

**Da Dalt de Mangione, Elizabeth ; Difabio de Anglat, Hilda**

*Evaluación de la competencia crítica a través del test Watson-Glaser. Exploración de sus cualidades psicométricas*

*Assessment of critical competence through the Watson-Glaser test. Exploration of the psychometric qualities*

Revista de Psicología Vol. 3 N° 6, 2007

Este documento está disponible en la Biblioteca Digital de la Universidad Católica Argentina, repositorio institucional desarrollado por la Biblioteca Central "San Benito Abad". Su objetivo es difundir y preservar la producción intelectual de la Institución.

La Biblioteca posee la autorización del autor para su divulgación en línea.

Cómo citar el documento:

Da Dalt de Mangione, E., Difabio de Anglat, H. (2007). Evaluación de la competencia crítica a través del test de Watson-Glaser : exploración de sus cualidades psicométricas [en línea]. *Revista de Psicología*, 3(6).  
Disponible en: <http://bibliotecadigital.uca.edu.ar/repositorio/revistas/evaluacion-competencia-critica-test-watson-glaser.pdf> [Fecha de consulta:.....]

## **Evaluación de la competencia crítica a través del test de Watson-Glaser. Exploración de sus cualidades psicométricas.**

Assessment of critical competence through the Watson-Glaser test.  
Exploration of the psychometric qualities.

Elizabeth Da Dalt de Mangione  
CONICET- UNCuyo. UCA

Hilda Difabio de Anglat  
Centro de Investigaciones Cuyo -CONICET

### **Resumen**

En una investigación piloto, se estudia la confiabilidad, la validez de constructo, la potencialidad discriminativa y el nivel de dificultad del Watson-Glaser Critical Thinking Appraisal, diseñado por Goodwin Watson y Robert Glaser (1980) para evaluar la capacidad de pensamiento crítico. El instrumento se aplicó a 102 alumnos de 16 años de edad en promedio. Evidencia muy buena confiabilidad ( $\alpha$  de Cronbach = 0,82), validez de constructo, homogeneidad en el 65% de los ítems, potencialidad discriminativa de nivel aceptable, bueno o muy bueno en el 71%, un adecuado nivel medio de dificultad (0,63).

### **Abstract**

A pilot research is set forth in which trustworthiness, construct validity, discriminative potentiality and difficulty level are studied in compliance with the Watson-Glaser Critical Thinking Appraisal, designed by Goodwin Watson and Robert Glaser (1980) to assess the ability of critical thinking. This test was conducted on 102 students, within an average age of 16 years. It showed very good trustworthiness (Cronbach's  $\alpha=0.82$ ), construct validity, homogeneity in 65% of items, discriminative potentiality in an acceptable level, 71% good or very good, and an adequate medium difficulty level (0,63).

Palabras Clave: Fiabilidad, Validez, Razonamiento, Pensamiento Crítico, Apreciación del Test Watson-Glaser.

Key Words: Test Reliability, Test Validity, Reasoning, Critical Thinking, Watson-Glaser Critical Thinking Appraisal

### **Introducción**

Gran impacto y desconcierto produce en la sociedad en general, y en docentes y padres, en particular, el alto índice de conductas de riesgo en los adolescentes y jóvenes, la agresividad atrozmente irracional, la desmotivación e indiferencia para construir un proyecto de vida personal, académica, laboral –más allá de lo esperable y razonable conforme a la etapa evolutiva que atraviesan. Por otra parte, es característico de la adolescencia el desmoronamiento de los valores que fueron promovidos

Correspondencia: Elizabeth Da Dalt de Mangione

Conicet

edalt@lab.cricyt.edu.ar

heterónomamente en la niñez –mediante la familia y la escuela– así como la reconstrucción de una escala valorativa desde la interioridad de la persona.

Sin embargo, es evidente que las condiciones socio-culturales contemporáneas (el predominio de la imagen, la fantasía, los sentidos; el profuso y creciente caudal de información de la llamada sociedad del conocimiento; el culto a lo efímero, la moda, el consumo; la búsqueda de evasión; la erosión del principio de realidad) dificultan dicha reconstrucción y exigen habilidades superiores de pensamiento de modo más decisivo que en cualquier época histórica anterior. Los nuevos contextos plantean nuevos desafíos y puesto que “la educación encierra un tesoro...se ve obligada a proporcionar las cartas náuticas de un mundo complejo y en perpetua agitación y, al mismo tiempo, la brújula para poder navegar por él” (Delors, 1996: 95). Es decir, a fin de contrarrestar las condiciones negativas señaladas y para responder a las demandas de una sociedad donde el cambio acontece con una velocidad vertiginosa, debemos aprender y enseñar a pensar y vivir crítica y reflexivamente, a ser capaces de discernir lo permanente, lo nuclear, lo valioso, de lo cambiante, accesorio y superficial. Ello exige del educador un compromiso renovado en ayudar al adolescente a cruzar el río para llegar a la otra orilla.

Por tanto, el problema último por investigar es cómo promover en el adolescente la competencia crítica reflexiva y las actitudes relacionadas con la misma –y no sólo las habilidades o procesos intelectuales, a fin de evitar el formalismo que entraña el llamado “movimiento del pensamiento crítico”<sup>i</sup>- a través de una incidencia sistemática mediante un contenido disciplinar significativo y valioso<sup>ii</sup>. Ello implica la necesidad de seleccionar un instrumento adecuado, válido y confiable, porque en esta área -como en otras metas educativas complejas- la carencia de métodos apropiados de evaluación constriñe el currículum: “... lo que es fácilmente evaluado se convierte en más importante” (Nagy, 1994: 164; el destacado es nuestro)<sup>iii</sup>.

Watson y Glaser, pioneros en la conceptualización y evaluación del pensamiento crítico, en su última formulación (1994, citados por Loo y Thorpe, 1999: 995) proporcionan la siguiente definición descriptiva:

1) Actitudes de indagación que implican la habilidad para reconocer la existencia de problemas y la aceptación de la necesidad general de evidencia en apoyo de lo que se asevera como verdadero; 2) conocimiento de la naturaleza de las inferencias válidas, abstracciones y generalizaciones en las que se determinan lógicamente la fuerza y atingencia de diferentes clases de evidencia; 3) habilidades para emplear y aplicar dichas actitudes y conocimiento.”

Nosotros lo hemos definido como el “pensamiento reflexivo que, mediante el análisis cuidadoso de los argumentos, busca evidencia válida y conclusiones fundamentadas” (Difabio de Anglat, 2005a, p. 181).

Desde sus componentes, la competencia crítica alude al pensamiento reflexivo que involucra una duda, perplejidad o dificultad mental que origina la indagación para: 1) interpretar o clarificar el significado de conceptos, juicios, analogías, reglas, símbolos; 2) identificar las diversas posturas que articula de modo conciso y coherente en argumentos; 3) analizar el/los argumentos para identificar el postulado o conclusión y las razones de apoyo o premisas y 4) Evaluar la validez y fuerza de las inferencias y deducciones.

Como señaláramos, una cuestión de importancia en la investigación del pensamiento crítico es la evaluación de sus dimensiones. En lengua inglesa (únicos instrumentos estandarizados existentes, hasta donde hemos podido indagar), entre los llamados tests generales o comprensivos<sup>iv</sup>, para nivel medio y universidad, se cuenta con los siguientes instrumentos (Ennis, 1991)<sup>v</sup>:

- Watson-Glaser Critical Thinking Appraisal de Goodwin Watson y Robert Glaser (1980);
- Cornell Critical Thinking Test, Level X y Level Z de Robert Ennis, Jason Millman y Thomas Tomko (1985);
- The Ennis-Weir Critical Thinking Essay Test de Robert Ennis y Eric Weir (1985);
- The California Critical Thinking Skills Test - College Level de Peter Facione y Noreen Facione (1990);
- New Jersey Test of Reasoning Skills de Virginia Shipman (1983).

En nuestro caso hemos optado por el test de Watson y Glaser (1980) porque -aunque prácticamente desconocido en nuestro medio- es el instrumento estandarizado de mayor uso en la investigación empírica en los 90' (por citar sólo algunos trabajos: Davou, Taylor y Worrall, 1991; Norris, 1992) y recientemente, en su forma original de 80 ítems o en una versión abreviada de 40 (Clifford, Boufal y Kurtz, 2004; Karma y Ghida, 2007; Loo y Thorpe, 2005; Moutafi, Furnham y Crump, 2003; Williams y Stockdale, 2003, entre varios otros). Los autores son coincidentes en que las cualidades psicométricas probadas del test lo recomiendan sobre otras opciones: "Para este estudio, se selecciona el Watson-Glaser Critical Thinking Appraisal a partir de la evaluación empírica de tres instrumentos de pensamiento crítico [se refiere a: Cornell Critical Thinking Test y California Critical Thinking Skills Test] ... Su consistencia interna, estimada por la correlación entre mitades, evidencia un promedio de 0.85 en las cinco muestras, con un rango desde 0.78 a 0.87" (Clifford, Boufal y Kurtz, 2004, p.171).

Este artículo se propone aportar algunos datos sobre aspectos también psicométricos y estructurales en su aplicación en español.

Es un estudio de naturaleza exploratoria, cuyos resultados pudieran servir para la realización de trabajos posteriores de carácter más específico destinados a ofrecer una herramienta útil para conocer y evaluar de manera más ágil y fiable la competencia crítica de estudiantes de nivel medio.

### **Descripción del instrumento**

En sus dos formas alternativas -A y B-, el Watson-Glaser Critical Thinking Appraisal original (en adelante: WGCTA) consta de 80 ítems, distribuidos en cinco dimensiones (Difabio de Anglat, 2005b: 229)<sup>vi</sup>:

- 1) Inferencia, definida como la conclusión que puede extraerse de hechos observados o supuestos, evalúa la discriminación de los grados de validez de inferencias inmediatas con cinco opciones para cada ítem (válido, probablemente válido, insuficiencia de datos, probablemente inválido e inválido).
- 2) Reconocimiento de supuestos o enunciados implícitos en las aseveraciones proporcionadas.
- 3) Deducción, determinación de la atingencia lógica de conclusiones a partir de las premisas dadas.
- 4) Interpretación, ponderación de la evidencia para juzgar si las generalizaciones propuestas se siguen de los datos más allá de una duda razonable; entonces, se distingue del subtest anterior en el cual la tarea consiste en determinar si las conclusiones se siguen necesariamente de la información proporcionada.
- 5) Evaluación de argumentos, distinción de los argumentos en "fuertes" o "débiles" en función de su relevancia para el asunto en discusión.

Empleamos la forma A en la traducción de Mónica Descalzzi (Ciafic -Centro de Investigaciones en Antropología Filosófica y Cultural-, Buenos Aires) y la adaptación de Chalupa-Vázquez (2006) que reduce el test a 66 ítems.

### **Muestra**

Está compuesta por 102 alumnos de primer y segundo año del polimodal (16 años de edad en promedio) de tres colegios de gestión privada de la provincia de Mendoza. Las tres escuelas son urbanas y los alumnos pertenecen a familias de nivel socio-cultural-económico medio y medio alto.

El estudio se lleva a cabo en el marco de la cátedra “Métodos y Diseños de Investigación” -a cargo de la Dra. Da Dalt de Mangione- de la Facultad de Humanidades y Ciencias de la Educación, U.C.A sede Mendoza, en el que las alumnas -docentes que realizan la licenciatura en Ciencias de la Educación- aplican y evalúan el instrumento como actividad práctica para abordar la determinación de las cualidades psicométricas de un instrumento.

Dado que es un Test de habilidad más que de velocidad, se administra -en forma colectiva, en dos partes con un intervalo de dos semanas- sin límite de tiempo. Se observan diferencias en el tiempo empleado por los alumnos para completar cada parte del instrumento: en las escuelas A y C, los alumnos emplean 40 minutos como máximo; en la B, 90 minutos (aunque en las tres instituciones se detectan diferencias importantes en esta variable de sujeto a sujeto).<sup>vii</sup>

La muestra es estratégica u ocasional, seleccionada en función del fin principal de la investigación: probar un instrumento diseñado para otra población y explorar sus cualidades psicométricas específicas para su uso particular (la evaluación de la capacidad de pensamiento crítico en alumnos de nivel medio).

### **Análisis de los resultados**

Se efectúan los siguientes análisis: descripción de la variable “pensamiento crítico” y su asociación con institución educativa, edad, sexo y rendimiento escolar, determinación de la confiabilidad, la correlación subescalas-puntuación total y la validez de constructo, estudio de la potencialidad discriminativa y del nivel de dificultad de los ítems.

#### **1) Descripción de la variable “pensamiento crítico” y su asociación con institución educativa, edad, sexo y rendimiento escolar**

Sobre un total posible de 66 puntos, la media es 41,3; su desvío estándar, 9,4; la mediana, 41,33; el puntaje mínimo, 17 y el máximo, 61.

Sintetizando la información en tres categorías, se obtienen los siguientes porcentajes:

Gráfico 1

## Gráfico 2

El bloque rectangular representa el 50% de los casos centrales de la distribución (desviación semiintercuartil o dispersión intercuartílica); la raya representa la mediana y los signos exteriores, los *outliers* (datos aberrantes). Es evidente la superioridad en competencia crítica de la escuela A y la más baja en la escuela B, según lo manifiesta el valor de la mediana (escuela A, 50 puntos; escuela B, 33,5; escuela C, 41) y la ubicación del bloque (límites inferior y superior del bloque en la escuela A, 46-53; en la escuela B, 31-36; en la escuela C, 36-45); diferencias que son estadísticamente significativas (Chi cuadrado = 39,65, con 2 grados de libertad,  $p = 0,00$ ).

## Gráfico 3

En relación con la cristalización de las capacidades intelectuales hacia el término de la enseñanza media, se esperaba una superioridad del grupo de 17 años en comparación con el de 16; sin embargo, si bien se advierte una leve superioridad en aquel, esta diferencia no es estadísticamente significativa (Chi cuadrado = 0,88, con un grado de libertad,  $p = 0,348$ ).

La diferencia entre los sexos tampoco es estadísticamente significativa (Chi cuadrado = 2,78, con un grado de libertad,  $p = 0,09$ ), resultado en la línea de investigaciones con estudiantes universitarios en otros contextos -por ejemplo, en el Líbano (Karma y Ghida, 2007)- y contradictoria con estudios tempranos (entre otros, Simon y Ward, 1974). En su representación gráfica:

## Gráfico 4

En cambio, se halla una diferencia estadísticamente significativa (Chi cuadrado = 14,895, con dos grados de libertad,  $p = 0,006$ ) en función de tres niveles de rendimiento, que se establecen para toda la muestra (sin distinción del colegio al que asiste el alumno) con la práctica de sumar y restar media desviación estándar al promedio, método que tomamos del modo de determinar un estilo intermedio en el constructo “dependencia-independencia respecto del campo (Meng y Del, 1991, p. 184):

## Gráfico 5

Se observa claramente el descenso en la puntuación total en el test en función del nivel de rendimiento, según lo manifiestan los valores de la mediana (Alto: 47 puntos; Intermedio: 40; Bajo: 36) y la ubicación de los bloques (límites inferior y superior del bloque en nivel Alto, 41-52; Intermedio, 35-45,5; Bajo, 32-46,5).

## 2) Confiabilidad

Para calcular la confiabilidad, se aplica el alpha de Cronbach en tanto requiere una sola administración del instrumento.

Se obtiene un alpha de 0,82 (ningún ítem sin variabilidad) para el test como un todo. Luego, según este estadístico (el más usual en la investigación cuantitativa contemporánea, según lo evidencian las publicaciones periódicas), se considera que el instrumento evidencia muy buena confiabilidad ya que *una* aplicación expresa casi el 68 % de la información que arrojaría administrado nuevamente al mismo grupo.

Por subescala, los coeficientes son más bajos:

- 1) Inferencia, alpha de 0,59, que se eleva a 0,66 si se eliminan los cinco ítems defectuosos, con una correlación ítem-puntuación total negativa o cercana a 0 (cfr. Anexo II, Tabla).
- 2) Reconocimiento de supuestos, alpha de 0,53; sin los dos ítems defectuosos, 0,61.
- 3) Deducción, alpha de 0,34; de 0,48 prescindiendo de los cuatro ítems defectuosos.
- 4) Interpretación, alpha de 0,45; sin el ítem defectuoso, 0,51.
- 5) Evaluación de argumentos, 0,68; aumenta a 0,70 si se anula el ítem defectuoso.

Luego, tres subescalas (Inferencia, Reconocimiento de supuestos y Evaluación de argumentos) satisfacen la confiabilidad mínima aceptable.

## 3) Correlaciones entre las subescalas y la puntuación total

El análisis de los resultados por dimensión de la variable (Tabla 1) evidencia que los subtests correlacionan con la puntuación total desde 0,54 a 0,73, bastante cercano al rango entre 0,62 y 0,73 que informan sus autores. Por otra parte, correlacionan entre sí, a excepción de deducción con evaluación de argumentos. Los coeficientes más altos se verifican entre inferencia y evaluación de argumentos y entre aquella y reconocimiento de supuestos; también son estos tres los que correlacionan mejor con el puntaje total en el test.

Tabla 1

#### 4) Validez de constructo

Para determinar la validez de construcción, resulta apropiado un análisis factorial exploratorio<sup>viii</sup> ya que la medida de adecuación muestral KMO es de 0,776 (según sus autores se requiere una puntuación mínima de 0,50 para poder realizar el análisis) y la prueba de esfericidad de Bartlett arroja un valor para Chi-cuadrado de 128,111 ( $p = 0,000$ ), lo que permite rechazar la hipótesis de que la matriz muestral provenga de una población en la que las variables no estén correlacionadas.

Con el método de extracción de factores basado en componentes principales, se generaron cinco componentes con autovalores superiores a 0,35, que explican el 52,26 % de la varianza, con fuerte peso del factor N° 1 en dicha explicación. Las subescalas saturan significativamente en las cinco dimensiones originales del test:

Tabla 2

#### 5) Homogeneidad de los ítems

Para determinarla, se emplea la correlación entre cada ítem y el puntaje total corregido o método de la consistencia interna, así llamado porque se evalúa si cada elemento se diferencia en el mismo sentido que el instrumento entero (Anastasi, 1973: 122). [Cfr. Anexo II, Tabla]. De los 66 ítems, 43 ítems -en letra negrita en dicha Tabla- (el 65 % del test) presentan una correlación significativa a nivel  $p < 0,05$ .

#### 6) Potencialidad discriminativa de los ítems

Se establece la discriminación de los ítems mediante el procedimiento de ordenar los resultados en la muestra total en forma decreciente; computar el porcentaje de respuestas correctas por ítem en el 27% superior y en el 27% inferior, porcentaje que permite formar grupos extremos tan grandes como es posible y al mismo tiempo, tan diferentes como es posible (Vianna, 1983)<sup>ix</sup>; y hallar el “índice de discriminación” (*ÍD*), resultado de la resta del número de aciertos entre el grupo superior y el inferior, dividido el número de sujetos que constituyen el 27% de la muestra -en este caso, 28 alumnos- (Cfr. Anexo II). El 50% del *test* (33 ítems) alcanza potencialidad discriminativa buena o muy buena.

Combinando ambos criterios -homogeneidad y discriminación-, 47 ítems (el 71% del test) resultan homogéneos y/o discriminativos; los ítems se distribuyen de la siguiente manera:

Tabla 3

Si se recalcula el alpha de Cronbach desestimando los 21 ítems defectuosos, porque no satisfacen -al menos- alguna de las condiciones (8 de 20 en Inferencia -Nº 1, 2, 3, 6, 9, 10, 11 y 20-; 2 de 10 en Reconocimiento de Supuestos -Nº 21 y 26-; 4 de 10 en Deducción -Nº 33, 34, 37 y 41-; 5 de 12 en Interpretación -Nº 43, 44, 47, 51 y 54-; 2 de 12 en Evaluación de Argumentos -Nº 56 y 60-), el coeficiente se eleva a 0,87, lo que significa que el instrumento en una aplicación muestra más del 75% de lo que arrojaría administrado nuevamente al mismo grupo.

### 7) Nivel de dificultad de los ítems

La dificultad de un ítem se determina en función de la proporción de personas que lo responden correctamente. Es una relación inversa: entre mayor sea esta proporción, menor será su dificultad. Para calcular la dificultad de un ítem, simplemente se divide el número de alumnos que contestó correctamente el ítem entre el número total de la muestra (Cfr. Anexo II).

El nivel medio de dificultad del test es 0,63; luego, satisface el requerimiento de la teoría estadística porque dicho nivel medio debe oscilar entre 0,5 y 0,6. Por otra parte, se recomienda que en los instrumentos para el ámbito educativo se empleen ítems cuyo intervalo de dificultad esté entre 0,20 y 0,80. “Los ítems con dificultad fuera del intervalo apuntado deben emplearse en un número limitado” (Vianna, op. cit.: 231). Se aconseja (e.g., Backhoff, Larrazolo y Rosas, 2000) un 5% de ítems muy fáciles (ND: de 0,87 a 1), 20% fáciles (ND: de 0,74 a 0,86), 50% con una dificultad media o moderados (ND: de 0,53 a 0,73), 20% difíciles (ND: de 0,33 a 0,52) y 5% muy difíciles (ND: de 0 a 0,32). En líneas generales, el test de Watson-Glaser satisface estos requisitos porque los resultados son: 1 ítem, muy fácil; 29% fáciles; 44% moderados; 21% difíciles; 4% muy difíciles.

## Discusiones

Una meta educativa de importancia requiere de un procedimiento apropiado de evaluación; en este sentido, según Oakland y Eu (1993), el interés creciente por el desarrollo de instrumentos con cualidades psicométricas comprobadas, se funda en la preocupación internacional por la calidad educativa en la cual la evaluación juega un importante papel.

De allí que el juicio acerca del valor de un instrumento para un propósito particular sólo debería emitirse una vez determinadas, de manera objetiva y tan exacta como fuera posible, dichas cualidades psicométricas, porque este conocimiento es necesario para que tanto el instrumento como los datos obtenidos con el mismo puedan usarse significativamente.

El WGCTA manifiesta una confiabilidad de 0,82 según el *alpha* de Cronbach, coeficiente que la teoría psicométrica juzga “considerable”.

Manifiesta adecuada homogeneidad de los ítems, evaluada a través del índice de correlación ítem-instrumento, en el 65% del instrumento: 43 enunciados (de 66) parecen diferenciarse en el mismo sentido que el test en su totalidad.

En tercer lugar, el 71% tiene una potencialidad discriminativa de nivel aceptable, bueno o muy bueno, y el nivel medio de dificultad es adecuado (0,63).

Por otra parte, si se desestiman los 21 ítems defectuosos (porque no satisfacen, al menos, alguna de ambas condiciones), el alpha de Cronbach se eleva a 0,87.

Parece, entonces, un instrumento potente en tanto, por un lado, evalúa una competencia unitaria según manifiesta la validación de constructo y, por otro, resulta sensible a distintos niveles de logro en la competencia crítica de alumnos de 1° y 2° años de nivel polimodal.

Finalmente, resulta de ágil aplicación (genera empatía en el alumno) y calificación, ambas cualidades de importancia en el ámbito escolar.

En suma, en nuestros resultados, el Watson-Glaser Critical Thinking Appraisal se inscribe en la línea de la llamada evaluación auténtica (Bateson, 1994), que se caracteriza por tareas que el alumno juzga valiosas y significativas: son en sí mismas instancias reales de la capacidad bajo examen.

### **Bibliografía**

- ANASTASI, A. (1947/1968). *Psychological Testing*. New York: Prentice-Hall. (Trad. Riesco Hernández, C.). Tests psicológicos. Madrid: Aguilar, 1958.
- BACKHOFF, E., LARRAZOLO, N. Y ROSAS, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). Revista Electrónica de Investigación Educativa, 2(1). (<http://redie.ens.uabc.mx/vol2no1>, 08/06/2004).
- BATESON, D. (1994). Psychometric and Philosophic Problems in “Authentic” Assessment: Performance Tasks and Portfolios. The Alberta Journal of Educational Research, XL(2), 233-245.
- CLIFFORD, J., BOUFAL, M. Y KURTZ, J. (2004). Personality Traits and Critical Thinking Skills in College Students: Empirical Tests of a Two-Factor Theory. Assessment, 11(2), 169-176.
- CHALUPA, M. (2006). El pensamiento crítico en la enseñanza de la Filosofía en el nivel medio. Tesis de Licenciatura no publicada, La Plata, Univ. Católica de La Plata, dirigida por la Dra. Stella Maris Vázquez.
- DAVOU, B., TAYLOR, F.Y WORRALL, N. (1991). The interplay of knowledge and abilities in the processing of text. The British Journal of Educational Psychology, 61(3), 310-318.
- DELORS, J. (1996). L’Education, un trésor est caché dedans. París: UNESCO/Odile Jacob. (Trad. Cast. No figuran datos) Informe Delors. La educación encierra un tesoro. Madrid: Unesco-Santillana.
- DIFABIO DE ANGLAT, H. (2005a). Competencias para la comprensión de texto y el pensamiento crítico en el nivel medio y universitario. Tesis Doctoral no publicada, Mendoza, Univ. Nacional de Cuyo.
- (2005b). El critical thinking movement y la educación intelectual. Estudios sobre Educación, Pamplona, España, 9, 167-187.
- ENNIS, R.(1991). Critical Thinking Tests. En: A. Costa. Developing Minds (368-69). Vol. 1, Alexandria, Virginia: ASCD (Association for Supervision and Curriculum Development).
- ENNIS, R., MILLMAN, J. Y TOMKO, T. (1985). Cornell Critical Thinking Tests Level X y Level Z. Manual (3a. ed.). Pacific Grove: Midwest Publications.

- ENNIS, R. Y WEIR, E. (1985). The Ennis-Weir Critical Thinking Essay Test. Test, Manual, Criteria, Scoring Sheet. An instrument for teaching and testing. Pacific Grove: Midwest Publications.
- FACIONE, P. Y FACIONE, N. (1990). Test Manual: The California Critical Thinking Skills Test (CCTST). Form A and Form B. Millbrae: The California Academic Press.
- KARMA, H. Y GHIDA, M. (2007). Validating the Watson Glaser Critical Thinking Appraisal. Higher Education, 54(3), 361-383.
- LOO, R. Y THORPE, K. (1999). A Psychometric Investigation of Scores on the Watson-Glaser Critical Thinking Appraisal New Form S. Educational and Psychological Measurement, 59, 995-1003.
- (2005). Relationships Between Critical Thinking and Attitudes Toward Women's Roles in Society. The Journal of Psychology, 139(1), 47-55.
- MENG, K. Y DEL, P. (1991). Field Dependence and Contextual Organizers. Journal of Educational Research, 84(3), 183-189.
- MOUTAFI, J., FURNHAM, A. Y CRUMP, J. (2003). Demographic and Personality Predictors of Intelligence. European Journal of Personality, 17, 79-94.
- NAGY, PHILIP (1994). Cognitive Assessment of Solutions to an Ill-structured Problem Using Forced Classification Analysis. The Alberta Journal of Educational Research, XL(2), 163-175.
- NORRIS, S. (1985). Synthesis of Research on Critical Thinking. Educational Leadership, 43(3), 40-45.
- NORRIS, S. Y ENNIS, R. (1989). Evaluating Critical Thinking. Pacific Grove, CA: Midwest Publications.
- OAKLAND, T. Y HU, S. (1993). The Top 10 Tests used with children and youth worldwide. Bulletin of the International Test Commission, 19(1), 99-120.
- SIMON, A. Y WARD, L. (1974). The Performance on the Watson-Glaser Critical Thinking Appraisal of University Students Classified According to Sex, Type of Course Pursued, and Personality Score Category. Educational and Psychological Measurement, 34, 957-960.
- SHIPMAN, V. (1983). New Jersey Test of Reasoning Skills. Form B. Upper Montclair, NJ: Institute for the Advancement of Philosophy for Children, Montclair State College.
- VIANNA, H. (1982). Testes em Educação. São Pablo: Editora Ibrasa. (Trad. Cast. No figura datos) Los tests en educación. Pamplona, España: EUNSA, 1983.
- WATSON, G. Y GLASER, E. (1980). Critical Thinking Appraisal. San Antonio: Harcourt Brace Jovanovich.
- WILLIAMS, R. Y STOCKDALE, S. (2003). High-performing students with low critical thinking skills. The Journal of General Education, 52(3), 200-226.

## Anexo I

### Inferencia

#### Instrucciones

Una inferencia es una conclusión que puede extraerse de ciertos hechos observados o supuestos. Por ej., si las luces están prendidas en una casa y se oye música proveniente de la misma, podríamos inferir que alguien está en la casa. Pero esta inferencia puede o no ser correcta. Tal vez sus ocupantes no apagaron las luces ni la radio al irse.

En este test cada ejercicio debe comenzar con una afirmación que debe considerar verdadera. Después de cada afirmación encontrará varias inferencias posibles, es decir conclusiones que podrían extraerse de los hechos afirmados. Examine cada inferencia por separado.

Para cada una encontrará un casillero en la hoja de respuestas, designados con la letra V, PV, ID, F; haga una marca en el casillero adecuado del siguiente modo:

V Si piensa que la inferencia es indudablemente Verdadera o correcta; que se sigue más allá de una duda razonable de los hechos dados.

PV Si, a la luz de los hechos dados, piensa que la inferencia es Probablemente Verdadera; que es más probable que sea correcta que incorrecta.

ID Si decide que hay Insuficiencia de Datos; que usted no puede saber, partiendo de los hechos dados, si es probable que la inferencia sea correcta o incorrecta; si los hechos no proporcionan una base para juzgar en uno u otro sentido.

F Si piensa que una inferencia es indudablemente Falsa; que es incorrecta, ya sea porque malinterpreta los hechos dados, o porque contradice los hechos o inferencia que se siguen necesariamente de los hechos.

### Ejercicio

Hace algún tiempo una multitud se congregó en Uruguay para oír hablar al nuevo presidente de la Cámara de Comercio local. El presidente dijo: “No estoy pidiendo sino exigiendo, que los sindicatos acepten completamente la responsabilidad que les corresponde por el mejoramiento cívico y el bienestar de la comunidad. No estoy pidiendo sino exigiendo, que se unan a la Cámara de Comercio”. Los miembros de los Sindicatos Centrales que estaban presentes aplaudieron con entusiasmo. Tres meses más tarde todos los sindicatos de Uruguay estuvieron representados en la Cámara de Comercio. Estos representantes trabajaron con representantes de otros grupos en Comités, expresaron abiertamente sus opiniones, participaron activamente en los proyectos de mejoramiento cívico, y ayudaron a la Cámara a lograr las metas establecidas en relación con dichos proyectos.

11) Tanto los representantes de los sindicatos como los otros miembros de los Comités llegaron a un mejor conocimiento de sus recíprocos puntos de vista a través de sus contactos en la Cámara de Comercio.

12) La participación de los sindicatos en la Cámara de Comercio de Uruguay redujo grandemente las disputas entre los trabajadores y los dirigentes empresariales de Uruguay.

13) La participación activa de los sindicatos resolvió muchas controversias en todas las reuniones de los comités de la Cámara de Comercio.

14) La mayoría de los representantes de los sindicatos lamentaron haber aceptado la invitación a participar en la Cámara de Comercio.

15) Algunos de los miembros de la Cámara de Comercio llegaron a pensar que su presidente había estado desacertado al pedirles a los representantes sindicales que se unieran a la Cámara.

16) El nuevo presidente señaló en su discurso que los sindicatos de la ciudad no habían aceptado todavía su completa responsabilidad por el mejoramiento cívico.

### **Reconocimiento de supuestos**

#### Instrucciones

Un supuesto es algo afirmado implícitamente o dado por sentado. Cuando usted dice: “Me recibiré en junio”, usted da por sentado o asume que estará vivo en junio, que su escuela juzgará que usted reúne las condiciones para recibirse y cosas similares.

Abajo hay algunas afirmaciones. Cada afirmación es seguida de varios supuestos. Debe decidir con respecto a cada uno si una persona, al hacer la afirmación, está realmente implicando este supuesto; esto es, dándolo por sentado, justificadamente o no.

### Ejercicio

Afirmación: “Esquel es la ciudad para mudarse: tiene los impuestos más bajos”.

Supuestos:

24) Impuestos más bajos implican administración eficiente de la ciudad.

25) Al decidir dónde vivir, es importante evitar impuestos elevados.

26) La mayoría de los residentes de Esquel están contentos con el actual gobierno de su ciudad.

### Deducción

#### Instrucciones

En este *test* cada ejercicio comprende afirmaciones (premisas) seguidas de conclusiones. A los efectos de este test considere las afirmaciones de cada ejercicio como Verdaderas. Lea cada conclusión. Si piensa que se sigue necesariamente de las afirmaciones dadas, haga una marca (en su hoja de respuesta) en el casillero **SÍ**. Si piensa que no se sigue necesariamente, marque **NO**, aunque usted la considere verdadera por su cultura general.

Trate que sus prejuicios no influyan en su juicio: simplemente aténgase a las afirmaciones dadas (premisas) y juzgue cada conclusión en cuanto a si se sigue o no necesariamente de ellas. Las palabras “alguno/a/os/as” significan una parte o cantidad indefinida de cosas “alguno/a/os/as” significa al menos una porción, y tal vez toda la clase. De este modo, “Algunos días feriados son lluviosos” significa al menos uno, posiblemente más de uno y tal vez aún todos los días feriados son lluviosos.

### Ejercicio

El arroz y el apio necesitan mucha humedad para crecer bien; por su parte, el centeno y el algodón crecen mejor en lugares relativamente secos. El arroz y el algodón crecen en lugares cálidos; el apio y el centeno en lugares frescos. En Timbuctú hace mucho calor y humedad. Por lo tanto:

38) Ni las condiciones de temperatura ni las de humedad de Timbuctú son favorables para el cultivo del apio

39) Las condiciones de temperatura y humedad de Timbuctú son más favorables para el cultivo del arroz que para el cultivo del apio, algodón o centeno.

40) Las condiciones de Timbuctú no son del todo favorables para el cultivo del algodón ni del centeno.

### Interpretación

#### Instrucciones

Cada uno de los ejercicios siguientes comprende un párrafo corto seguido de conclusiones. A los efectos de este test asuma que todo lo contenido en el párrafo es verdadero. Debe juzgar si cada una de las conclusiones propuestas se sigue lógicamente más allá de una duda razonable de la información suministrada en el párrafo o no. Si piensa que la conclusión propuesta se sigue más allá de una duda razonable (aunque pueda no seguirse absoluta y necesariamente), entonces haga una marca en el casillero **SÍ** (la conclusión se sigue) en su hoja de respuestas. Si piensa que la conclusión no se

sigue más allá de una duda razonable, entonces marque NO. Recuerde que debe juzgar cada conclusión independientemente.

### Ejercicio

En cierta ciudad donde las normas de asistencia a clase se aplican estrictamente, se halló que sólo el 15% de los alumnos tenía asistencia perfecta en un semestre. Sin embargo, entre alumnos que trabajaban, el 25% tenía asistencia perfecta en el mismo semestre.

52) Es más probable que los alumnos que trabajan tuvieran asistencia perfecta en el semestre que la tuvieran los alumnos que no trabajan.

53) La estricta aplicación de las normas de asistencia a clase en esta ciudad no impidió que el 85% de los alumnos estuviera ausente alguna vez durante el semestre.

54) Si a los alumnos que faltan a clase sin justificación se les diera trabajo, su asistencia a clase mejoraría.

### Evaluación de argumentos

#### Instrucciones

Al tomar decisiones importantes, es necesario distinguir entre argumentos que son fuertes y argumentos que son débiles. Para que un argumento sea fuerte debe ser importante y estar directamente relacionados con la cuestión. Un argumento es débil si no está relacionado directamente con la cuestión (aunque pueda ser de una gran importancia general) o si se relaciona solamente con aspectos triviales de la cuestión. Se presentan preguntas seguidas de varios argumentos. A los efectos de este test usted debe considerar cada argumento como verdadero. Su tarea es entonces decidir si es un argumento fuerte o débil. Haga una marca en la hoja de respuestas debajo de SÍ, si piensa que el argumento es fuerte o debajo de NO, si piensa que es débil. Juzgue cada argumento por separado, de acuerdo con su propio mérito. Trate de que su actitud personal con relación a la cuestión no influya en su evaluación del argumento, ya que cada argumento debe considerarse verdadero.

### Ejercicio

¿Promovería un partido ecologista el bienestar general del pueblo argentino?

55) No; un partido ecologista en el gobierno hace que los inversores en compañías industriales no arriesguen su dinero en emprendimientos comerciales, causando de este modo un sostenido desempleo en gran escala.

56) Sí; hoy las diferencias entre justicialistas y radicales no son tan grandes como las diferencias entre sus respectivas líneas internas.

57) No; los ecologistas han convocado protestas frente a algunas industrias importantes.

## ANEXO II

## Tabla 1. Anexo.

## Notas

<sup>i</sup> El movimiento emerge en Estados Unidos en la década del 80 el seno de la Psicología cognitiva. Aparece como dato relevante en su gestación el reconocimiento de las deficiencias en los procesos de pensamiento de alto nivel, denominador común en las aulas escolares, comprobación que se verifica después en otros países.

<sup>ii</sup> Este trabajo es parte de uno mayor de la Dra. Da Dalt de Mangione, que se halla en proceso, con una muestra de aproximadamente 900 sujetos de escuelas rural-marginales.

<sup>iii</sup> Es un hecho constatado por los docentes y los expertos en evaluación que ésta produce una especie de *selección natural* de fines, objetivos y contenidos, ya que el alumno atiende a los que son evaluados y desestima los demás, que quedan en el plano de meros enunciados desiderativos, con el consiguiente empobrecimiento del currículum real.

<sup>iv</sup> Según Norris y Ennis (1989: 37-38), un *test comprehensivo* es una técnica que incluye tareas en cada una de las grandes áreas del pensamiento crítico. Aunque ningún instrumento estandarizado cabe completamente en la categoría, se considera que aquellos que cubren tres o cuatro áreas centrales satisfacen las condiciones mínimas para ser llamados comprensivos. Llamamos *test específico* al que se centra en un aspecto del pensamiento crítico, útil para proporcionar un diagnóstico detallado de habilidades complejas (como juzgar la credibilidad de una fuente).

<sup>v</sup> pp. 368-369

<sup>vi</sup> Cfr. Anexo I donde se ejemplifican estas dimensiones.

<sup>vii</sup> Por ejemplo, en el Colegio A, mientras dos adolescentes tardan 15 minutos en completar la primera parte del instrumento, otras tres necesitaron 40 minutos. Esta misma apreciación se registra en la aplicación del *test* New Jersey de Virginia Shipman (1983), en cinco colegios rural-marginales.

<sup>viii</sup> Agradecemos este análisis a la Dra. Stella Maris Vázquez.

<sup>ix</sup> P. 227

Gráfico 1

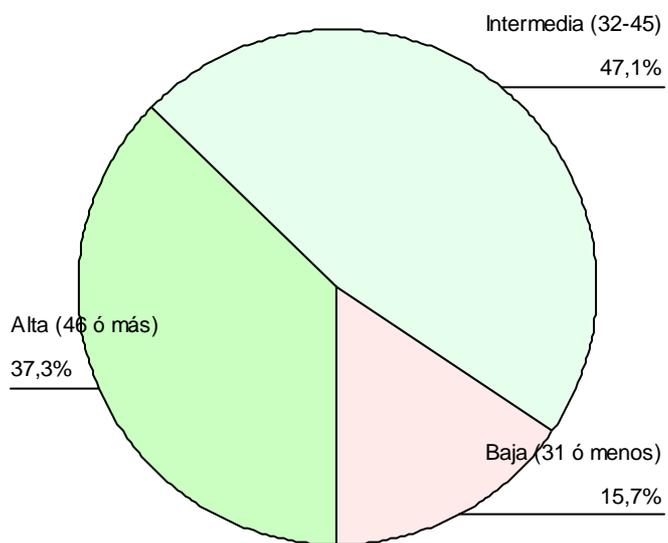


Gráfico 2.

Por institución educativa, los resultados en un box-plot son:

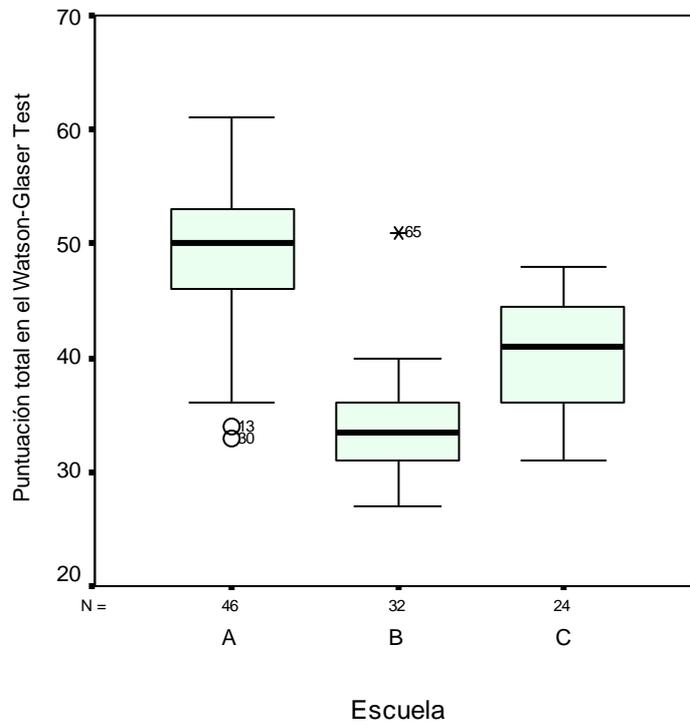
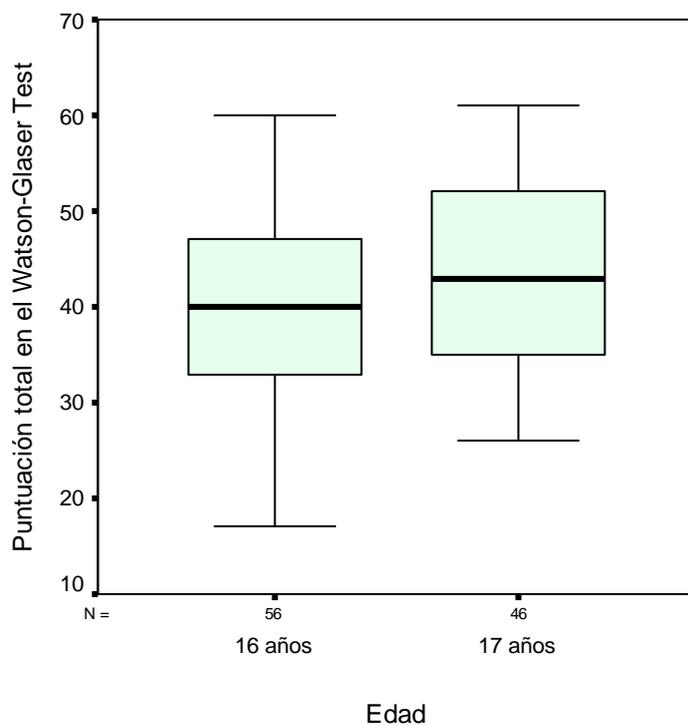


Gráfico 3

Por edad<sup>1</sup>, los resultados son:



<sup>1</sup> A fin de simplificar el análisis, hemos subsumido los cinco alumnos con 18 años en el grupo de 17 años.

Gráfico 4

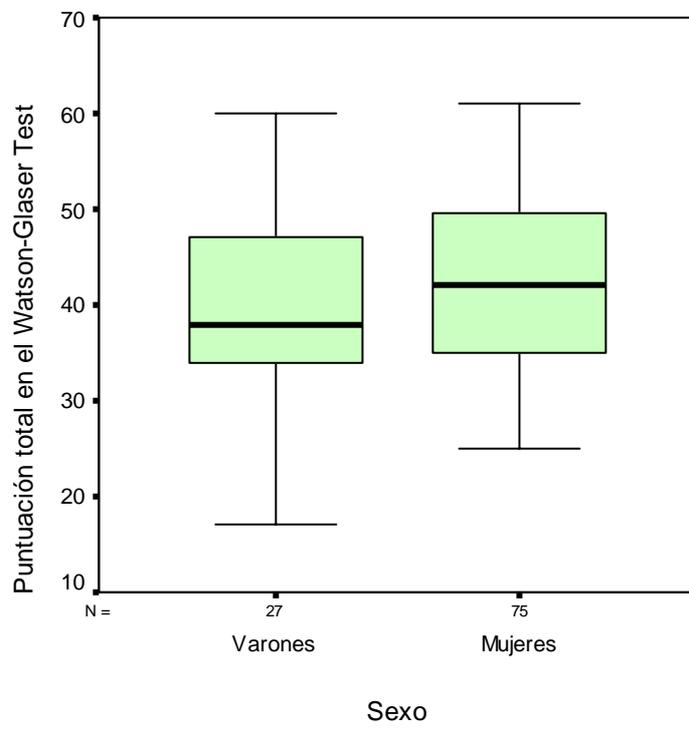


Gráfico 5

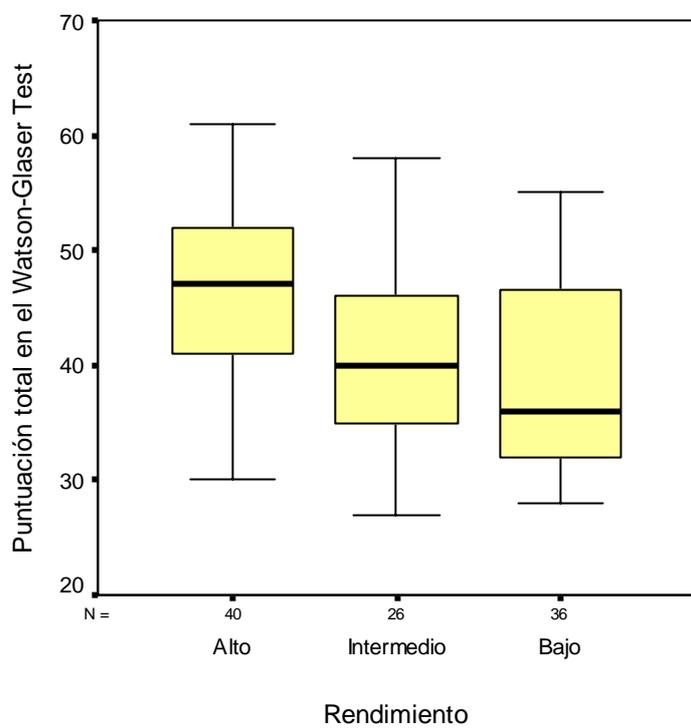


Tabla 1  
 Correlaciones entre las subescalas y la puntuación total del  
 WGCTA

<i>Dimensión</i>	1)	2)	3)	4)	5)
<b>1) Inferencia</b>	1	<b>0,51</b>	<b>0,35</b>	<b>0,36</b>	<b>0,54</b>
<b>2) Reconocimiento de supuestos</b>		1	<b>0,38</b>	<b>0,43</b>	<b>0,52</b>
<b>3) Deducción</b>			1,00	<b>0,42</b>	0,18
<b>4) Interpretación</b>				1,00	<b>0,26</b>
<b>5) Evaluación de argumentos</b>					1,00
<b>Total</b>	<b>0,73</b>	<b>0,76</b>	<b>0,54</b>	<b>0,62</b>	<b>0,67</b>

Los coeficientes destacados son significativos a nivel  $p < 0,05$ .

Tabla 2  
 Matriz de componentes

	Componente
	1
Reconocimiento de supuestos	0,809
Inferencia	0,781
Evaluación de argumentos	0,715
Interpretación	0,683
Deducción	0,610

Método de extracción: Análisis de componentes principales.  
 a 1 componentes extraídos

Tabla 3

Discriminación Homogeneidad	Muy buena (índice de 40% o más)	Buena (índice entre 30% y 39%)	Aceptable (índice entre 20% y 29%)
Muy buena (correlación de 0,40 o más)	17, 18, 19, 23, 29, 30, 42, 50, 57, 61, 65, 66	40, 49, 62	-----
Buena (correlación entre 0,30-0,39)	5, 13, 22, 24, 27, 45, 46, 52, 64	7, 15, 16, 36	-----
Aceptable (correlación entre 0,21-0,29)	-----	4, 8, 12, 55	14, 25, 28, 38, 39, 48, 53, 58, 59, 63

ANEXO II

Tabla 1. Anexo.

Análisis cuantitativo de los ítems: Índices de homogeneidad y discriminación, nivel de dificultad

<i>Ítem N°</i>	<i>Correlación ítem-puntuación total</i>	<i>N° de aciertos grupo superior</i>	<i>N° de aciertos grupo inferior</i>	<i>ID</i>	<i>ND</i>
1	0,17	16	10	0,2 2	0,42
2	0,04	16	15	0,0 4	0,43
3	0,02	17	15	0,0 7	0,54
4	<b>0,26</b>	15	6	<b>0,3 2</b>	0,38
5	<b>0,38</b>	17	4	<b>0,4 6</b>	0,35
6	0,15	20	15	0,1 8	0,52
7	<b>0,30</b>	12	1	<b>0,3 9</b>	0,33
8	<b>0,29</b>	14	5	<b>0,3 2</b>	0,28

<i>Ítem Nº</i>	<i>Correlación ítem- puntuación total</i>	<i>Nº de aciertos grupo superior</i>	<i>Nº de aciertos grupo inferior</i>	<i>ID</i>	<i>ND</i>
				<b>2</b>	
<b>9</b>	0,07	10	7	0,1 1	0,29
<b>10</b>	0,05	10	8	0,0 7	0,35
<b>11</b>	-0,04	14	15	- 0,0 4	0,55
<b>12</b>	<b>0,29</b>	20	9	<b>0,3 9</b>	0,54
<b>13</b>	<b>0,33</b>	23	11	<b>0,4 3</b>	0,48
<b>14</b>	<b>0,26</b>	22	16	0,2 2	0,74
<b>15</b>	<b>0,32</b>	24	15	<b>0,3 2</b>	0,69
<b>16</b>	<b>0,35</b>	22	13	<b>0,3 2</b>	0,49
<b>17</b>	<b>0,57</b>	24	5	<b>0,6 8</b>	0,33
<b>18</b>	<b>0,58</b>	22	3	<b>0,6 8</b>	0,37
<b>19</b>	<b>0,40</b>	22	7	<b>0,5 4</b>	0,31
<b>20</b>	0,15	16	10	0,2 2	0,38
<b>21</b>	0,00	23	21	0,0 7	0,77
<b>22</b>	<b>0,37</b>	25	13	<b>0,4 3</b>	0,69
<b>23</b>	<b>0,52</b>	26	10	<b>0,5 7</b>	0,70
<b>24</b>	<b>0,33</b>	26	13	<b>0,4 6</b>	0,65
<b>25</b>	<b>0,23</b>	28	22	0,2 2	0,85
<b>26</b>	0,06	18	14	0,1 4	0,69
<b>27</b>	<b>0,38</b>	26	12	<b>0,5 0</b>	0,61

<i>Ítem Nº</i>	<i>Correlación ítem- puntuación total</i>	<i>Nº de aciertos grupo superior</i>	<i>Nº de aciertos grupo inferior</i>	<i>ID</i>	<i>ND</i>
28	0,27	25	19	0,2 2	0,83
29	0,40	27	15	0,4 3	0,80
30	0,45	26	14	0,4 3	0,69
31	0,39	28	25	0,1 1	0,95
32	0,19	26	13	0,4 6	0,64
33	-0,19	6	13	- 0,2 5	0,45
34	0,07	22	20	0,0 7	0,75
35	0,18	24	17	0,2 5	0,77
36	0,30	23	14	0,3 2	0,54
37	0,15	21	18	0,1 1	0,59
38	0,28	25	19	0,2 2	0,76
39	0,27	27	20	0,2 5	0,86
40	0,40	25	15	0,3 6	0,73
41	0,07	24	24	0	0,76
42	0,53	25	7	0,6 4	0,59
43	0,07	15	12	0,1 1	0,48
44	0,06	22	22	0	0,76
45	0,35	28	14	0,5 0	0,73
46	0,33	26	14	0,4 3	0,61
47	0,15	25	21	0,1 4	0,81
48	0,22	22	14	0,2 0	0,55

<i>Ítem Nº</i>	<i>Correlación ítem- puntuación total</i>	<i>Nº de aciertos grupo superior</i>	<i>Nº de aciertos grupo inferior</i>	<i>ID</i>	<i>ND</i>
				9	
<b>49</b>	<b>0,40</b>	27	16	<b>0,3 9</b>	0,79
<b>50</b>	<b>0,41</b>	24	8	<b>0,5 7</b>	0,56
<b>51</b>	0,16	24	23	0,0 4	0,84
<b>52</b>	<b>0,37</b>	28	16	<b>0,4 3</b>	0,72
<b>53</b>	<b>0,21</b>	23	17	0,2 2	0,67
<b>54</b>	0,05	19	20	- 0,0 4	0,73
<b>55</b>	<b>0,23</b>	23	14	<b>0,3 2</b>	0,68
<b>56</b>	0,16	27	20	0,2 5	0,81
<b>57</b>	<b>0,48</b>	26	10	<b>0,5 7</b>	0,66
<b>58</b>	<b>0,23</b>	27	20	0,2 5	0,76
<b>59</b>	<b>0,25</b>	28	20	0,2 9	0,82
<b>60</b>	0,02	22	21	0,0 4	0,77
<b>61</b>	<b>0,45</b>	27	13	<b>0,5 0</b>	0,71
<b>62</b>	<b>0,41</b>	26	15	<b>0,3 9</b>	0,73
<b>63</b>	<b>0,24</b>	25	17	0,2 9	0,70
<b>64</b>	<b>0,39</b>	26	13	<b>0,4 6</b>	0,75
<b>65</b>	<b>0,44</b>	26	9	<b>0,6 1</b>	0,69
<b>66</b>	<b>0,46</b>	25	8	<b>0,6 1</b>	0,54

Hemos destacado en letra negrita los índices de discriminación buena -entre 0,30 y 0,39- y muy buena -de 0,40 o más-; en cursiva, los coeficientes entre 0,20 y 0,29 que significan discriminación aceptable.