

## **EVALUACIÓN FORMATIVA Y NUEVAS TECNOLOGÍAS**

Ángel Sobrino, Concepción Naval

Universidad de Navarra

### **RESUMEN**

La evaluación formativa, en tanto que proceso continuo de detección de dificultades en el aprendizaje del alumno, es uno de los elementos fundamentales de la individualización. Las posibilidades de gestión de la información convierten a las nuevas tecnologías en medios muy apropiados para facilitar a los alumnos una evaluación, todo lo frecuente que se quiera, de su aprendizaje. Presentamos en este trabajo un proyecto de desarrollo de una herramienta para la generación de pruebas objetivas en Internet, desarrollado por el Departamento de Educación y el Centro de Tecnología Informática de la Universidad de Navarra. El objetivo del proyecto es implementar un procedimiento de examen y autoevaluación a través del ordenador, que permita a los profesores disponer de un método de valoración inmediata de los conocimientos de los alumnos, y a éstos disponer de un sistema de autoevaluación fiable y de fácil acceso.

### **PALABRAS CLAVE**

Tecnología educativa. Evaluación formativa. Exámenes

## **FORMATIVE EVALUATION AND NEW TECHNOLOGIES**

### **ABSTRACT**

Formative evaluation is one of the most important factors for individualization. By formative evaluation we mean the continuous process of detection of the students' learning difficulties.

New technologies, because of their power to deal with information, have become very appropriate media to provide the student with academic assessments as frequent as desired.

We offer in this paper a project of a tool for generating tests through Internet. This project has been well-developed by the Education Department and the Computer Technology Center at the University of Navarra -Spain-.

The goal of the project is to implement a procedure of exam and self-assessment through the computer. It allows the teachers to have an immediate method of assessment of students' knowledge, and the students to have a reliable and of easy access self-assessment system.

### **KEY WORDS**

Educational Technology. Formative evaluation. Testing.

## INTRODUCCIÓN: EVALUACIÓN FORMATIVA E INSTRUCCIÓN EFICAZ

Si hay un tema que puede originar apasionadas discusiones en el mundo de la educación es el de la eficacia, criterio que se suele considerar aplicable a la instrucción, pero que parece contrapuesto al concepto de formación. Sin embargo, esta supuesta oposición puede encontrar la reconciliación en lo que podríamos denominar evaluación formativa. Entendemos por evaluación formativa un proceso que busca la eficacia, pero que no se opone, sino que precisamente complementa, la mejora de la calidad de la formación de nuestros alumnos.

Cuando Guskey (1) intenta responder a la difícil cuestión de las características de una enseñanza eficaz lo hace de forma indirecta: selecciona a un grupo de profesores que la comunidad educativa considera excepcionales y, a continuación, les entrevista planteándoles cuestiones acerca de su praxis. Son cuatro las características comunes que encuentra en todos ellos: la peculiar organización del curso, la motivación, la participación de los alumnos y la constante retroalimentación al alumno (feedback).

Centrémonos en este último punto. No es ni mucho menos mérito de la Psicología Cognitiva la insistencia en la importancia del feedback en el proceso de enseñanza-aprendizaje, aunque es preciso reconocer la acertada operacionalización que aquélla realiza. El conductismo ya había dejado su impronta en este tema desde los trabajos pioneros de Skinner (2) pasando por el sistema PSI de Keller) o el ya más maduro, desde el punto de vista pedagógico, del *Mastery learning* (4). Habrá sin embargo que esperar a la revolución cognitiva, y posteriormente a la constructivista, para que se acentúe la dimensión informativa frente a la meramente reforzante (5).

De esta forma, dar información precisa a los estudiantes sobre el progreso en su aprendizaje, a intervalos regulares a lo largo de la secuencia instructiva, se convierte en uno de los puntos clave de una buena enseñanza, sea cual fuere la escuela psicológica que sigamos, el modelo didáctico que elijamos o el sistema organizativo que empleemos. Esta información será ocasión para los estudiantes de "identificar lo que es importante aprender, lo que han aprendido bien y a qué necesitan dedicar más tiempo" (6).

Pues bien, el problema de la necesidad de retroalimentación se puede analizar desde varios enfoques. Desde el psicológico, se hará especial hincapié en la adecuación del mensaje a diferentes variables del alumno (edad, nivel previo de conocimientos, dificultades de aprendizaje, actitudes,...); desde el organizativo se tendrá en cuenta especialmente la redefinición del tiempo escolar, la adaptación de espacios, la planificación de materiales instructivos y otros recursos; desde la perspectiva didáctica interesará el concepto de *evaluación formativa*. Y aquí es donde nos vamos a detener.

## COMPONENTES DE LA EVALUACIÓN FORMATIVA

Scriven habla de evaluación formativa (7) refiriéndose a un proceso continuo de detección de dificultades de aprendizaje en el alumno, con la finalidad de determinar los tratamientos pedagógicos que le permitirán progresar. Es por tanto uno de los elementos fundamentales de la individualización.

López (8) destaca tres requisitos del proceso evaluador formativo: la frecuencia, la exigencia y la información. Detengámonos en el análisis que hace este autor de cada uno de ellos, especialmente en el primero y último. La medición frecuente, también denominada por algunos "evaluación continua", es una característica fundamental del proceso instructivo individualizador. Este punto es tan obvio que parece indiscutible, pero una duda práctica puede acechar inmediatamente: ¿cada cuánto tiempo conviene medir? o, si se nos permite la redundancia, ¿cuál es la frecuencia adecuada de la evaluación frecuente? Desde la "inflación de medida" del programa skinneriano al examen único de final de curso hay toda una graduación respecto al número de mediciones del rendimiento del alumno.

Bangert-Drowns, Kulik y Kulik realizan una síntesis sobre los efectos de la frecuente medición en el aula. Una de sus conclusiones sugiere que el aumento en la frecuencia de medición puede incrementar regularmente el rendimiento al final de la enseñanza, pero a partir de cuatro o cinco pruebas el incremento es progresivamente menor (9). Por supuesto, coincidirá el lector con Kulik y Kulik que "para incrementar la efectividad instructiva se precisa algo más que un simple incremento del número de pruebas aplicadas a los estudiantes (10). Ese "algo más" se refiere al uso que se haga de la medición, es decir, a la retroalimentación y consiguiente proceso correctivo, si hiciera falta (11). Es el tercer requisito al que antes nos referíamos.

En líneas generales, el feedback incluye información sobre la ejecución y, además, ofrece la oportunidad de interactuar con el docente y con el propio proceso de aprendizaje. Pero la masificación de las aulas —especialmente las universitarias— convierte la labor evaluadora del profesor. y la consecuentemente correctiva, en un difícil cometido. Sobre todo cuando se intenta aumentar la frecuencia de la valoración del proceso de aprendizaje del alumno. Una solución, desde el punto de vista instrumental, puede proceder del mundo de las Nuevas Tecnologías.

## **NUEVAS TECNOLOGÍAS Y EVALUACIÓN FORMATIVA**

Las posibilidades de gestión de la información convierten a los ordenadores en medios muy apropiados para facilitar a los alumnos una evaluación, todo lo frecuente que se quiera, de su aprendizaje. La máquina tiene aquí una triple ventaja: es capaz de generar un número muy elevado de preguntas, corregirlas inmediatamente y proporcionar al alumno —o al profesor, según sea el caso— información detallada del dominio de los contenidos.

De entre toda la maraña de acrónimos procedentes del mundo anglosajón —que tan acertadamente han estructurado Repáraz y Tourón (12)— las siglas CMI (Computer-Managed Instruction) identifican las tareas de apoyo al proceso de enseñanza-aprendizaje. Y dentro de éstas, las aplicaciones informáticas para el diagnóstico y valoración de sujetos han cumplido ya las tres décadas, con todo un conjunto de experiencias que durante este período han configurado un corpus de conocimiento para el que el mundo anglosajón acuñó el término CAT (Computer-Assisted Testing). El proceso de automatización en la generación, aplicación y corrección de las pruebas estandarizadas es hoy una realidad aceptada entre los profesionales de la docencia (13).

Por otro lado, la "revolución telemática" aporta una plataforma de trabajo muy beneficiosa para la evaluación formativa. Internet —en este caso mejor decir Intranet— constituye una mejora no solo cuantitativa sino de orden cualitativo en la esfera del CAT. Son dos las ventajas que proporciona. En primer lugar, no se necesitan conocimientos de ninguna aplicación concreta de CAT, puesto que a través de cualquier navegador se accede a un interface de usuario extremadamente sencillo, de forma que docentes y alumnos no necesiten ningún conocimiento informático para generar y resolver, respectivamente, las pruebas estandarizadas. Y en segundo lugar, ofrece la posibilidad de centralizar toda una serie de recursos en un único servidor; con acceso remoto desde cualquier ordenador conectado a la red. Además de estos dos aspectos, las ventajas de la "tele-educación" en algunos campos como la formación de adultos son tan obvias que no parece necesario insistir más sobre ello en estas líneas.

Pero, ¿se puede reducir la actividad evaluadora a las pruebas estandarizadas u objetivas? Quizá merezca la pena hacer un paréntesis en nuestra argumentación antes de continuar.

### **LAS PRUEBAS OBJETIVAS**

El problema de la estandarización de las pruebas de evaluación de alumnos es de una controversia considerable en el actual panorama pedagógico. Por un lado, la masificación de la enseñanza empuja a los docentes a utilizar técnicas que permitan una gestión automática —o al menos semiautomática— de los exámenes. Pero, por otra parte, pueden resultar escasamente representativos de algunos de los conocimientos de los alumnos, en especial en el área de expresión.

La cuestión que subyace en esta controversia es la dialéctica entre estandarización e individualización. Sin embargo debe tenerse en cuenta que ambos no son conceptos necesariamente contradictorios; más bien es el tipo de uso que se realice de las pruebas, lo que puede enfrentarlos (14). El debate es harto complejo y excede las pretensiones de estas páginas, pero nos parece adecuado sentar una serie de principios. En primer lugar, apostar por la complementariedad —y no la exclusión— entre pruebas objetivas y no objetivas (ensayo, exámenes orales, etc.). En segundo lugar, tener presente que se trata de técnicas de evaluación y que, como tales, deben subordinarse no sólo a la naturaleza de la materia, sino sobre todo a los objetivos de formación que pretendamos.

Aclarada, o al menos introducida, esta cuestión, vamos a detenernos en las posibles aplicaciones de las pruebas objetivas a lo largo del proceso de enseñanza-aprendizaje. Como puede observarse en el cuadro adjunto (Tabla 1), son al menos cinco las finalidades que puede perseguir una prueba estandarizada. En el contexto en el que nos movemos, tres son especialmente importantes para la docencia universitaria: las referentes a las pruebas de nivel, las pruebas de clase y las de propósito diagnóstico.

Por lo que se refiere a las "pruebas de amplio espectro" son conocidas las virtualidades de este sistema para generar y corregir pruebas de admisión, con decisiones sumativas asociadas.

Dimensión de valoración	Tipo de prueba estandarizada				
	1. AMPLIO ESPECTRO Indicador de resultados Certificación Admisión	2. DE NIVEL O DOMINIO	3. DE CLASE	4. DE PROPÓSITO DIAGNÓSTICO	5. INDIVIDUALIZADAS
Amplitud del dominio educativo	Amplio	Mixto	Reducido	Mixto	Reducido
Límites del dominio educativo	No finitos difusos	Mixto	Finitos concretos	Finitos concretos	Finitos concretos
Dimensión	Multidimensional	Multidimensional	Unidimensional	Mixtos	Unidimensional
Amplitud de la población	Amplia/ Muy amplia	Intermedia/ amplia	Reducida/ muy reducida	Amplia/ Reducida	Reducida
Grado de la diversidad de la población	Heterogénea	Mixto	Homogénea	Heterogénea	Homogénea
Decisiones Asociadas	Formativas/sumativas Sumativas	Sumativas	Formativas/ sumativas	Formativas	Formativas/ sumativas
Unidades sobre las que se decide	Grupos Individuos	Individuos	Individuos	Individuos	Individuos y/o grupos
Tipo de Estándar	Mixto: normativo Normativo criterial	Criterial	Criterial	Normativo criterial	Criterial

Tabla 1. Tipos fundamentales de pruebas estandarizadas valorados respecto a dimensiones de caracterización (Jornet y Suárez, 1992)

Se consiguen valorar, de forma rápida y poco costosa, conocimientos en diversas áreas de una población amplia y heterogénea. En algunas Facultades de la Universidad de Navarra ya se tiene experiencia de este tipo de aplicaciones con resultados satisfactorios.

Las así denominadas "pruebas de clase" se refieren, en nuestro ámbito universitario, a los exámenes sumativos —en febrero, junio y septiembre— de las diversas asignaturas. Aunque dirigidas a una población sensiblemente menor que en el caso anterior, ésta sigue siendo de un volumen considerable. Además, la premura de tiempo por razones administrativas invita a adoptar formas de examen asistidas por ordenador.

Pero quizá las pruebas de propósito diagnóstico sean las de más interés pedagógico si, como estamos intentando mostrar, la evaluación formativa es uno de los elementos fundamentales de cualquier modelo de enseñanza-aprendizaje de calidad. La posibilidad que ofrece este sistema para la autoevaluación y el conocimiento inmediato de los resultados va a permitir que los alumnos conozcan cuál es su nivel de dominio de la materia. De esta manera se satisfacen los requisitos de inmediatez, continuidad y carácter formativo del proceso evaluador, que se tornan dificultosos cuando el profesor se ve obligado a valorar pruebas no estandarizadas.

Vamos a presentar a continuación un proyecto para el desarrollo de una herramienta de gestión de pruebas objetivas desarrollado por el Departamento de Educación y el Centro de Tecnología Informática de la Universidad de Navarra. El objetivo del proyecto es implementar un procedimiento de examen y autoevaluación a través del ordenador, que permita a los profesores disponer de un método de valoración inmediata de los conocimientos de los alumnos, y a éstos disponer de un sistema de autoevaluación fiable y de fácil acceso.

## UNA APLICACIÓN PARA LA GESTIÓN DE PRUEBAS OBJETIVAS

El proyecto, en fase de realización, pretende desarrollar una aplicación interactiva y multimedia en Internet para generar, gestionar, administrar y corregir de forma automática pruebas objetivas de diferentes modalidades. Como lenguaje multiplataforma se ha utilizado Java, modularizado en applets independientes que se comunican mediante CGI's programados en Perl y con utilidades escritas en JavaScript.

El sistema informático se implementará en Internet utilizando como soporte una base de datos MySQL, consultada a través de hojas Web que están situadas en un servidor HTTP Apache instalado en una máquina UNIX Digital. Sobre este servidor y a través de cualquier ordenador conectado a la red de la Universidad, el profesor podrá seleccionar el tipo de examen que quiere realizar. Esto da lugar a que el profesor pueda configurar el examen desde cualquier punto de la Universidad, y mantener unas preguntas disponibles siempre que quiera realizar una prueba. La figura 1 representa la página html de Internet de entrada al generador de exámenes.

### FUNCIONES DEL SISTEMA

#### 1. Construcción del banco de items

El profesor debe ir creando, a lo largo del tiempo, un banco de items que la aplicación gestiona como una base de datos. El sistema permite mantener esta base de datos remota y personalizada (añadiendo, modificando o eliminado items), sin que el usuario disponga de conocimientos informáticos ni software especializado.

Figura 1: Pantalla de acceso al programa

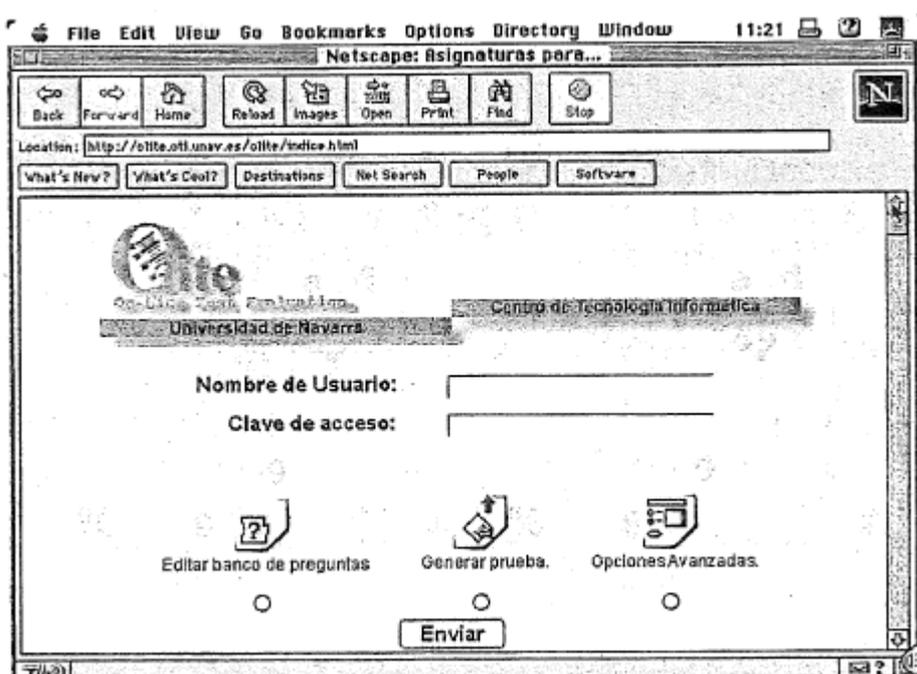
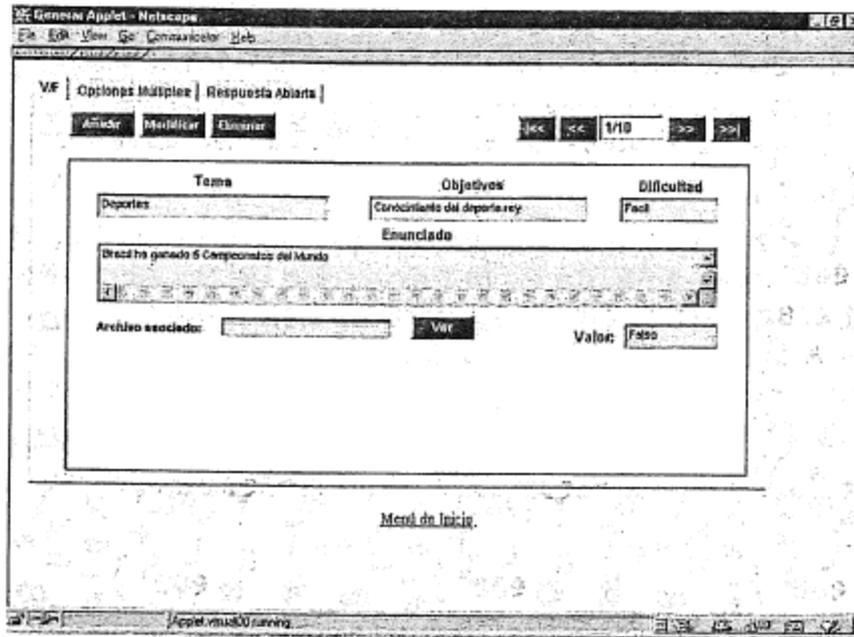


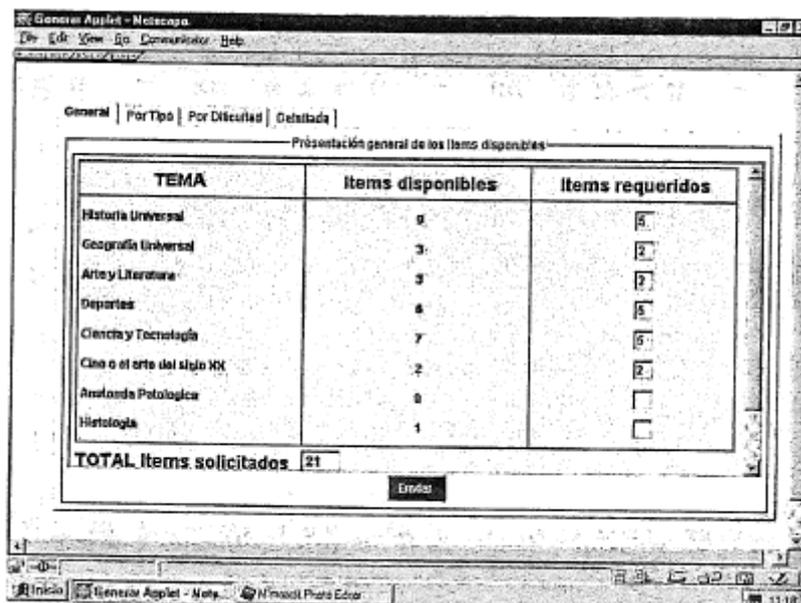
Figura 2: Introducción de un ítem (modalidad verdadero/falso)



## 2. Generación de una prueba

Una vez que se define el tipo de examen, el programa concreta la prueba para esa asignatura y ese modelo de preguntas. A partir de este momento, se determinan otros datos necesarios, como el formato y número de preguntas por tema, su dificultad, el número de alternativas, la duración del examen, y los datos del aula y la fecha de realización. Si se elige el modo de generación automático, el programa selecciona de manera aleatoria el número total de ítems requeridos del total de disponibles en la base de datos. Por el contrario, si la generación es manual, el usuario podrá elegir uno a uno todos los ítems que compondrán la prueba (cfr. Fig 3).

Figura 3: Generación de una prueba: elección manual de ítems



### **3. Administración de la prueba**

Una vez preparado el examen, el alumno accederá al mismo con un navegador de internet (Nescape, Internet Explorer, etc.), el día de la prueba (en el caso de pruebas sumativas) o bien durante un período determinado (pruebas formativas), desde cualquier lugar del campus o incluso su propia casa. El alumno se identifica y automáticamente se genera un conjunto de ítems, extraídos de la base de datos, según el perfil que ha diseñado el profesor.

### **4. Corrección y análisis de resultados**

Una vez finalizada, el sistema enviará por correo electrónico los resultados del examen y si lo pide, el análisis de las pruebas. Se podrán analizar las características psicométricas de la prueba: fiabilidad (consistencia interna), análisis de los ítems (los distractores y el índice de dificultad y homogeneidad) e ITR (Item Response Theory).

### **5. Seguimiento del alumno**

Este sistema permite registrar todas las respuestas que el alumno realiza, de tal forma que posibilita un análisis y seguimiento de la trayectoria de los alumnos en distintas pruebas. Además corrige la posibilidad de pérdida de exámenes o equivocaciones en los resultados ya que son obtenidos de manera automática.

Aunque estamos aún en fase de realización, este proyecto promete ser una ayuda adecuada para la mejora de la calidad del proceso de enseñanza-aprendizaje.

(1) Cfr. Guskey, T. R., *Improving Student Learning in College Classrooms*, New York. Thomas Charles C. Pub.. 1987.

(2) Cfr. Skinner, B. F.: "The Science of Learning and the Art of Teaching", *Harvard Educational Review*, 24 (2) (1954) pp. 86-97.

(3) Cfr. Keller, E S.: "Good bye, teacher...". *Journal of Applied Behavior Analysis*, 1 (1968) pp. 79-89.

(4) Cfr., por ejemplo, Anderson. L. W. y Block. J. H.: *Mastery learning model of teaching and learning*, en: T. Husen, T y Postlewait, T. N. (eds.): *The International Encyclopedia of Education*, Oxford, Pergamon, 1985, pp. 3219-3230.

(5) Cfr. Sobrino. A.: *Nuevas Tecnologías aplicadas a la educación universitaria. Evaluación de un sistema hipermedia*, Pamplona, Departamento de Educación. Universidad de Navarra, Tesis doctoral inédita, 1995.

(6) Guskey. T. R.: "The essentials elements of mastery learning", *Journal of Classroom Interaction*, 22 (1987) pp. 19-22.

(7) Cfr. Scriven, M.: *Evaluation Thesaurus*, California. Edge Press. 1981.

- (8) Cfr. López, E.: Proyecto docente e investigador de Pedagogía Diferencial, Madrid. Universidad Complutense. Memoria de Cátedra, 1996.
- (9) Cfr. Bangert-Drowns. R. L. y Kulik. J. A., y Kulik, C. C.: "Effects of Frequent Classroom Testing". The Journal of Educational Research, 85 (2) (1991) pp. 89-99.
- (10) Kulik, J. A. y Kulik, C. C.: "Meta-analysis in education". International Journal of Educational re-search, 13 (3) p. 295.
- (11) Cfr. López, E.: op. cit.
- (12) Cfr. Reparaz, Ch. y Tourón, J.: El aprendizaje mediante ordenador en el aula, Pamplona. Eunsa, 1992.
- (13) Cfr. Sampson, J. P.: "Computer Applications and Issues in Using Test in Counselling", en Watkins, C. E. y Campbell. V. L (eds.): Testing in Counseling Practica, Hillsdale NJ. Lawrence Erlbaum, 1990.
- (14) Cfr. Jornet, J. M. y Suárez. J. M.: "Pruebas estandarizadas y evaluación del rendimiento". Revista de Investigación Educativa, 14 (2) (1996) pp. 141-163.