

Tests Adaptativos Informatizados

Francisco J. Abad, Julio Olea y Vicente Ponsoda

Dto. Psicología Social y Metodología

"En las últimas dos décadas los tests informatizados han pasado de ser un procedimiento experimental a ser empleado por cientos de programas de evaluación que evalúan a millones de personas cada año" ... "ser evaluado mediante un ordenador puede pronto llegar a ser incluso más natural que ser evaluado en papel" (p. 358).

Davey, T. (2005). *Computer-based testing*, En B. S. Everitt y D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science*. Hoboken, NJ: Wiley.

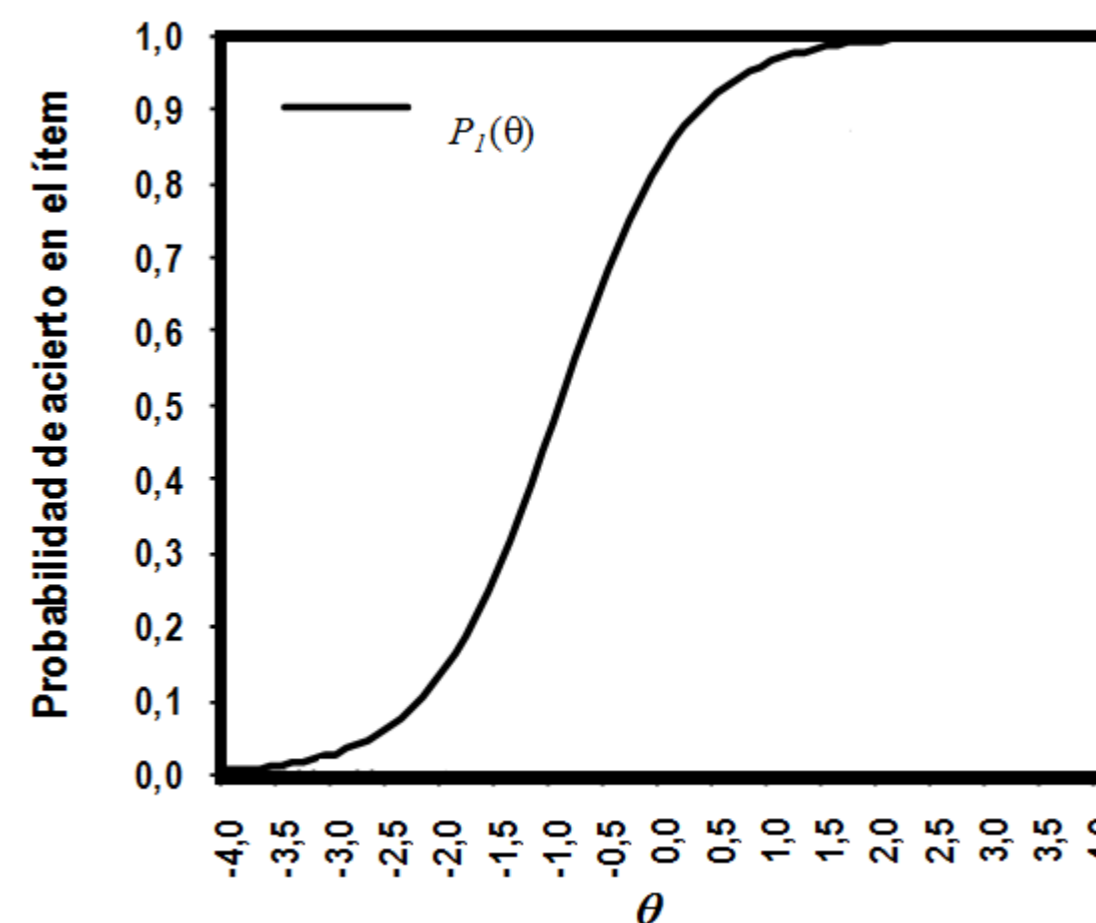
Test fijos vs Tests adaptativos informatizados (TAIs)

En los tests fijos se aplican los mismos ítems a todos los evaluados. Esto constituye una pérdida de eficacia porque los ítems que nos resultan demasiado fáciles o demasiado difíciles no informan bien de nuestro nivel de aptitud.

La principal ventaja de un TAI es que se genera un test "a medida" (ajustado en dificultad) para cada evaluado. Un TAI ofrece una gran ventaja en relación a un test fijo: permiten mantener la precisión reduciendo a aproximadamente la mitad el número de ítems administrados y obtener estimaciones más precisas, especialmente para los niveles más bajos o altos de aptitud.

La base matemática de un TAI

Los TAIs se basan en un modelo estadístico propuesto a finales de los años 60: La **teoría de respuesta al ítem (TRI)**.

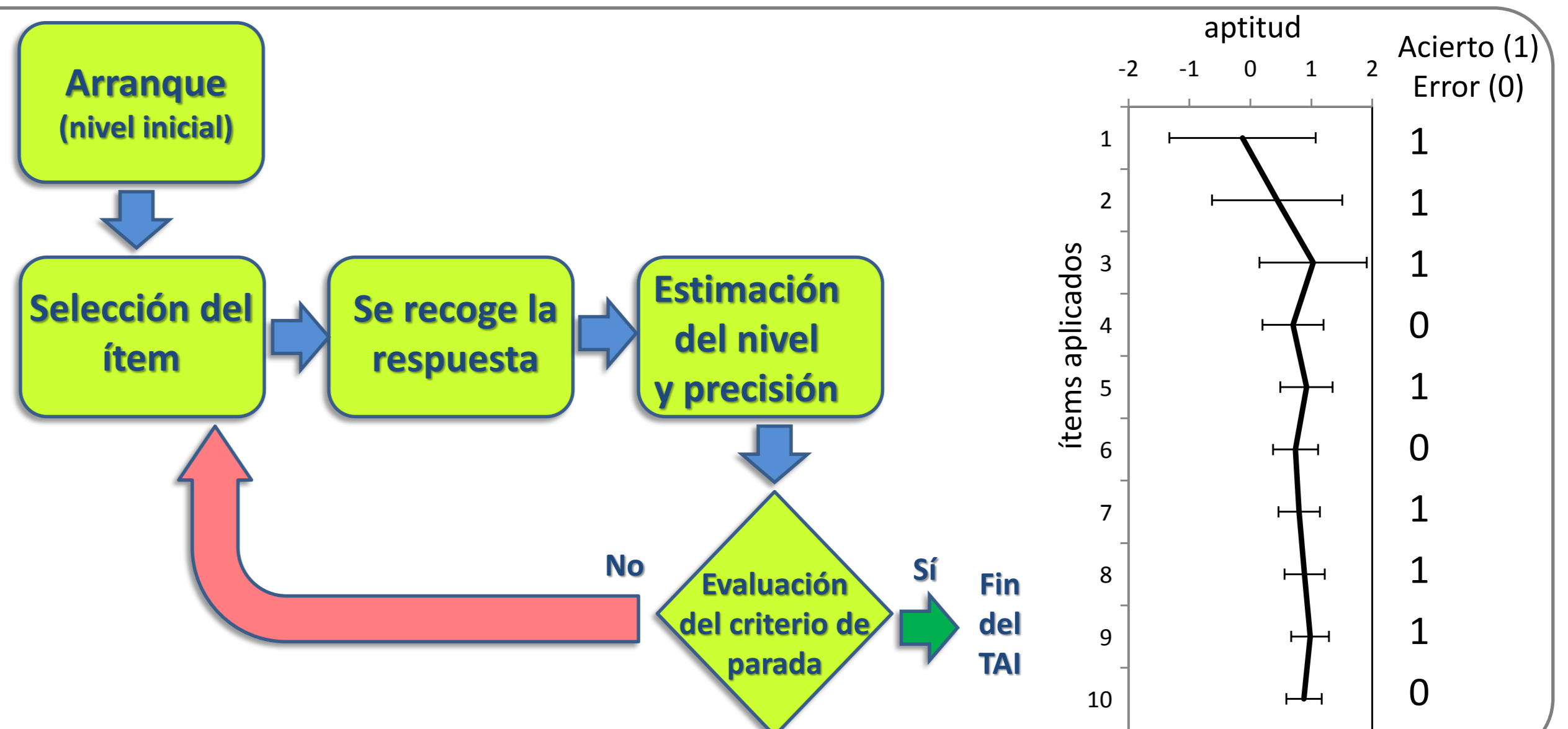


Después de aplicar las preguntas de un test en una muestra de personas, la TRI permite saber la dificultad de un ítem para cada nivel de aptitud.

Es la base matemática de un TAI.

El funcionamiento de un TAI

- Arranque:** al principio del test no sabemos nada sobre el nivel de aptitud del evaluado. Por tanto, se le asigna un nivel intermedio.
- Selección del ítem:** Se selecciona el ítem más "informativo" para ese nivel. En contextos de acreditación, promoción o selección es importante que los ítems no resulten "sobrepuestos" y se incluyen restricciones en su tasa de exposición (p.ej., que cada ítem no sea administrado en más del 20% de los tests).
- Estimación del nivel de rasgo:** mediante procedimientos estadísticos se estima qué nivel de aptitud hace más verosímil el patrón de aciertos y fallos.
- Evaluación del criterio de parada:** Si se ha alcanzado una precisión adecuada (o se ha llegado al máximo número de ítems) el TAI termina. De lo contrario, se vuelve al paso 2.



Pasado, presente y futuro de los TAIs

Un ejemplo concreto: eCat

El pasado de los TAI

La idea de adaptar la dificultad del test al nivel de aptitud del evaluado se remonta a Binet, pero fue **Frederick M. Lord** en el Educational Testing Service, quien a partir de años 70 sentó la base matemática que permitió su desarrollo (Lord y Novick, 1968).

El primer test operacional se desarrolló en el ETS y su versión actual se llama **ACCUPLACER**, que es un test de clasificación (Placement Tests) que permite asignar a los examinados a cursos que se ajusten a su experiencia y sus destrezas educativas (Rojas, 2001).

El presente de los TAI

Algunos ejemplos de TAIs actuales son:

A nivel internacional:

ASVAB Armed Services Vocational Aptitude Battery: para evaluar aptitudes.

ETS TOEFL Test of English as a Foreign Language: para evaluar el nivel de inglés.

GMAT Graduate Management Admission Test: prueba de acceso a escuelas de negocios.

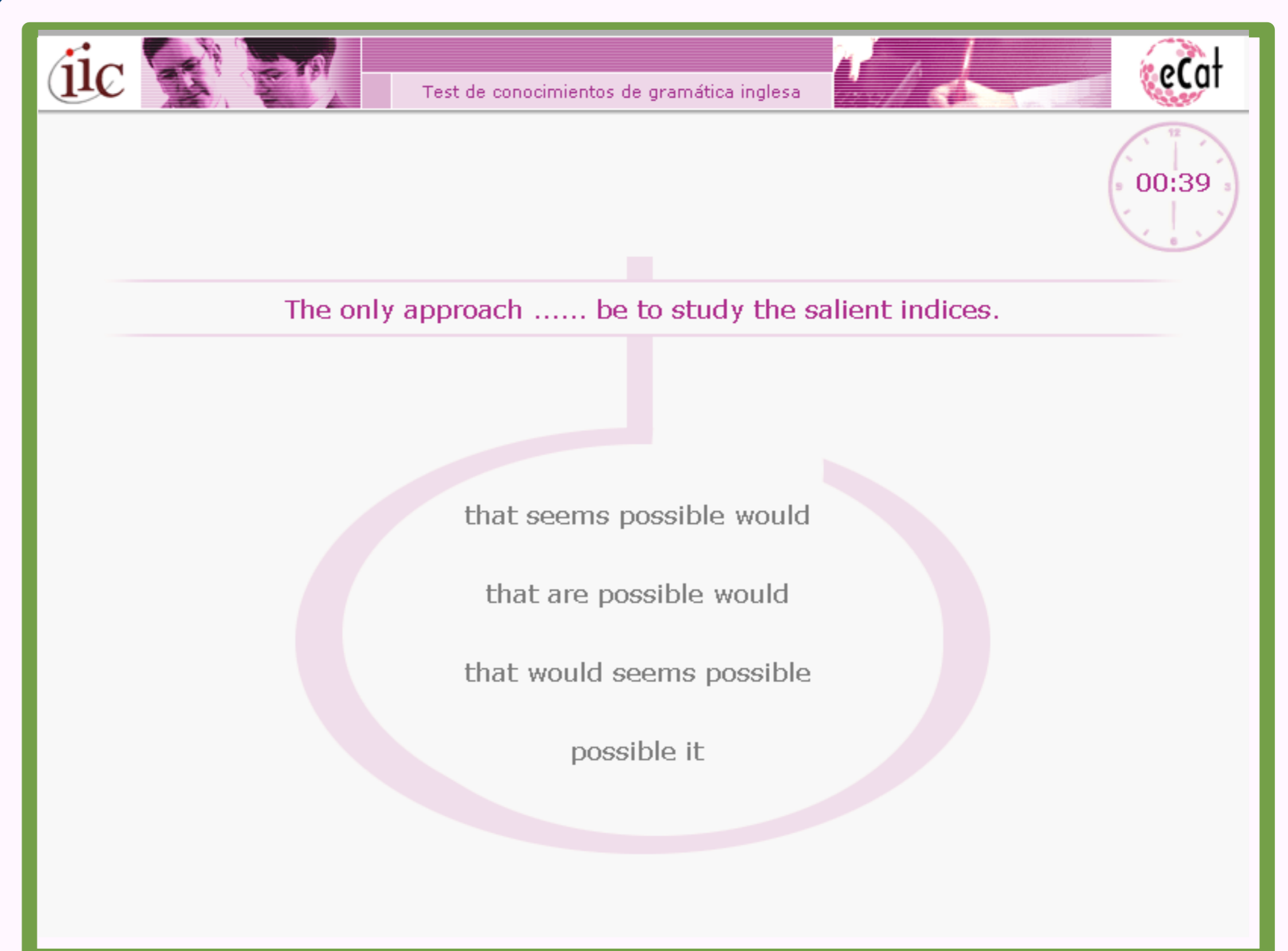
En España:

TRASI (Rubio y Santacreu, 2004): para medir la capacidad de razonamiento secuencial e inductivo)

CAT (Rebollo et al., 2009): para evaluar calidad de vida relacionada con la salud.

El futuro de los TAI

Los TAIs no han estado exentos de críticas, especialmente debidas a su coste (**necesidad de mantener grandes bancos, de desarrollar, evaluar y actualizar software, de disponibilidad de ordenadores para la aplicación, etc.**). El experto **Howard Wainer** (2000) concluye que antes de utilizar un TAI hay que valorar: (a) ¿Con qué frecuencia se aplicará el test? (b) ¿El contenido del test requiere una aplicación informatizada?



eCAT es una prueba de evaluación del nivel de inglés que se compone de dos test: eCAT-grammar (Olea, Abad, Ponsoda y Ximénez, 2004), que mide el nivel de comprensión escrita, y eCAT listening (Olea, Abad, Ponsoda, Barrada, y Aguado, 2011), que mide el nivel de comprensión oral.

Esta prueba se ha construido conjuntamente por el grupo de investigación de la **Cátedra "Modelos y Aplicaciones Psicométricas"** y el **Instituto de Ingeniería del Conocimiento**. Su funcionamiento y las propiedades psicométricas se describen en los siguientes trabajos:

Olea, J., Abad, F.J., Ponsoda, V., & Ximénez, M.C. (2004). Un test adaptativo informatizado para evaluar el conocimiento del inglés escrito: diseño y comprobaciones psicométricas. *Psicothema*, 16, 519-525.

Olea, J., Abad, F. J., Ponsoda, V., Barrada, J. R., & Aguado, D. (2011). eCAT-Listening: design and psychometric properties of a computerized adaptive test on English Listening. *Psicothema*, 23(4), 802-807.

<http://www.iic.uam.es/catedras/map>

Referencias bibliográficas:

- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Olea, J., Abad, F. J., & Barrada, J. R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del psicólogo: revista del Colegio Oficial de Psicólogos*, 31(1), 97-107.
- Olea, J. y Ponsoda, V. (2003). *Tests adaptativos informatizados*. Madrid: UNED.
- Olea, J., Ponsoda, V. y Prieto, G. (1999). *Tests informatizados Fundamentos y aplicaciones*. Madrid: Pirámide.
- Rojas, A. J. (2001). Pasado, presente y futuro de los Tests Adaptativos Informatizados: entrevista con Isaac I. Bejar. *Psicothema*, 13(4), 685-690.
- van der Linden, W. J. y Glas, C. A. W. (2000). *Computerized adaptive testing: Theory and practice*. Londres: Kluwer Academic.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B., Mislav, R., Steinberg, L. y Thissen, D. (2000). *Computerized adaptive testing: A primer* (2ª ed.). Mahwah, NJ: LEA.
- Wainer, H. (2000). CAT: Wether and Whence. *Psicologica*, 21, 121-133

