y el aprendizaje?

La pregunta que da nombre a este apartado es fundamental para determinar el verdadero valor de un TAI. Puede ser básico conocer si un instrumento de este tipo es capaz de ayudar a

orientadores y educadores a conocer las estrategias de aprendizaje de los alumnos, a determinar las habilidades de pensamiento que utiliza el aprendiz, a reconocer la transferencia de lo aprendido por un estudiante o si, gracias a los TAIs, podemos aproximarnos al rendimiento de un alumno en una materia y, por ende, ayudar a

sus profesores a adaptar el curriculum o las estrategias de enseñanza, verdaderamente estaremos ante un recurso útil y valioso.

### 3.2.1. TAI y estrategias de aprendizaje

Los tests con referencia a un criterio se crearon con la idea de promover un tipo de enseñanza adaptativa e individualizada (Glaser, 1963). Desde una perspectiva cognitiva, estos tests deberían proporcionarnos un conocimiento de los modelos mentales, los errores conceptuales o las habilidades de los estudiantes para solucionar problemas. Una información de esta naturaleza sería de extraordinaria utilidad a la hora de planificar el tipo de tareas que debería desarrollar un estudiante para progresar en su aprendizaje. No obstante, en buena medida, la gran promesa que representaron los tests criterios para crear nuevas formas de enseñanza individualizada capaces de asegurar la competencia de todos los estudiantes no se han hecho realidad (Frederiksen y White, 1990).

La introducción de los TAI ha supuesto una novedad tecnológica de primera magnitud en el campo de la medición, que ha superado algunas de las limitaciones presentes en los tests tradicionales de lápiz y papel, de elección múltiple utilizados en su momento para desarrollar una evaluación basada en criterios. No obstante, permanece el reto de lograr construir pruebas capaces de determinar los métodos y estrategias del estudiante en la solución de problemas, sus modelos mentales y sus falsas ideas o concepciones dentro de un dominio de conocimiento.

Asumiendo este reto, los TAI deberían ser valorados en función de su *posibilidad para describir las diferentes estrategias de aprendizaje utilizadas por el estudiante*: exploratorias, basadas en explicaciones, inductivas, deductivas, guiadas, etc. De igual modo, la utilidad de estos instrumentos de medida debería ser valorada teniendo en cuenta que los resultados que proporcionan fueran capaces de facilitar la intervención sobre las estrategias del estudiante en relación con la capacidad o capacidades medidas. Es decir, *un TAI sería instruccionalmente válido en la medida en que hiciese posible diseñar microestrategias (de repetición o de elaboración) y macroestrategias de aprendizaje*

(*de organización, de regulación y afectivo-emocionales*).

### 3.2.1. TAI y ambiente de aprendizaje

La introducción de ordenadores dentro de las aulas proporciona sin duda una oportunidad para reconsiderar en toda su amplitud la relación entre enseñanza y aprendizaje (Collins, 1990). En este nuevo marco educativo, es posible detectar no pocos problemas relativos al modo en que se utilizan los tests; de hecho, en los sistemas que incorporan materiales instruccionales para el alumno y procedimientos de evaluación (p.e. en la enseñanza de idiomas), es posible observar cómo se ha puesto un mayor énfasis en la valoración de las habilidades de orden inferior en detrimento de las de orden superior, al tiempo que se abandonan a su suerte a aquellos estudiantes que no logran pasar las puntuaciones de corte.

En lugar de crear ambientes educativos más próximos a la capacidad del estudiante, los nuevos sistemas tutoriales de enseñanza están aún lejos de lograr acercarse a las verdaderas habilidades del alumno. La incorporación de TAI puede allanar la pendiente que permita acercarse a la capacidad del estudiante en un dominio determinado y, con ello, favorecer el progreso de su aprendizaje. Si esa es una de las metas fundamentales que guían a los constructores de TAI sería pertinente valorar si el producto conseguido es uno en el que puede constatar que:

- a. El estudiante comienza con habilidades de un nivel dado y, dependiendo de su ejecución, los items siguientes incorporan habilidades de orden inferior o superior.
- b. Los items del test están correctamente situados dentro de la secuencia que, en la capacidad medida, va desde las habilidades de nivel inferior hasta las habilidades de nivel superior.
- c. El estudiante ha alcanzado un determinado nivel de progreso dentro de la secuencia de habilidades que comprende la capacidad medida.
- d. La información aportada tras la administración del test (tiempos de respuesta,

curvas características, funciones de información, porcentajes de error, indicadores sobre las habilidades adquiridas, etc.), resultan útiles para situar al estudiante dentro de una secuencia que describe la capacidad medida.

No obstante, modificar el ambiente de aprendizaje supone algo más que obtener la capacidad del estudiante en un dominio dado, si bien ese logro no es en sí nada sencillo. Crear un ambiente de aprendizaje adecuado supone asociar a la administración de los TAIs elementos fundamentales para el profesor, el alumno y la mediación que los une. Podremos conocer si un TAI es algo más que la mera constatación de un dominio si:

- a. El profesor obtiene información sobre el nivel de dominio que cada uno de sus alumnos/clientes posee sobre determinadas habilidades y, gracias a ella, puede rediseñar su enseñanza/terapia.
- b. El alumno puede conocer qué nuevas habilidades debe aprender.
- c. El tiempo invertido en la evaluación no supone interrumpir la secuencia instruccional ni el tipo de organización que habitualmente se sigue en la clase. El TAI está integrado dentro del proceso de enseñanza.
- d. El TAI puede facilitar, con las herramientas instruccionales/técnicas oportunas, una mejor monitorización de la enseñanza y un tratamiento más personalizado del alumno.
- e. El TAI no dificulta la interacción entre los estudiantes o de éstos con el profesor. Por el contrario, gracias al desarrollado de sistemas tutoriales adecuados, el test aumenta la frecuencia y la calidad de dichas interacciones.

### 3.2.2. TAIs y transferencia del aprendizaje

Si la meta que nos hemos trazado es la desarrollar procedimientos de medición que vayan más allá de la información ofrecida por las pruebas al uso, deberíamos construir instrumentos de medida que fuesen capaces de informar

el proceso de enseñanza. En esa dirección, una de las informaciones claves para cualquier orientador o profesor es, sin duda, conocer qué es lo que lleva a explotar a estudiantes con buen y mal rendimiento académico determinados procesos o contenidos de forma tan diferente.

El desarrollo de TAIs debería permitir a los profesores integrar mejor los procesos de enseñanza y aprendizaje, haciendo posible que los docentes lleguen a conocer:

- a. En qué grado el estudiante posee las claves o indicios (elementos) que resultan básicos para adquirir y utilizar la capacidad medida.
- b. Qué claves o indicios debería enseñar al estudiante en relación con dicha capacidad.
- c. En qué medida las habilidades adquiridas configuran una capacidad de uso habitual en el estudiante y no un mero repertorio de habilidades muertas.

Para ello, resulta fundamental que los TAIs permitan determinar la eficacia de transferencia de lo aprendido por el estudiante. Con un diseño adecuado de tales pruebas, el profesional de la educación tras la ejecución del alumno debería tener acceso a informaciones relativas a (Campione y Brown, 1990):

- a. Las nuevas habilidades que está en condiciones de adquirir ese alumno sin ayuda (eficiencia del aprendizaje).
- b. La posibilidad de que el estudiante pueda aplicar las habilidades adquiridas a diversas situaciones no necesariamente recogidas en el TAI (virtualidad de la transferencia).
- c. El nivel en que un estudiante comprende las reglas y principios asociados a la capacidad medida por el TAI (ejecución de la transferencia).
- d. El nivel en que la capacidad adquirida resulta un aprendizaje transversal (extensión de la transferencia).

### 3.2.3. Los TAIs y el diagnóstico de habilidades

Uno de los campos en los que los TAIs han creado más expectativas es el diagnóstico de las habilidades de los estudiantes. Estas expectativas encuentran una justificación en el hecho de que algunos de los objetivos asociados a los TAIs están próximos a aquellos que se establecen en el ámbito del diagnóstico. Un TAI se construye bajo el supuesto de que la presentación de un ítem cualquiera a un sujeto es el resultado de un proceso de análisis en el que se tienen en cuenta las características del ítem y las del propio sujeto. Pues bien, el problema básico del diagnóstico que se apoya en el uso de tests es determinar bajo qué condiciones, restricciones o reglas se elige un ítem de modo que nos permita profundizar en el proceso que sigue un individuo para afrontar una tarea. En ambos campos, el problema es un problema no sólo de contenido sino, sobre todo, de proceso: ¿cuál es el siguiente ítem? y ¿qué características debe poseer? Un TAI que pretenda utilizarse con fines diagnósticos debería responder satisfactoriamente a esas dos cuestiones.

*A. ¿Cómo debería ser un buen ítem diagnóstico?*

Si deseamos averiguar no sólo qué respuestas de un sujeto son correctas y cuáles erróneas sino que, además, queremos conocer cómo procede ese sujeto ante el problema planteado, entonces debemos construir ítems que sean capaces de ofrecernos la mayor información posible sobre la respuesta de un sujeto en el tiempo que invierte en responderlos. Los ítems, primeramente, deben ser suficientes y eficientes.

No obstante, estas dos cualidades estadísticas son tan sólo una parte de la cuestión. Algunos ítems excelentes no son, sin embargo, apropiados para evaluar a determinados sujetos. También debemos, pues, construir ítems que estén basados en el conocimiento que se tiene sobre un individuo en particular. Parece razonable que los buenos ítems de diagnóstico “puedan obtener un amplio rango de información y la habilidad del estudiante para dar una respuesta” (Marshall, 1990: 437).

Si estamos interesados en identificar qué habilidades de un individuo están bien asentadas y

cuáles están inoperantes o mal orientadas, más que clasificar a una persona como intelectualmente buena, mala o mediocre, de poco nos sirven las medidas tradicionales de dificultad del ítem o de habilidad intelectual.

La TRI sí que tiene una aplicabilidad potencial ante este tipo de exigencias. En su forma más simple, la TRI determina la probabilidad de que un individuo acierte un ítem en función de las características del ítem (normalmente, la dificultad) y de las características de la persona que responde (generalmente, algún tipo de medida de su habilidad). Llegados a este punto, cabe preguntarse ¿es posible incorporar las características cognitivas de los ítems dentro de las estimaciones de la TRI, de modo que puedan elaborarse mejores ítems diagnósticos?

Marshall (1990) recoge varias tentativas en esta dirección. Una de las propuestas es la que plantea Fischer (1973), que este autor denominó «modelo de test logístico lineal» y que se empleó inicialmente para analizar la complejidad psicológica de los problemas de cálculo. En el modelo, los parámetros del ítem están apoyados en la estructura cognitiva de los reactivos, definida de acuerdo con las operaciones que se requerían para aportar la solución. De igual forma, el modelo incorpora tanto las estimaciones de la dificultad (según Rasch) como los parámetros de la operación, derivados éstos últimos de la estructura cognitiva asociada a las posibles estrategias de solución.

El modelo de parameterización de las estructuras cognitivas ideado por Fischer implica clasificar los problemas de acuerdo con la presencia o ausencia de un conjunto definido de características. Después, se trabaja con una matriz de orden  $n \times p$ , donde  $n$  es el número de ítems y  $p$  es el número de características psicológicas. Cada ítem se define mediante un vector de “ceros” y “unos”, que indica la presencia o ausencia de los factores psicológicos.

Ahora bien, ¿qué características de los individuos podrían incorporarse dentro de las estimaciones de la TRI?

Lo habitual en los TAIs construidos desde la TRI es que las características cognitivas de los

individuos se estimen a partir de sus habilidades en relación con un dominio específico. Se obtiene, de ese modo, una medida del rendimiento que indica la proporción de ítems de un test que un individuo puede responder correctamente. No obstante, en principio no habría objeción para que también se incorporasen en las estimaciones de la TRI medidas sobre las habilidades cognitivas, mejorando así la generación de buenos ítems de diagnóstico. Algunas de estas mediciones ya han sido comentadas en este trabajo: el tiempo de reacción de la memoria a corto plazo, el tiempo utilizado en la recuperación de la información de la memoria a largo plazo y la estimación del procesamiento de imágenes o de la organización espacial podrían ser buenos ejemplos.

#### B. *¿Cómo deberían seleccionarse los ítems diagnósticos?*

El orden en que van a presentarse los ítems a un sujeto es crucial, según hemos comentado, tanto en un proceso diagnóstico como en la elaboración de cualquier TAI. Salvando los problemas estrictamente psicométricos, el modo en el que se organiza el conocimiento en un dominio específico y el modelo de aprendizaje del estudiante son los que determinan el proceso de selección de un ítem. (Marshall, 1990: 438). Examinaremos dos posibilidades básicas: el conocimiento/aprendizaje estructurando y el conocimiento/aprendizaje en forma de red enmarañada o quasi-jerárquico.

En una estructura jerárquica, el objeto a evaluar mediante el test constaría de un conjunto formado por destrezas cognitivas,  $C_1$  y  $C_2$ , y su correspondiente combinación,  $C_{12}$ . Si partimos de la existencia de una relación jerárquica entre estas habilidades, contestar correctamente a una combinación implica tener éxito en todos los componentes de esa combinación. Por tanto, un alumno que responda correctamente a  $C_{12}$  es un alumno que domina todas las destrezas incluidas en esa combinación.

En una habilidad que conste de pocas destrezas no es difícil determinar el diseño óptimo de selección de ítems apoyándose en los principios de la programación dinámica. El sistema es un

proceso de Markov. Así, para cualquier tipo de ítem relativo a una destreza dada ( $C_1$ ,  $C_2$  o  $C_{12}$ ) existen tres posibles resultados: se domina, no se domina o todavía no se ha presentado. El número de estados en el modelo de Markov se determina por  $3^k$ , donde 3 es el número de resultados y  $k$  el número de ítems diferentes a los que puede responderse. Se trataría, entonces, de estimar la probabilidad de que un sujeto haya aprendido cualquier destreza concreta o su combinación.

Cuando aumentamos el número de destrezas no resulta fácil, sin embargo, establecer cuál es la mejor forma de seleccionar los ítems (v.g. destrezas a medir) que debe responder un sujeto en un test. Un primer grupo de soluciones a este problema pasa por la consideración de diseños típicos de ramificación fija o tests secuenciales: nivel flexible (Lord, 1970), piramidal (Weiss, 1982), en estructura de árbol (Hulin, Drasgow y Parsons, 1983), "stradaptive" (Weiss, 1982), ramificado y "testlets" (Wainer y Kiely, 1987) o de búsqueda dicotomizada (Xiao, 1990).<sup>[2]</sup> Otro grupo de soluciones pasa por la consideración de diferentes heurísticos apoyados en un enfoque bayesiano, en los que la probabilidad de responder correctamente a un ítem depende de las respuestas previas del sujeto (Marshall, 1981).

Aunque el modelo de Markov y los tests secuenciales aportan algunas soluciones al problema de la selección de ítems cuando se construyen tests para evaluar habilidades jerárquicas, no podemos olvidar que sólo en determinadas aplicaciones tiene sentido considerar ese tipo de habilidades. Estamos hablando, por ejemplo, de campos como el de los procedimientos matemáticos (operaciones básicas, resolución de algunos tipos de problemas, etc.) o de la electricidad básica (p.e. circuitos eléctricos), en los que se requiere una composición jerárquica de las habilidades y destrezas.

La representación del conocimiento en dominios más complejos requiere formas de aprendizaje igualmente complejas y, por ende, medidas de logro sensiblemente distintas. Consideremos, por ejemplo, una estructura particular de conocimiento a la que la psicología cognitiva

ha prestado especial atención: el esquema. Como es sabido, un esquema alude a una representación genérica de una experiencia, situación o concepto cuya estructura puede representarse mediante una red. En esa red pueden identificarse términos descriptivos, condiciones previas y reglas relativas a una meta o propósito. Las formas de las redes pueden ser simples, de modo que cada elemento de la red sólo se relaciona con otros elementos dentro de la red (ver Figura 1 A), pero también pueden existir redes enmarañadas en las que algún elemento de la red esté compartido por una segunda red (Figura 2 B).

En general, cualquier parte específica de un esquema sólo activa una muestra de los elementos que caracterizan al esquema en su totalidad. Considerando que al menos existen tres grupos diferentes de elementos: los hechos declarativos, las condiciones previas y los subsiguientes procedimientos o reglas, cualquier item de un test dado debería considerarse como un subconjunto de estos tres elementos. El problema al que se enfrenta el diagnóstico es el de cómo medir varios subconjuntos y, por tanto, estimar de forma eficiente la integridad del conocimiento que tiene un sujeto de un esquema.

Según Marshall (1990) son tres las cuestiones generales que cabría formularse en relación con la comprensión y el uso de esquemas por parte

de un sujeto: (1) ¿Cuál es el nivel de competencia que demuestra un individuo al asociar un item con un esquema?; (2) ¿Qué elementos críticos de un esquema forman parte del conocimiento almacenado por un individuo?; y, (3) ¿Están los elementos críticos de un esquema ligados entre sí en la forma correcta?

Valorar la competencia que una persona tiene para resolver un problema que supone la aplicación de un esquema implica al menos dos cosas: primero, confirmar que el individuo está intentando acceder al esquema y, segundo, valorar su competencia para asociar el item al esquema. La cuestión, entonces, es encontrar diseños de presentación de items que permitan seguir el proceso de construcción y aplicación de esquemas por un sujeto. Los diseños de ramificación variable, como la estrategia autodaptable (Rocklin, 1989) o el método de máxima información (Weiss, 1982) podrían aportar alguna solución al problema, sin embargo están aún por explorar sus verdaderas posibilidades en el análisis de los elementos críticos en la formación de esquemas o en el estudio del almacenamiento de esquemas en la memoria a largo plazo.

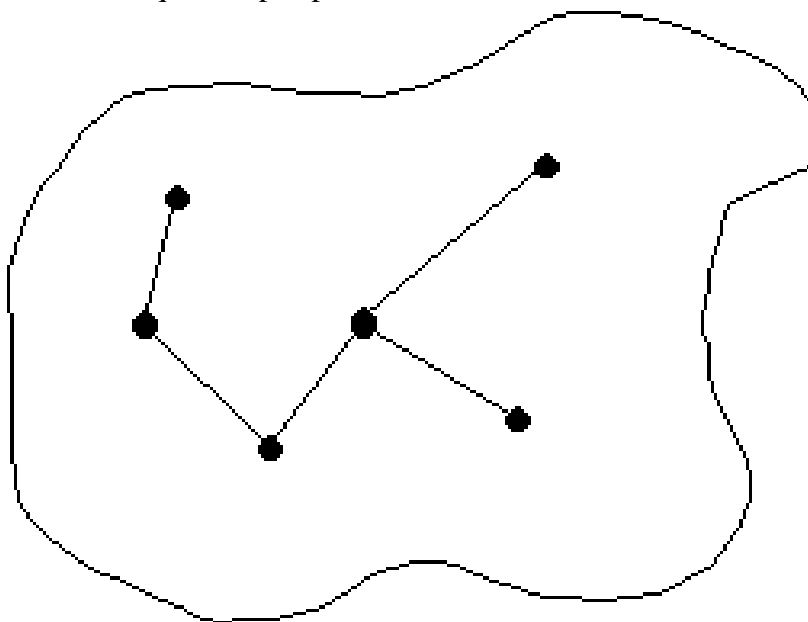


Figura 1 A. Red Simple: Todos los nódulos forman parte de un esquema simple

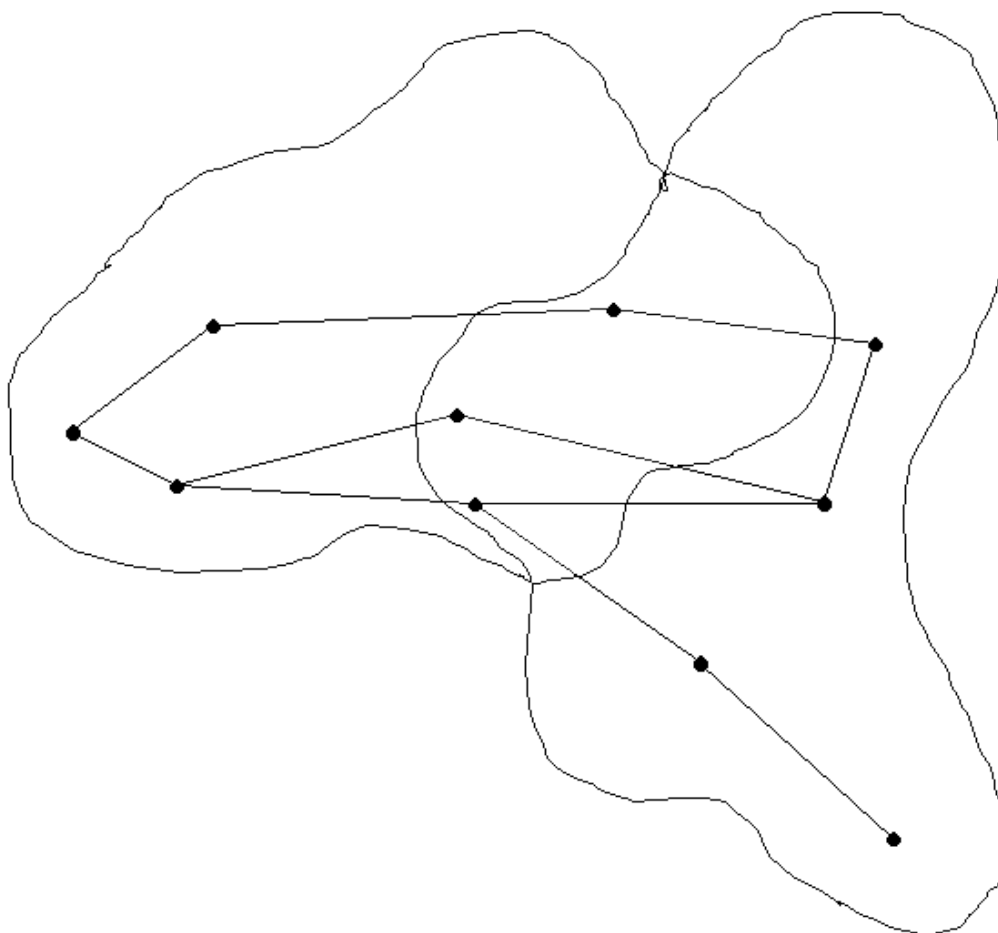


Figura 1B. Red enmarañada: Algunos nódulos son comunes a ambos esquemas

En resumen, si un TAI aspiran a convertirse en instrumento verdaderamente adaptativo a las condiciones del proceso de enseñanza y aprendizaje debería superar determinados criterios o indicadores. El Cuadro 5 representa una síntesis de las dimensiones consideradas

hasta aquí: Estrategias de aprendizaje de los alumnos, ambientes de aprendizaje del aprendiz, transferencia de lo aprendido, diagnóstico de habilidades.



Cuadro 5. Orientaciones para la evaluación de TAIs, según su vinculación al proceso de enseñanza-aprendizaje

DIMENSIONES	Orientaciones
Estrategias de aprendizaje	<ol style="list-style-type: none"> <li>1. Deberían describir las diferentes estrategias de aprendizaje utilizadas por el estudiante: exploratorias, basadas en explicaciones, inductivas, deductivas, guiadas.</li> <li>2. Es necesario que los TAIs hagan posible diseñar microestrategias (de repetición o de elaboración) y macroestrategias de aprendizaje (de organización, de regulación y afectivo-emocionales).</li> <li>3. Los resultados de los TAIs deberían ofrecer información sobre las ideas o concepciones erróneas que sostienen los alumnos en un determinado dominio ó ámbito de conocimiento.</li> </ol>
Ambientes de aprendizaje	<ol style="list-style-type: none"> <li>4. Los TAIs deberían favorecer la creación de ambientes educativos próximos a la capacidad del estudiante.</li> <li>5. El profesor está en disposición de obtener información sobre el nivel de dominio que cada uno de sus alumnos/clientes posee sobre determinadas habilidades y, gracias a ella, puede rediseñar su enseñanza/terapia.</li> <li>6. El alumno puede conocer qué nuevas habilidades debe aprender.</li> <li>7. El tiempo invertido en la evaluación no supone interrumpir la secuencia instruccional ni el tipo de organización que habitualmente se sigue en la clase. El TAI está integrado dentro del proceso de enseñanza.</li> <li>8. El TAI puede facilitar, con las herramientas instruccionales/técnicas oportunas, una mejor monitorización de la enseñanza y un tratamiento más personalizado del alumno.</li> <li>9. El TAI no dificulta la interacción entre los estudiantes o de éstos con el profesor. Por el contrario, gracias al desarrollado de sistemas tutoriales adecuados, el test aumenta la frecuencia y la calidad de dichas interacciones.</li> </ol>
Transferencia del aprendizaje	<ol style="list-style-type: none"> <li>10. Los TAIs deberían permitir a los profesores conocer en qué grado el estudiante posee las claves o indicios (elementos) que resultan básicos para adquirir y utilizar la capacidad medida.</li> <li>11. Los profesores deberían tener acceso a las claves o indicios en que los apoyarse para enseñar al estudiante en relación con la capacidad medida.</li> <li>12. Los profesores deberían conocer en qué medida las habilidades adquiridas configuran una capacidad de uso habitual en el estudiante y no un mero repertorio de habilidades muertas.</li> <li>13. El profesional de la educación tras la ejecución del alumno debería tener acceso a informaciones relativas las nuevas habilidades que está en condiciones de adquirir ese alumno sin ayuda (eficiencia del aprendizaje).</li> <li>14. Un TAI debería ofrecer información sobre la posibilidad de que el estudiante pueda aplicar las habilidades adquiridas a diversas situaciones no necesariamente recogidas en el TAI (virtualidad de la transferencia).</li> <li>15. Tras las ejecuciones de un estudiante, el profesor debería tener información sobre el nivel en que un estudiante comprende las reglas y principios asociados a la capacidad medida por el TAI (ejecución de la transferencia).</li> <li>16. El profesor debería tener información sobre el nivel en que la capacidad adquirida resulta un aprendizaje transversal (extensión de la transferencia).</li> </ol>
Diagnóstico de habilidades	<ol style="list-style-type: none"> <li>17. Los parámetros de los items deberían apoyarse en la estructura cognitiva exigida a los individuos en cada uno de los reactivos.</li> <li>18. Las estimaciones de la TRI deberían incorporar medidas sobre las habilidades cognitivas de los sujetos (v.g. tiempo de reacción, tiempo de recuperación de la información, etc.).</li> <li>19. Las estrategias de selección de los items deberían tener en cuenta el tipo de aprendizaje/conocimiento objeto de medida.</li> <li>20. Para seleccionar un item en un TAI debería considerarse el nivel de competencia que exige al sujeto, en función del tipo de aprendizaje o conocimiento requerido.</li> </ol>

## Referencias

Campione, J.C. y Brown, A.L. (1990). Guided learning and transfer: Implications for approaches to assessment, en N. Frederiksen, Glaser, R., Lesgold, A. y Shafto, M.G. (Eds.). Diagnostic monitoring of skill and

knowledge acquisition. Hillsdale, New Jersey, Lawrence Erlbaum Ass., 141-172.

Collins, A. (1990). Reformulating testing to measure learning and thinking. Comments on chapter 1, 2 and 3, en N. Frederiksen, Glaser, R., Lesgold, A. y Shafto, M.G.

- (Eds.). Diagnostic monitoring of skill and knowledge acquisition. Hillsdale, New Jersey, Lawrence Erlbaum Ass., 75-87.
- Fisher, G. (1973). The linear logistic test model as an instrument of educational research. *Acta Psychologica*, 37, 359-374.
- Frederiksen, J. R. y White, B.Y. (1990). Intelligent tutors as intelligent testers, en N. Frederiksen, Glaser, R., Lesgold, A. y Shafto, M.G. (Eds.). Diagnostic monitoring of skill and knowledge acquisition. Hillsdale, New Jersey, Lawrence Erlbaum Ass., 1-25.
- García, E. Gil, J. y Rodríguez, G. (1996). Los tests adaptativos computerizados (TACs). Su aplicación al área de Geografía, Historia y Ciencias Sociales en la Educación Secundaria Obligatoria. Informe Final de Investigación. Madrid: CIDE. Documento inédito.
- García, E. y otros (1998). LEX. Un test adaptativo informatizado de Legislación Educativa Básica. Sevilla: documento inédito.
- Gil, J. y otros (1998). ESTADIS. Un tests adaptativo informatizado de Estadística Descriptiva. Sevilla: documento inédito.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, 18, 510-522.
- Green, B.F. y cols. (1983). Applications of item response theory. Vancouver, Canada: Educational Research Institute of British Columbia.
- Green, B.F. (1988). Critical problems in computer-based psychological measurement. *Applied Measurement in Education*, 1, 223-231.
- Hambleton, R.K. Zaal, J.N. y Pieters J.P.M. (1990). Computerized adaptive testing: Theory, applications and standards, en R.K. Hambleton y J.N.Zaal (Ed.). *Advances in educational and psychological testing*. Boston: Kluwer Academic Press, 341 – 366.
- Hulin, C.L., Drasgow, F. Y Parsons, C.K. (1983). *Item Response Theory: Applications to Psychological Measurement*. Homewood, Illinois: Dow Jones-Irwin.
- Lord, F.M. (1970). Estimating item characteristic curves without knowledge of their mathematical form, *Psychometrika*, 39, 247-264.
- Marshall, S.P. (1981). Sequential item selection: Optimal and heuristics policies. *Journal of Mathematical Psychology*, 23, 134 – 152.
- Marshall, S.P. (1990). Generating good items for diagnostic tests, en N. Frederiksen, Glaser, R., Lesgold, A. y Shafto, M.G. (Eds.). *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale, New Jersey, Lawrence Erlbaum Ass., 433-452.
- Rocklin, T. (1989). Individual differences in item selection in Computerized adaptive Testing. Paper presented at the Annual Meeting of the A.E.R.A. San Francisco, 27 al 31 Marzo.
- Sternberg, R.J. (1990). Cognitive theory and psychometrics, en R.K. Hambleton y J.N.Zaal (Ed.). *Advances in educational and psychological testing*. Boston: Kluwer Academic Press, 367-393.
- Wainer, , H. y Kiely, G. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24, 3, 185-201.
- Weiss, D.J. (1982). Improving measurement quality and efficiency with adaptive testing. *Computerized Adaptive Testing*. New York: Academic Press.
- Xiao, B. (1990). Dichotomous search strategies for Computerized Adaptive Testing. Paper presented at the Annual Meeting of the A.E.R.A. Boston, 16 al 20 Abril.

## ANEXO 1

### CUESTIONARIO PARA LA EVALUACIÓN DE LA PRUEBA COMTEX

Centro:

Curso:

Experiencia previa en el trabaja con ordenadores:    SI        NO    (Tacha lo que proceda)

Tras terminar tu trabajo con el COMTEX, nos gustaría que emitieras una opinión sobre los distintos aspectos de esta prueba. Es algo tan sencillo como rodear con un círculo la opción que consideres correcta en cada una de las afirmaciones que aparecen a continuación:

	Mucho	Bastante	Algo	Poco	Nada
1. El modo en que ha comenzado la prueba me ha gustado.....	5	4	3	2	1
2. La forma de presentar los textos en pantalla facilita su lectura.....	5	4	3	2	1
3. Me gusta trabajar en una prueba como ésta cuando se presenta en un ordenador.....	5	4	3	2	1
4. Contestar a estas preguntas por ordenador es más sencillo que hacerlo con lápiz y papel .....	5	4	3	2	1
5. Trabajar con el TAC-1 requiere tener conocimientos de informática.....	5	4	3	2	1
6. Las instrucciones son claras y fáciles de seguir .....	5	4	3	2	1
7. En todo momento he sabido lo que tenía que hacer con el ordenador.....	5	4	3	2	1
8. Los resultados que se obtienen en la prueba son fáciles de comprender.....	5	4	3	2	1
9. Obtener los resultados de una prueba nada más terminarla me ayuda a aprender.....	5	4	3	2	1
10. Es ideal que las preguntas de una prueba se ajusten a lo que yo puedo hacer.....	5	4	3	2	1
11. Me parece bien que cada alumno haga una prueba diferente según la capacidad que demuestre.....	5	4	3	2	1
12. La prueba me ayuda a saber lo que soy capaz de hacer y lo que no.....	5	4	3	2	1
13. Desde el principio sabía lo que la prueba trataba de evaluar.....	5	4	3	2	1
14. Las preguntas te resultan fáciles.....	5	4	3	2	1
15. Esta prueba me ha ayudado a saber desde dónde debo empezar a aprender.....	5	4	3	2	1
16. Las preguntas de esta prueba me han ayudado a comprender lo que leo.....	5	4	3	2	1
17. He aprendido mientras contestaba la prueba.....	5	4	3	2	1
18. El COMTEX me ha ayudado a organizar mejor mi forma de trabajar con un texto.....	5	4	3	2	1
19. Me he divertido contestando a las preguntas.....	5	4	3	2	1
20. Mientras contestaba a las preguntas, se me ha pasado el tiempo rápido.....	5	4	3	2	1
21. Es suficientemente grande el número de preguntas que he contestado.....	5	4	3	2	1

## **NOTA**

[1] Hay quienes reconocen a Binet la idea de administrar a cada sujeto un test hecho a su medida.

[2] Para una revisión de estas estrategias para la selección de nuevos items ver Renom, J. (1993). *Tests adaptativos computerizados. Fundamentos y aplicaciones*. Barcelona: PPU

---

## **ABOUT THE AUTHORS / SOBRE LOS AUTORES**

**García Jiménez, Eduardo** ([egarji@cica.es](mailto:egarji@cica.es)). Universidad de Sevilla.

**Gil Flores, Javier** ([jflores@cica.es](mailto:jflores@cica.es)). Universidad de Sevilla.

**Rodríguez Gómez, Gregorio** ([gregorio.rodriguez@uca.es](mailto:gregorio.rodriguez@uca.es)). Universidad de Cádiz.

---

## **RELIEVE**

**Revista E**lectrónica de **I**nvestigación y **E**valuación **E**ducativa  
*E-Journal of Educational Research, Assessment and Evaluation*

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).