

EFFECTOS DE VARIABLES ASOCIADAS AL CAPITAL CULTURAL DE LOS JUECES SOBRE LA VALIDACIÓN DE UNA ESCALA DE VALORACIÓN DE REACTIVOS OBJETIVOS*

Rodrigo García Leija**

Resumen

Se validó una escala para evaluar fuentes de contenido en reactivos objetivos y se investigó la incidencia del capital cultural de los jueces sobre esta validación. Participaron 18 jueces con experiencia en medición de resultados de aprendizaje. Se utilizó un diseño de dos factores (2x3), donde el factor "A" fue la disciplina (psicología y pedagogía) y el factor "B" fue el nivel académico (licenciatura y posgrado) de los jueces. Los acuerdos entre jueces validaron las definiciones dadas previamente a las seis dimensiones que conforman la escala ($Q = 20.3$, 22 gl., $p = 0.56$), pero no a los puntajes asignados ($Q = 40.93$, 20 gl., $p = 0.004$). El análisis de regresión múltiple mostró que las incidencias del capital cultural de los jueces afectaron la validación sólo en la asignación de puntajes y no en las definiciones previamente dadas.

Palabras claves: Medición, resultados de aprendizaje, análisis de regresión múltiple.

Fecha de recepción: Agosto de 2004
Fecha de aceptación: Octubre de 2004

* Esta investigación fue financiada por el proyecto CONACYT 40608 H.

** Psicólogo, Facultad de Psicología UNAM. rogalei@yahoo.com.mx

Abstract

A scale was validated to evaluate content sources in reactive objective and the incidence of the cultural capital of the judges on this validation was investigated. 18 judges with experience in measurement of learning results participated. A two-factor design was used (2x3), where factor "A" was discipline (psychology and pedagogy) and factor "B" was the academic formation (B.A. or graduate degree) of the judges. The agreements between judges validated the definitions given previously to the six dimensions that conform the scale ($Q = 20,3, 22 \text{ gl.}, p = 0,56$), but not to the assigned scales ($Q = 40,93, 20 \text{ gl.}, p = 0,004$). The multiple regression analysis showed that the incidence of cultural capital of the judges affected the validation only in the allocation of points and not in the definitions previously given.

Key words: Measurement results of learning, multiple regression analysis.

Dentro de la medición de resultados de aprendizaje es común medir sólo lo que los alumnos saben y dejar de lado la forma en la que los alumnos aprenden, recuerdan y aplican los conocimientos ante contextos diversos. Una característica de los exámenes, en nuestro sistema educativo, es que pueden predecir el fracaso pero son insensibles para predecir el éxito académico porque no miden los mecanismos que lo producen (Castañeda, 1993).

Sin embargo y a pesar de las fuertes críticas hechas a los exámenes objetivos estandarizados, éstos continúan siendo los instrumentos más utilizados en la medición de resultados de aprendizaje a gran escala. Los reactivos más utilizados en ellos son los de opción múltiple, y es común que éstos presenten errores en su construcción. Si como dijo Messick (1989), la validez del atributo que se mide es una propiedad de las mediciones, los puntajes reflejan entonces una intrincada relación entre el atributo que se mide y el error de medición.

Así, estimar la cantidad de error contenida en las mediciones constituye una necesidad que debe ser satisfecha, dado que el error de medición es el principal responsable de la falta de precisión de los puntajes. Pero el error puede generarse desde la misma construcción

del reactivo, por ejemplo, demandar habilidades lingüísticas superiores en ítems de resolución de problemas trigonométricos. Este error puede afectar, sistemáticamente, el puntaje en el examen.

Con base en lo descrito hasta aquí, estudios que caractericen tareas y reactivos constituye una importante línea de generación de evidencia a favor del atributo que se intenta medir y aun cuando se recurre, usualmente, a expertos para generar las evidencias, en variadas ocasiones, se presenta un buen número de limitaciones: insensibilidad de los jueces para identificar gradientes de dificultad en las diversas fuentes que componen el ítem o la tarea, fatiga generada por analizar bancos extensos de reactivos y niveles de rigor diferenciales entre jueces.

El problema

A partir de lo planteado hasta aquí, un problema que debe ser considerado –debido a la importancia que sus efectos revisten sobre la validez– es el hecho de que la medición a gran escala descansa, mayoritariamente, en formatos objetivos que intentan medir aprendizajes complejos. En la actualidad, necesidades prácticas acerca de: a) estandarizar procesos de medición y b) abatir costos en calificar formatos de respuesta construida, ha generado el uso indiscriminado de este tipo de reactivos objetivos.

Los exámenes objetivos son instrumentos conformados por ítems que miden la habilidad de los sustentantes para ganar puntajes, pero ¿cuáles de sus rasgos facilitan o dificultan resolverlos? Los exámenes miden el logro de los sustentantes para hacer las cosas que las tareas requieren y, estas cosas, son las demandas que cada ítem hace al sustentante. Si las teorías que fundamentan los exámenes no pueden explicar por qué determinado ítem es más difícil que otro, entonces no podrán ser útiles para entender el proceso de responder, y menos aun para derivar recomendaciones que impacten los cambios requeridos.

MÉTODO

La investigación empírica persiguió dos objetivos, validar una escala de valoración de fuentes de facilidad - dificultad de reactivos objetivos, mediante validación de jueces con experiencia en construcción de reactivos objetivos e identificar incidencias de tres variables del capital cultural de los jueces al validar los componentes de la escala en estudio.

Diseño de investigación

Para llevar a cabo la investigación se utilizó un diseño de dos factores, A x B (2x3), donde el factor "A" fue la disciplina (psicología y pedagogía) y el factor "B" fue el *nivel académico* (licenciatura y posgrado). Se controló la experiencia (en años de construcción de reactivos).

Participantes y sus funciones

Participaron 18 jueces, éstos conocieron la escala de valoración de fuentes de dificultad - facilidad de reactivos objetivos y validaron las definiciones y la asignación de puntajes dados a las seis dimensiones que contiene la escala: *formato del reactivo, operación cognitiva demandada, contexto en el que se presenta el reactivo, campo de conocimiento que mide, redacción del reactivo y el uso técnico del contenido del reactivo.*

Herramientas

Se utilizaron tres instrumentos de lápiz y papel. Dos para los que los jueces respondieron a preguntas relacionadas con: **a)** reactivos referidos a su capital cultural y **b)** reactivos para validar los componentes de la escala en cuestión. El tercer instrumento fue **c)** la Escala de Valoración de Fuentes de Facilidad - Dificultad incluidas en reactivos objetivos de exámenes a gran escala. Se describen, a continuación, en el orden en el que fueron aplicados.

1) *Cuestionario de Capital Cultural*. El objetivo del cuestionario es identificar las características del capital cultural de los jueces. Está

compuesto de cinco áreas: datos personales, familiares, formación académica y trayectoria laboral, productividad académica y divulgación. Los reactivos que integran el instrumento son, en total, 68, distribuidos en número diferente en las cinco áreas. Los reactivos son de dos tipos: semiabiertos y de elección forzada.

2) *Escala de Valoración de Fuentes de Facilidad - Dificultad de Reactivos Objetivos de exámenes a gran escala.* Es una escala conformada por 55 reactivos; distribuidos en dimensiones como: formato del reactivo, operación cognitiva, contexto, campo de conocimiento, redacción y uso técnico apropiado. Los primeros 21 reactivos pertenecen a las dimensiones: Tipo de formato, con 5 reactivos, Operación cognitiva, con 9 reactivos, Contexto, con 4 reactivos, y Campo de conocimiento, con 3 reactivos; éstos están compuestos por la interacción entre el reactivo y el tipo de conocimiento con el que interactúa, ya sea factual, conceptual o procedimental. Los 24 reactivos restantes se reparten en la dimensión de redacción, con 22 reactivos, y en la dimensión de Uso técnico apropiado, con 12 reactivos, éstos reactivos se estructuran en forma lineal, es decir, sólo tienen un valor, su función es identificar fuentes de facilidad o dificultad en reactivos objetivos de exámenes a gran escala. La escala fue validada previamente por cinco expertos, todos con doctorado, en construcción de reactivos y diseño de exámenes.

3) *Escala de validación de los componentes de la escala arriba descrita:* se compone de dos dimensiones que refieren a las definiciones que se manejan en la escala de valoración de fuentes de facilidad - dificultad de reactivos objetivos de exámenes a gran escala y a la asignación de puntajes que maneja esta escala. Además, incluye una porción en la que se pide a los jueces su opinión acerca de la utilidad que representan cada uno de los indicadores.

- La primera dimensión presenta 23 reactivos que validan la *definición* que los autores de la escala le dieron a cada uno de los componentes que la conforman. Por ejemplo, la definición al Tipo de formato utilizado en el reactivo con sus correspondientes indicadores: simple, ordenamiento, apareamiento, FVM y canevá, la

definición dada al Nivel Cognitivo que se demanda para contestar el reactivo con sus indicadores en comprensión y organización, en aplicación y en resolución, la definición en Tipo de contexto que se refiere al tipo y ubicación de las declaraciones y relaciones semánticas que deben procesarse durante la tarea, con sus correspondientes indicadores E - R, R - E, C E - R y C R - E, la definición de Campo de conocimiento dependiendo del conjunto de saberes evaluados en el reactivo, con sus correspondientes indicadores conocimiento teórico, técnico y combinado, la definición de Redacción refiriéndose a la forma en la que se expresan las demandas textuales en el reactivo y la definición de Uso técnico apropiado utilizado en el reactivo. Los jueces valoran las definiciones que subyacen a cada dimensión y a cada indicador que incluye la escala. Los jueces califican con 0 si están en desacuerdo y 1 si están de acuerdo con las definiciones.

- La siguiente dimensión es la correspondiente a la *asignación de puntajes* que se les da a los diferentes indicadores de la escala, con base en los niveles de demanda solicitados en el reactivo. Así, los puntajes que maneja la escala de valoración de fuentes de facilidad - dificultad, fueron pensados para asignar el valor del puntaje dependiendo de la dimensión que mide el indicador en interacción con las demandas derivadas de los tipos de conocimiento que se miden.

Por ejemplo, el indicador de identificación dentro de la dimensión de operación cognitiva con un conocimiento factual, tendrá un valor de dificultad de 1, es decir, la demanda cognitiva en este reactivo será fácil; en cambio, si en la misma dimensión nos encontramos con un reactivo con la demanda de jerarquización en un conocimiento procedimental, el valor de dificultad será de 9, es decir, un reactivo con una demanda cognitiva de muy alta dificultad. La actividad solicitada en este rubro a los jueces es valorar si el puntaje que determina la escala de valoración de fuentes de facilidad - dificultad es el adecuado o no. Califican con 0 si están en desacuerdo y con 1 si están de acuerdo.

Procedimiento

Para llevar a cabo la investigación se utilizó un diseño factorizado, A x B (2x3), donde uno de los factores fue la disciplina profesional de los jueces, con dos valores (*Psicología y Pedagogía*), otro factor fue el grado académico alcanzado por los jueces, con dos valores (*licenciatura y posgrado*). Los años de experiencia que los jueces tienen acerca de construcción de reactivos, fue una variable atributiva que se controló pero no entró en el diseño. Se consideraron dos valores (*poca experiencia (-) de 1 a 2 años y mucha experiencia (+) de 2 años*).

Definición de variables

- VD 1. Número de acuerdos y desacuerdos entre jueces en la valoración de la definición de dimensiones e indicadores asociados a las fuentes de facilidad - dificultad incluidas en la escala de valoración puesta a prueba.
- VD 2. Número de acuerdos y desacuerdos entre jueces en la valoración de la asignación de puntajes a los indicadores asociados a las fuentes de facilidad - dificultad incluidas en la escala de valoración puesta a prueba.

Variables atributivas

- Variable atributiva 1: Disciplina (D) a la que pertenecen los jueces, con dos valores psicología y pedagogía. Representa la carrera profesional que los jueces han cursado y en la que desarrollaron, hipotéticamente, conocimientos y habilidades acerca del diseño de exámenes y construcción de reactivos.
- Variable atributiva 2: Grado académico (GA) alcanzado por los jueces con dos valores: licenciatura y posgrado.
- Variable atributiva 3: Experiencia en la construcción de reactivos objetivos y diseño de exámenes (EDCR). Refiere a los años que los jueces han invertido en medición de resultados de aprendizaje con

dos valores: poca experiencia (1 a 2 años), mucha experiencia ((+) de 2 años).

RESULTADOS

Proceso de validación de la escala

Los jueces validaron la escala en estudio conforme a dos aspectos de interés: la definición de cada dimensión y de los indicadores que las constituyen y la de los puntajes asignados a cada indicador de la escala.

La mayoría de las definiciones dadas a los indicadores muestra buen acuerdo entre jueces (tabla 1). El valor de Q con 18 jueces fue de $Q = 20.3$; $gl = 22$, con un nivel de significancia = 0.56. El índice obtenido muestra que no existen diferencias significativas entre las frecuencias de acuerdos y desacuerdos de los jueces, lo que permite establecer que las definiciones fueron consideradas válidas por ellos.

La frecuencia de desacuerdos de los jueces aumentó en la validación de los puntajes asignados a los indicadores de las cuatro dimensiones que interactúan con el tipo de conocimiento (tabla 2). La muestra de jueces también fue de 18 y los valores arrojados por el análisis ($Q = 40.934$; $gl = 20$ con un nivel de significancia = .004) identificaron diferencias significativas en las valoraciones de los jueces con relación a los puntajes asignados previamente a los indicadores por lo que los puntajes asignados no fueron validados por los jueces.

Cuando se corrió el análisis para las dimensiones en las que la asignación de puntajes no interactuó con el tipo de conocimiento que se está evaluando, la Q no fue significativa. Como se puede observar (tabla 3), la mayoría de los puntajes asignados a los indicadores de estas dos dimensiones muestra acuerdos entre jueces.

En resumen, y con base en los análisis realizados, es posible establecer que los jueces validaron las definiciones dadas a todas las dimensiones y a todos los indicadores de la escala. En cambio, los

puntajes asignados a los indicadores de formato del reactivo, de operación cognitiva, de contexto y de campo de conocimiento no fueron validados por los jueces, pero los puntajes asignados a los indicadores de redacción y uso técnico apropiado se consideraron válidos.

Incidencia del Capital Cultural de los jueces al validar la asignación de puntajes de la escala en estudio

Se corrieron análisis de regresión lineal múltiple stepwise con los datos de la valoración de la asignación de puntajes para cada variable del capital cultural.

El análisis de las coocurrencias entre *Nivel académico de los jueces y la asignación de puntajes* identificó, en la dimensión tipo de formato, que el indicador formato simple tiene una R (correlación parcial) = 0.535, significativa al .02, es decir, explica la varianza en el 53%. El coeficiente de determinación fue de $R^2 = 0.241$.

El análisis de las coocurrencias entre *Experiencia en construcción de reactivos de los jueces y la asignación de puntajes* identificó, en la dimensión Uso técnico apropiado, que el indicador U2 explicó el 53% de la varianza ($R = 0.535$, significativa al 0.02, con $R^2 = 0.24$).

El análisis de las coocurrencias entre *Disciplina profesional de los jueces y la asignación de puntajes* se encontraron más indicadores que explican la varianza, éstos pertenecientes a las siguientes dimensiones: el indicador Traducir, técnicas, procedimientos y rutinas (TTPR) de la dimensión Operación Cognitiva en el nivel de aplicación del conocimiento ($R = 0.47$, $p = .048$); el indicador Caso regla - ejemplo (CR - E), de la dimensión contexto que acompaña al reactivo ($R = 0.64$, $p = .004$), donde la R^2 fue de 0.38 para todas las variables que incluyó el modelo de regresión; el indicador No contiene todas las palabras comunes a las opciones (A4) de la dimensión de redacción ($R = 0.50$, $p = .03$) y el indicador No emplea opciones sinónimas (U6) dentro de la dimensión uso técnico apropiado ($R = 0.50$, $p = .03$).

También se realizaron análisis estadístico de ji cuadrada para determinar si existían diferencias significativas entre las variables del capital cultural de los jueces. Diferencias de los acuerdos entre psicólogos y pedagogos (disciplina); entre jueces con licenciatura y jueces con posgrado (nivel académico) y entre jueces con menos experiencia en construcción de reactivos y jueces con más experiencia en construcción de reactivos.

En resumen, no se encontraron diferencias significativas para las definiciones de los indicadores a partir de las variables del capital cultural de los jueces, pero sí en la asignación de puntajes.

CONCLUSIONES

La forma en la que se realiza la medición de resultados de aprendizaje, en la actualidad, ha llevado a buscar evaluaciones que tomen en cuenta no sólo el producto que el estudiante crea, sino además comprender el proceso por el que pasa para llegar a dicho producto. Este nuevo enfoque de la medición de resultados de aprendizaje dará luz a la comprensión de los mecanismos que le facilitan o le obstaculizan al estudiante realizar una tarea en particular.

El reactivo, herramienta con la que los exámenes evalúan los resultados de aprendizaje de los estudiantes, requiere ser valorado en una diversidad de fuentes de contenido. De aquí que la escala construida para hacerlo permita valorarlo y, en consecuencia, diseñarlo tomando en cuenta componentes que aseguren una medición más precisa.

La validación de la escala, como primer objetivo de la tesis, se logró de manera integral, en cuanto a las Definiciones de las dimensiones y sus indicadores. En cambio, en la Asignación de puntajes sólo se logró en dos dimensiones, la de redacción del reactivo y la del Uso técnico apropiado. Es importante destacar que en esta valoración, los jueces fueron insensibles a la demanda cognitiva generada en la estimación de la asignación de puntajes, dado que ésta requería tener en cuenta la interacción entre el tipo de formato, la operación

cognitiva demanda, el contexto que acompaña al reactivo o el campo de conocimiento evaluados con el tipo de conocimiento que se utiliza en esos componentes. Esta situación se pudo deber a que los jueces estén más habituados a formas tradicionales de evaluar reactivos donde sólo se maneja una propiedad a la vez y de manera aislada de las demás.

Poder afirmar que sí existe una relación entre el capital cultural de los jueces y la manera en la que validan la escala, hace notoria la necesidad de capacitarlos en construcción de reactivos basada en los nuevos paradigmas que se manejan en la medición de resultados de aprendizaje en la actualidad.

Referencias

- Ahmed, A. & Pollit, A. (2002). The Support Model for Interactive Assessment. Paper presented at the 2002 *International Association for Educational Assessment Conference*.
- (1999). Curriculum Demands and Question Difficulty. Paper presented at the 1999 *International Association for Educational Assessment Conference*.
- Aiken, L. R. (1996). *Tests psicológicos y evaluación* (caps. 3, 4 y 5). México: Prentice May.
- Basil, Bernestein, (1988). *Poder, Educación y Conciencia: Sociología de la Transmisión Cultural*. Santiago, Chile: CIDE.
- Berger L, P. & Luckmann, T. (1991). *La construcción social de la realidad*. Buenos Aires: Amorrortu.
- Bourdieu, P. & Passeron, J.-C (1977). *La Reproducción, elementos para una teoría del sistema de enseñanza*. Barcelona: Laia.
- Carreño, H. F. (1990). *Instrumentos de medición del rendimiento escolar* (caps. 2, 3, 4). México: Trillas.
- Castañeda, S. (en revisión). Evaluación y fomento de la actividad cognitiva constructiva y autorregulatoria en actividades educativas ordinarias. En V. M. Alcaraz & L. Reidl, *Las aportaciones mexicanas a la Psicología*. México: UNAM, U. de G.
- Castañeda, S. (2004). Enseñanza estratégica. Guía abreviada para el docente y el tutor. En S. Castañeda, *Educación, Aprendizaje y Cognición. Teoría en la práctica* (pp. 393-421). México: UNAM-U. de G.-Manual Moderno.
- (2003a). Construyendo puentes entre la teoría y la práctica. *Pensamiento Educativo*, 32, 155-176. Chile: Universidad Pontificia de Chile.

- (2002). A cognitive model for learning outcomes assessment. *International Journal of Continuing Engineering Education and Life-long Learning*, 12(1-4), 106. UNESCO:UK.
- (2002). Diagnostividad e interpretabilidad: retos a la medición de resultados de aprendizaje. En A. Bazán & A. Arce (Eds.). *Estrategias de Evaluación y medición del comportamiento* (pp. 247-268). México: ITSON y UADY.
- (2000). Un modelo cognitivo para la medición de resultados de aprendizaje. *Revista de Psicología Contemporánea*, 7 (7), 92-96.
- (2000b). *La medición de resultados de aprendizaje en la enseñanza de ciencias*. Enseñanza-Aprendizaje de las Ciencias (p. 39-53). México: SISIERRA-CONACYT.
- (2000). *Análisis cognitivo de tareas*. Laboratorio de Desarrollo e Innovación Tecnológica Posgrado, Facultad de Psicología, UNAM.
- (1998). Evaluación de resultados de aprendizaje en escenarios educativos. *Revista Sonorense de Psicología*, 12(2), 57- 67.
- (1994). *La Enseñanza Estratégica. Un modelo de instrucción al servicio del docente universitario*. Reporte interno del Posgrado de la Facultad de Psicología de la UNAM. México: UNAM.
- (1993a). El papel del Desarrollo Cognitivo en la Psicología Escolar mexicana. Trabajo presentado en XXIV Congreso Interamericano de Psicología, 4-9 de julio, Santiago de Chile.
- (1993c). *Herramientas de Estudio. Modelo inductivo de estrategias de Aprendizaje*. Documento interno. Posgrado de la Facultad de Psicología. UNAM: México.
- (1992). *Aplicando la Cognición a materiales Instruccionales*. Tercera Unidad de la Cátedra Especial "José Gómez Robleda", UNAM, México.
- Castañeda, S. & Arce, A. (2001). Construct empirical validation of the general examination for students ending the psychology major. Memorias del VIII Congreso Europeo de Psicología, organizado por la Asociación Europea de Psicología y realizado en Londres, Reino Unido, el 4 de julio de 2001.
- Castañeda, S., Bazán, A., Sánchez, B. A. & Ortega, I. (en revisión). Validez apriorística y empírica de constructos. Modelamiento estructural de porciones extensas de exámenes objetivos a gran escala. Enviado a la *Revista Latina de Pensamiento y Lenguaje* en abril de 2003.
- Castañeda, S., Hernández, J., Smet, M. & Arriola, A. (1995). Evaluating and modeling learning strategies: successes and failures. Memorias de la Conferencia Europea de Investigación en Aprendizaje e Instrucción (p. 333-334). Agosto 26 a 31 de 1995, Nijmegen, Netherlands.
- Castañeda, S. & López, O. M. (1999). Elaboración de un instrumento para la medición de conocimientos y habilidades en estudiantes de psicología. *Revista Intercontinental de Psicología y Educación*, 1, 9-15.
- Castañeda, S. & López, O.M. (1994). Desarrollo cognitivo vía la evaluación y modelamiento de la comprensión de textos instruccionales. Memorias de la Primera Reunión Universitaria sobre la Enseñanza de la Lengua Materna. Memoria de la reunión. Coordinación de Programas Académicos, Secretaría General, UNAM, pp. 96-128.

- Castañeda, S. & López, M. (1994). *How to evaluate and to model learning strategies: convergences, divergences, successes and prospectives*. Memorias de EARLI 4. Aix Provence, France.
- Castañeda, S., López, O. M., Orduña, J. & Pineda, L. (1993c). Un marco de trabajo experimental y neurocomputacional para el estudio de la estructuración del conocimiento a partir de lo leído. *Revista Latina de Pensamiento y Lenguaje*, 1(2), 201-232.
- Castañeda, S. & López, O. M. (1992). Psicología instruccional mexicana. *Revista Intercontinental de Psicología y Educación*, 5(1), 57 - 69.
- (1990). Modelamiento computacional de mecanismos de aprendizaje. De novato a experto. *Revista Mexicana de Psicología*, 7(1-2), 157-171.
- (1989). La psicología del aprendizaje escolar. En S. Castañeda & O. M. López (Eds.). *Antología: La Psicología Cognoscitiva del Aprendizaje*. Aprendiendo a Aprender (p. 25-56). México: UNAM.
- (1988a). *Estructuración de procesos cognoscitivos de bajo y alto nivel en la comprensión y recuerdo de textos de naturaleza científico-instruccional en el bachillerato*. (Serie: Sobre la Universidad N° 9). México: UNAM.
- Castañeda, S. López, O. M., Castro, L. & Heman, A. (1985). Estructuración del conocimiento. Resúmenes del *IV Congreso Mexicano de Psicología*, 5 a 8 de octubre de 1985.
- Castañeda, S. & Martínez, R. (1999). Enseñanza y Aprendizaje Estratégicos. Modelo integral de evaluación e instrucción. *Revista Latina de Pensamiento y Lenguaje*, 4, 251-278.
- Castañeda, S. & Ortega, I. (2003). *Manual de la Guía Integral para Tutores de Educación Superior*. Laboratorio de Evaluación y Fomento del Desarrollo Cognitivo y del Aprendizaje, Facultad de Psicología, UNAM, México.
- Castañeda, S. (2003). *Análisis Cognitivo de Tareas y el Proceso de Responder a Exámenes*. Laboratorio de Desarrollo Cognitivo e Innovación Tecnológica, Posgrado, Facultad de Psicología, UNAM, México.
- Castañeda, S. (2002). *Fundamentos de la medición cognitiva de resultados de aprendizaje*, Laboratorio de Evaluación y Fomento del Desarrollo Cognitivo y el Aprendizaje. Posgrado, Facultad de Psicología, UNAM, México.
- (1998). *Evaluación y fomento del desarrollo intelectual en la enseñanza de ciencias, artes y técnicas*. "Perspectiva internacional en el umbral del siglo XXI". Colección Problemas educativos en México, Facultad de Psicología UNAM, CONACYT y Porrúa. Cap. Modelos de evaluación a gran escala en México, aprendizaje estratégico: un modelo conceptual, instruccional y de evaluación.
- (2000). Un modelo cognitivo para la medición de resultados de aprendizaje. *Revista de Psicología Contemporánea*, 7(7), 92-96.
- Dall'Alba, G. & Edwards, J. (1981). *The scale of cognitive demands: An instrument for analyzing cognitive demand in secondary science*, *Educational Research and development unit*. Royal Melbourne Institute of Technology, Australia.

- Díaz Barriga, Arceo F. & Hernández Rojas, G. (2002). *Estrategias docentes para un aprendizaje significativo: "Una interpretación constructivista"*. México: McGraw-Hill.
- Embreston, S.E. (1985). *Test Design: Developments in psychology and psychometric*. E.U.A.: Academic Press.
- Flores Cruz, A. L. (2004). La calidad de la educación en México. *Observatorio ciudadano de la educación*, volumen IV, N° 102.
- Frederiksen, N., Glaser, R., Lesgold, A. & Shafto, M. G. (1990). *Diagnostic Monitoring of skill and knowledge Acquisition* (Introducción). Hillsdale, New Jersey: LEA.
- Gimeno Sacristán, J. (2002). *Educar y convivir en la cultura global* (caps. 2 y 3). Madrid: Morata.
- Gómez Arceo, B. M. (1990). *Evaluación Criterial: "Una metodología útil para diagnosticar el nivel de aprendizaje de los alumnos"* (cap. 2). Madrid: Nancea, Ediciones Madrid.
- Gronlund, N. (1973). *Medición y evaluación de la enseñanza* (Apéndices). México: Pax.
- Haladyna, T.M. (1994). *Developing and Validating Multiple - Choice test item* (caps. 8, 4 y 10). Hillsdale, New Jersey: Lawrence Erlbaum.
- Messick, S. (1989). Validity [Validez]. En R. Linn (Ed.), *Educational Measurement* (3ª ed., pp. 13-103). New York, NY: Macmillan.
- Pollit, A. & Ahmed, A. (2001). Science or Reading? How Students Think when Answering TMS Questions. Paper presented at the 2001 *International Association for Educational Assessment conference*.
- ____ (1999). A New Model of the Question Answering Process. Paper presented at the *International Association for Educational Assessment Conference*.
- Pollit, A., Marriott, C. & Ahmed, A. (2000). Language, Contextual and Cultural Constraints on Examination Performance. Paper presented at the 2000 *International Association for Educational Assessment Conference*.
- Pollit, A. & Ahmed, A. (2000). Comprehension Failures in Educational Assessment. Paper presented at the 2000 *European Conference on Educational Research*.
- Sampieri, H. & Collado, F. (2000). *Metodología de la investigación* (cap. 10). México D.F., McGraw-Hill.
- Santrock, J. W. (2002). *Psicología de la educación* (caps. 13 y 14). México: McGraw-Hill.
- Siegel, S. (1979). *Estadística no paramétrica "aplicada a las ciencias de la conducta"*. México: Trillas.

Tabla 1
Porcentajes de apreciación de jueces en definiciones organizados por disciplina, nivel académico y experiencia de los jueces

Indicadores	Psicología		Pedagogía		Licenciatura		Posgrado		(-)Exp.		(+)Exp.	
	D	A	D	A	D	A	D	A	D	A	D	A
Tipo formato	0	100	0	100	0	100	0	100	0	100	0	100
Simple	17	83	0	100	11	89	11	89	22	78	0	100
Ordenamiento	8	92	17	83	11	89	11	89	22	78	0	100
Apareamiento	0	100	17	83	11	89	0	100	11	89	0	100
FVM	17	83	0	100	11	89	11	89	11	89	11	89
Canevá	8	92	17	83	22	78	0	100	11	89	11	89
		92		92		89		94		87		96
Operación cognitiva	17	83	0	100	11	89	11	89	11	89	11	89
Comp. y org.	8	92	17	83	11	89	11	89	22	78	0	100
Aplicación	8	92	0	100	0	100	11	89	11	89	0	100
Resolución	0	100	17	83	11	89	0	100	11	89	0	100
		92		92		92		92		86		97
Contexto	0	100	0	100	0	100	0	100	0	100	0	100
E-R	0	100	0	100	0	100	0	100	0	100	0	100
R-E	0	100	0	100	0	100	0	100	0	100	0	100
C E-R	0	100	0	100	0	100	0	100	0	100	0	100
CR-E	0	100	0	100	0	100	0	100	0	100	0	100
		100		100		100		100		100		100
C. de conocimiento	0	100	17	83	11	89	0	100	11	89	0	100
Teórico	0	100	17	83	0	100	11	89	0	100	11	89
Técnico	0	100	0	100	0	100	0	100	0	100	0	100
Combinado	0	100	0	100	0	100	0	100	0	100	0	100
		100		92		97		97		97		97
Redacción	0	100	0	100	0	100	0	100	0	100	0	100
Sobre la base	0	100	17	83	11	89	0	100	11	89	0	100
Sobre las opciones	0	100	17	83	11	89	0	100	11	89	0	100
		100		89		89		100		89		100
Uta	8	92	0	100	0	100	11	89	11	89	0	100

Tabla 2

Porcentajes de apreciación de jueces en asignación de puntajes organizados por variables de capital cultural de los jueces, en las dimensiones *redacción del reactivo* y *uso técnico apropiado de la escala de valoración*

Indicadores	Psicología		Pedagogía		Licenciatura		Posgrado		(-)Exp.		(+)Exp.	
	D	A	D	A	D	A	D	A	D	A	D	A
Redacción del reactivo												
A1	8	92	33	67	11	89	22	78	11	89	22	78
A2	33	67	33	67	33	67	33	67	56	44	11	89
A3	17	83	17	83	11	89	22	78	33	67	0	100
A4	0	100	33	67	22	78	0	100	11	89	11	89
A5	8	92	33	67	33	67	0	100	22	78	11	89
A6	8	92	0	100	11	89	0	100	11	89	0	100
A7	0	100	33	67	22	78	0	100	0	100	22	78
A8	8	92	0	100	0	100	11	89	11	89	0	100
A9	8	92	17	83	11	89	11	89	11	89	11	89
B1	25	75	0	100	0	100	33	67	33	67	0	100
B2	8	92	33	67	22	78	11	89	22	78	11	89
B3	8	92	33	67	22	78	0	100	22	78	11	89
B4	0	100	17	83	11	89	0	100	0	100	11	89
B5	0	100	17	83	11	89	0	100	0	100	11	89
C1	0	100	33	67	22	78	0	100	11	89	11	89
C2	8	92	17	83	11	89	11	89	11	89	11	89
C3	0	100	33	67	22	78	0	100	11	89	11	89
C4	0	100	33	67	22	78	0	100	11	89	11	89
C5	8	92	0	100	0	100	11	89	11	89	0	100
C6	17	83	0	100	0	100	22	78	22	78	0	100
C7	8	92	17	83	11	89	11	89	11	89	0	100
C8	25	75	17	83	22	78	22	78	33	67	11	89
	9	91	20	80	15	85	10	90	17	83	9	91
Uso técnico apropiado												
U1	17	83	0	100	11	89	11	89	22	78	0	100
U2	25	75	17	83	22	78	22	78	44	56	0	100
U3	17	83	17	83	11	89	22	78	11	89	22	78
U4	17	83	0	100	0	100	22	78	22	78	0	100
U5	17	83	17	83	11	89	22	78	11	89	22	78
U6	0	100	33	67	22	78	0	100	11	89	11	89
U7	8	92	17	83	11	89	11	89	11	89	11	89
U8	0	100	0	100	0	100	0	100	0	100	0	100
U9	17	83	0	100	11	89	11	89	11	89	11	89
U10	25	75	0	100	0	100	11	89	11	89	0	100
U11	17	83	17	83	11	89	22	78	33	67	0	100
U12	25	75	17	83	11	89	33	67	33	67	11	89
	15	85	11	89	10	90	16	84	18	82	7	93

Tabla 3

Porcentajes de apreciación de jueces en asignación de puntajes organizados por variables de capital cultural de los jueces, en las dimensiones tipo de formato, operación cognitiva, tipo de contexto y campo de conocimiento

Indicadores	Psicología		Pedagogía		Licenciatura		Posgrado		(-)Exp.		(+)Exp.	
	D	A	D	A	D	A	D	A	D	A	D	A
Tipo de formato												
Simple	25	75	17	83	0	100	44	56	33	67	11	89
Ordenamiento	25	75	17	83	0	100	44	56	33	67	11	89
Apareamiento	25	75	17	83	0	100	44	56	33	67	11	89
FVM	50	50	17	83	22	78	44	56	44	56	22	78
Canevá	50	50	66	33	56	44	56	44	56	44	56	44
	35	65	27	73	16	84	46	54	40	60	22	78
Operación cognitiva												
Identificación	8	92	17	83	22	78	0	100	0	100	22	78
Ord. y clasif.	8	92	33	67	33	67	0	100	11	89	22	78
Jerarquización	8	92	33	67	33	67	0	100	11	89	22	78
Conc. y Princ.	25	75	50	50	44	56	22	78	33	67	33	67
TTPR	8	92	50	50	33	67	11	89	11	89	33	67
Modelos mentales	17	83	50	50	33	67	22	78	22	78	33	67
Ident. de errores	8	92	50	50	33	67	11	89	11	89	33	67
Acción con guía	17	83	50	50	33	67	22	78	22	78	33	67
Acción sin guía	8	92	50	50	33	67	11	89	11	89	33	67
	12	88	43	57	33	67	11	89	15	85	29	71
Contexto												
E-R	8	92	67	33	22	78	33	67	33	67	22	78
R-E	8	92	50	50	11	89	33	67	33	67	11	89
C E-R	17	83	67	33	22	78	44	56	44	56	22	78
CR-E	17	83	83	17	33	67	44	56	44	56	33	67
	13	87	67	33	22	78	39	61	39	61	22	78
Campo de conocimiento												
Teórico	8	92	0	100	0	100	11	89	11	89	0	100
Técnico	8	92	0	100	0	100	11	89	11	89	0	100
Combinado	8	92	0	100	0	100	11	89	11	89	0	100
	8	92	0	100	0	100	11	89	11	89	0	100