

EVALUACIÓN PSICOMÉTRICA DEL CUESTIONARIO DE ANSIEDAD Y RENDIMIENTO (CAR) EN UNA MUESTRA DE ESCOLARES

Pere Joan Ferrando, M^a Dolores Varca y Urbano Lorenzo
Universidad Rovira i Virgili

Se presenta un estudio psicométrico, que pretende evaluar las propiedades de un cuestionario de ansiedad de tarea diseñado como predictor de rendimiento académico. En particular, se estudian: (a) la estructura del cuestionario, (b) su invarianza, y (c) su validez predictiva. Aunque el trabajo es fundamentalmente substantivo, sirve también como ejemplo para ilustrar el uso de nuevas metodologías en este tipo de estudios. Los resultados se discuten en términos substantivos y de utilidad práctica.

A psychometric study of the Anxiety and Performance Questionnaire (CAR) in primary school students In this study, the psychometric properties of a Test Anxiety questionnaire, which was designed mainly as a predictor of academic performance, are assessed. In more detail, we study: its structure, its invariance across gender, and its criterion related validity. Even though the work is mainly substantive, it also serves to illustrate some new procedures which can be effectively used in this type of studies. Results are discussed mainly in terms of practical applications and usefulness of the questionnaire.

El estudio de las relaciones entre la ansiedad y el rendimiento escolar ha ocupado el interés de psicólogos y educadores durante más de 50 años, y la revisión de la literatura no muestra el menor indicio de que este interés vaya a decaer. El lector interesado en el tema puede obtener una visión bastante actual del estado de la cuestión consultando revisiones teóricas relativamente recientes (King, Ollendick y Gullone, 1991; Birenbaum y Nasser, 1994), así como los

meta-análisis llevados a cabo por Hembree (1988) y por Seipp (1991).

En general, los resultados de la mayor parte de estudios empíricos al respecto, indican que el interés antes mencionado está bien justificado. La ansiedad, o mejor, algunos componentes del constructo de ansiedad, muestran consistentemente relaciones negativas con diversas medidas del rendimiento académico (Seipp, 1991). Si se admite esta relación, parece claro que la evaluación fiable y válida de los niveles de ansiedad en las dimensiones relevantes es un tema de importancia, tanto en la práctica educativa como en la clínica.

El presente trabajo analiza las propiedades psicométricas de un cuestionario de an-

Correspondencia: Pere Joan Ferrando
Facultad de Psicología
Universidad Rovira i Virgili
43007 Tarragona (Spain)
E-mail: pjfp@fcep.urv.es

siedad y rendimiento académico, desarrollado en nuestro país, en el que se trabaja desde hace más de una década. Si bien la finalidad principal del estudio es substantiva, puede considerarse que el trabajo tiene también un componente metodológico importante, ya que ilustra la posibilidad de aplicar nuevas metodologías a la evaluación psicométrica de cuestionarios de este tipo. En particular, se muestra la posibilidad de calibrar un cuestionario multidimensional, con un número relativamente alto de reactivos, mediante un modelo de teoría de respuesta a los ítems (TRI).

Fundamentación Teórica del CAR

El marco teórico general en el que se desarrolla el 'Cuestionario de Ansiedad y Rendimiento' C.A.R. (Aguilar, 1984) es el de la 'Ansiedad de Prueba' o 'Ansiedad de Tarea' (Test Anxiety en la terminología original anglosajona), una teoría que empezaron a desarrollar Mandler y Sarason (1952), de la universidad de Yale, a principios de los 50. De acuerdo con estos autores, en sujetos caracterizados por una elevada ansiedad de prueba se produciría, en situaciones de evaluación, una activación de dos estados impulsivos de ansiedad: por una parte, se produciría un incremento, tanto de la activación autonómica como de pensamientos y preocupaciones internas, que harían que la atención se desviara de la tarea, por lo que se perderían energía y tiempo que deberían haber sido dedicados a la resolución de la misma (ansiedad perturbadora del rendimiento). Por otra, se activarían una serie de respuestas de ansiedad estrechamente relacionadas con la tarea (Alpert y Haber, 1960). Este segundo componente sería la ansiedad que Pelechano (1973, 1975) denomina 'facilitadora del rendimiento'. Las principales aportaciones de la teoría de la escuela de Yale pueden encontrarse en Sarason (1980).

El C.A.R. incorpora también una propuesta, posterior a la de Sarason, que distingue dos componentes de ansiedad perturbadora del rendimiento. Es la teoría de la ansiedad dual de Liebert y Morris (1967), según la cual los dos componentes de ansiedad perturbadora que se activan en situación de evaluación son el de preocupación (Worry, W) y el de emocionalidad (E). El primero se refiere a la focalización de la atención sobre pensamientos y preocupaciones irrelevantes para la tarea. El segundo se refiere al estado afectivo generado por la activación fisiológica, es decir, a la percepción que tiene el sujeto de su propia activación.

La sencilla teoría de la ansiedad dual se hizo muy popular dentro del marco teórico cognitivo-atencional (véase Morris, Davis y Hutchings, 1981, para una revisión) y sigue generando actualmente investigación en varias direcciones. Por una parte, se han venido llevando a cabo estudios empíricos transculturales para evaluar si la distinción entre W y E se mantiene en culturas no anglosajonas (p. ej. Araki, 1992). Por otra se han llevado a cabo estudios similares para evaluar posibles diferencias por sexos (Benson, Bandalos y Hutchinson, 1994, Everson, Millsap y Rodríguez, 1991), así como efectos diferenciales sobre distintas materias del currículum (Everson, Tobias, Hartman y Gourgey, 1993). En conjunto puede afirmarse que la distinción W-E está actualmente bien establecida (Hardy y Hagtvet, 1996). Sin embargo esto no significa que la teoría esté exenta de problemas.

Los principales problemas del modelo dual son, fundamentalmente, problemas de medida. Por una parte, si bien los análisis estructurales suelen separar con relativa claridad los ítems de preocupación de los de emocionalidad, en cambio, cuando se analizan diferentes instrumentos que poseen escalas W-E mediante matrices multimétod-multirrasgo no puede, en general, establecerse validez discriminante (Williams,

1994). Es decir, pesa más la especificidad del instrumento de medida que la distinción W-E. Por otra parte, existe el clásico problema de la escasa validez predictiva de las medidas derivadas de la teoría (Hardy y Hagtvet, 1996). Por último, algunos autores han cuestionado que el rasgo de preocupación sea unidimensional (Sarason, 1984, Birenbaum, 1990).

Desde los principios teóricos que se han descrito hasta ahora, Aguilar, (1984) empezó a desarrollar un instrumento de investigación, compuesto inicialmente por 110 ítems dicotómicos, que pretendía ser un predictor del desempeño académico tanto positivo como negativo. Los resultados obtenidos en una serie de investigaciones con niños (no publicadas), sugirieron acortar el cuestionario, eliminando sobre todo ítems que no parecían relevantes cara a la validez predictiva, ya que este era el criterio principal en el que se basó la elaboración del test. La versión sobre la que se trabaja en el presente estudio consta de 30 ítems, los cuales se presentan en la tabla 1.

Tal como puede apreciarse en la tabla 1, los ítems del CAR se refieren a situaciones imaginadas de evaluación y pretenden ser un reflejo del pensamiento verbal del sujeto en situación de examen. En esencia el CAR es una prueba específica y situacional, de carácter cognoscitivo y de alcance restringido.

Supuestos y Organización del Estudio

De acuerdo con los principios teóricos antes expuestos, y con la evidencia empírica, pueden plantearse una serie de supuestos que se espera tengan utilidad como guía de la evaluación psicométrica que seguirá. Así:

– El C.A.R. está diseñado para tener una estructura factorial tridimensional (Preocupación, Emocionalidad y Ansiedad Facilitadora), con factores correlacionados. Se espera que preocupación y emocionalidad co-

Tabla 1
Los 30 ítems del C.A.R.

Preocupación	
1.	Me desanima pensar que otros hacen cosas más difíciles.
2.	Pierdo el tiempo pensando en lo que me falta por hacer.
3.	Cuando los demás me miran no soy capaz de pensar correctamente.
4.	Siento frustración por no ser capaz de resolver la tarea.
5.	Durante la explicación que nos han dado pensaba que debía atender, pero no me he enterado de qué debo hacer.
6.	A veces resuelvo problemas muy difíciles y ahora no soy capaz de resolver éstos, los demás si lo harán.
7.	Por pensar en los fracasos anteriores no puedo concentrarme.
8.	Cuanto más me esfuerzo, menos me concentro en la tarea.
9.	Me gustaría consultar algo, pero me da miedo pasar por tonto.
10.	Pienso tanto en cada detalle, que no tendré tiempo de acabar.
11.	Me concentro peor cuanto más cerca estoy del examen.
12.	Estoy tan desalentado y convencido del mal resultado que desconfío de mi capacidad para preparar el examen.
13.	La preocupación por las notas me impide trabajar.
14.	En vez de concentrarme en la tarea pienso que es tan difícil que me saldrá mal. Me lamento constantemente.
15.	Me preocupan tantas cosas que no puedo concentrarme.
Emocionalidad	
16.	Con nervios no puedo hacer ni las cosas más fáciles.
17.	A lo largo de la tarea se me pone un peso en el estómago
18.	Tengo tantos nervios que no paro de moverme.
19.	Antes del examen tengo temblores, sudores, trastornos digestivos, dolores de cabeza.
20.	Los nervios y la emoción me impiden ordenar las ideas.
21.	Los nervios me han hecho olvidar hasta las cosas más sencillas.
22.	Me sudan o tiemblan las manos. El corazón late muy deprisa.
23.	Me tiemblan las manos sólo de pensar lo que dirán mis padres si no lo hago bien.
24.	No puedo pensar porque en este momento no me encuentro bien.(Durante el examen).
Ansiedad Facilitadora	
25.	Trabajo deprisa y bien, cuando lo descubro mejoro aún más.
26.	Cuanto más me estimulo, más rápido y mejor lo hago.
27.	A medida que voy entendiendo la tarea, siento como una excitación que me hace concentrar más y mejor en todo.
28.	La proximidad del examen me excita y entiendo mejor las cosas.
29.	Estoy tan excitado y despierto que resolvería cualquier problema.
30.	Cuanto más preocupado estoy, mejor me sale el examen.

relacionen fuerte y positivamente entre sí (Morris et al. 1981) y que la ansiedad facilitadora correlacione negativamente con los otros dos factores (Alpert y Haber, 1960).

– Las escalas de Preocupación y de Emocionalidad se espera que correlacionen negativamente con un criterio externo de rendimiento académico, en tanto que la escala de Ansiedad Facilitadora debería mostrar una correlación positiva con dicho criterio. Los estudios de revisión y meta-análisis (Deffenbacher, 1980; Morris et al. 1981; Hembree 1988, Seipp 1991) indican que la correlación entre la escala de preocupación y el criterio debe ser superior a la correlación entre la escala de emocionalidad y el criterio.

La investigación que sigue está estructurada en tres partes. En la primera parte se estudia la dimensionalidad y estructura del CAR y se justifica la metodología empleada. En la segunda parte se evalúa la invariancia de la solución en niños y niñas, justificándose la fundamentación substantiva del estudio. La tercera parte, por último, describe un estudio de validez predictiva, utilizando un criterio externo de rendimiento académico, en la que se comparan dos métodos para estimar los niveles de ansiedad.

Método

Participantes

La muestra total se compone de 1823 escolares de 12 a 14 años (895 niños y 928 niñas). Los datos se empezaron a recoger en 1993 y forman parte de un proyecto de investigación más general, una de cuyas etapas es el análisis y validación de los instrumentos de medida de la ansiedad.

Colaboraron un total de 11 centros escolares, 6 de ellos eran colegios públicos y los 5 restantes privados concertados; todos ellos ubicados en el núcleo urbano y alrededores de Tarragona.

Procedimientos

La puesta en contacto con los centros se hizo en un primer momento telefónicamente. Posteriormente había una entrevista entre el director o coordinador de la etapa a la que se le iba a administrar las pruebas y el investigador; en la cual se le explicaban los motivos de la investigación y de qué constaba la misma.

La prueba se administraba en el grupo clase formado por 25 o 30 alumnos, los profesores se ausentaban del aula quedándose sólo el investigador y los niños. La persona que recogió los datos fue siempre la misma.

En el primer contacto entre el investigador y los alumnos éste se presentaba, y les daba las siguientes consignas:

a) Lo que vamos a realizar a continuación nada tiene que ver con la escuela, es decir, los resultados de lo que hagáis no lo verán vuestros profesores y podéis estar tranquilos porque no forma parte de la evaluación académica. Además los datos sólo los verá yo.

b) Es importante que seáis sinceros en vuestras contestaciones, que tengáis una actitud positiva y que estéis concentrados en lo que hacéis.

c) Si hay algún alumno que no desee realizar las pruebas puede optar libremente por ello.

Una vez dadas las diferentes consignas se iniciaba la administración, en forma oral, es decir, el investigador leía la pregunta y ellos debían contestar en un folio con un «SÍ» o un «NO», si no entendían lo que se les preguntaba podían indicarlo y el investigador les explicaba la pregunta. Se les repetía dos veces cada ítem, y entre pregunta y pregunta se les daba un tiempo para que pudiesen contestar. Es importante indicar que antes de contestar a este test se les indicaba que las preguntas las debían interpretar situán-

dose en una situación de examen o de deberes, los cuales eran muy importantes para ellos.

Análisis y Resultados

Evaluación de la Estructura del CAR

La evaluación de la estructura del CAR se ha basado en un análisis factorial (AF) no-lineal sobre la matriz de productos momentos. Dado que este procedimiento no es demasiado habitual en investigación aplicada, se tratará a continuación de justificar su uso.

El AF lineal común, la técnica más habitual para evaluar la estructura de un test, es un modelo para variables continuas e ilimitadas, y relaciones lineales. Al utilizarlo sobre ítems binarios (que serían el caso extremo de variables discretas y limitadas), se pueden producir distorsiones en el caso de que las relaciones ítem-factor no sean lineales (véase McDonald y Ahlawat, 1974). Tales distorsiones consisten, por una parte en efectos de atenuación sobre las cargas factoriales (Collins, Cliff, MacCormick y Zatzkin, 1986) y por otra en evidencia espuria de multidimensionalidad, es decir, que para conseguir un buen ajuste del modelo, es necesario extraer más factores de los que realmente tienen significado substantivo.

El modelo AF no lineal utilizado en este estudio, asume que la regresión ítem factor adopta la forma de ojiva propia del modelo de TRI de dos parámetros. Esta es una relación bastante más plausible que la lineal cuando los ítems son binarios y, por tanto, el procedimiento parece ser a priori, teóricamente más adecuado.

Más en detalle, el procedimiento utilizado en el trabajo sigue el modelo general de AF no lineal de McDonald (1967), se basa en el análisis armónico y está implementado en el programa NOHARM (Fraser y McDonald, 1988). En primer lugar, NOHARM estima los parámetros de los ítems en el mo-

delo TRI de ojiva normal según el criterio de mínimos cuadrados, y después, mediante reparametrización, obtiene las saturaciones factoriales. Dado que los estimadores mínimo cuadráticos son consistentes, debe notarse que, si el modelo elegido es correcto y la muestra razonablemente grande, los valores estimados deben de ser muy similares a los que se hubiesen obtenido utilizando métodos más sofisticados.

Una limitación de NOHARM es la falta de un criterio formal para evaluar la bondad de ajuste. Sin embargo, recientemente, Gessaroli y De Champlain (1996) han propuesto un estadístico que sigue aproximadamente una distribución ji-cuadrado bajo la hipótesis nula de que los elementos no-diagonales de la matriz de correlaciones residuales son cero. El estadístico parece funcionar bastante bien en estudios de simulación. Por otra parte, en la tradición de los modelos de ecuaciones estructurales, McDonald y Mok (1995) sugieren utilizar la versión para mínimos cuadrados del índice normado de bondad de ajuste gamma, propuesto por Tanaka y Huba (1985). Estos nuevos criterios pueden complementar la información aportada por el output de NOHARM.

Para evaluar la dimensionalidad de los 30 ítems del CAR, se probaron modelos desde uno hasta cuatro factores. Aparte del criterio substantivo, se utilizaron los siguientes índices: (a) la suma de cuadrados de los residuales; (b) la media cuadrática residual; (c) el estadístico de Gessaroli y de Champlain y (d) el índice gamma de bondad de ajuste. Los indicadores (a) y (b) forman parte del output standard de NOHARM. Para los otros dos, se desarrolló un programa de cálculo en nuestro laboratorio.

Los resultados de la evaluación dimensional del CAR se presentan en la tabla 2.

En líneas generales, los resultados mostrados en la tabla 2 indican que el modelo unidimensional es inaceptable, en tanto que, a partir de dos factores, el ajuste empieza a

ser razonable. La inspección de las soluciones rotadas se puede resumir como sigue: en el modelo de un factor, los ítems de preocupación y emocionalidad tenían cargas positivas y los de ansiedad facilitadora negativas. Este patrón se descomponía en dos factores en el modelo siguiente (ansiedad perturbadora y ansiedad facilitadora). El modelo de tres factores será comentado a continuación. Por último el de cuatro factores era ininterpretable, produciéndose una atenuación general de las cargas que quedaban en su inmensa mayoría por debajo de .20. Claramente, extraer cuatro factores de esta matriz de datos es sobrefactorizar. Se decidió, por tanto, adoptar la solución en tres factores.

Modelo	SC _r	MC _r	$\chi^2_{(435)}$	γ
1 Factor	0.040	0.0096	2010.58	0.986
2 Factores	0.023	0.0073	810.79	0.994
3 Factores	0.014	0.0056	665.23	0.995
4 Factores	0.010	0.0049	484.31	0.996

Nota SC_r: Suma de cuadrados residual; MC_r: Media cuadrática residual; $\chi^2_{(435)}$: Estadístico aproximado ji-cuadrado con 435 grados de libertad; γ : Índice gamma de bondad de ajuste de Tanaka y Huba.

A continuación, el patrón directo en tres factores fue rotado utilizando el método Procrustes oblicuo semi-especificado (Browne, 1972). La matriz diana se especificó del siguiente modo: los 15 primeros ítems (preocupación) tenían cargas libres en el primer factor y cercanas a cero en los dos restantes. Los 9 siguientes ítems (emocionalidad) tenían cargas libres sólo en el tercer factor, y cerca de cero en los dos restantes. Por último, los 6 ítems finales (ansiedad facilitadora) saturaban libremente en el segundo factor y cerca de cero en los otros dos. La rotación se llevó a cabo mediante el programa TARROT (Browne, 1991).

El método de rotación descrito no es habitual en investigación aplicada. Sin embargo, los autores creemos que es apropiado en este caso ya que lo que se pretende es ver si existe (o no) una posición de rotación que concuerde con la estructura hipotetizada desde la teoría. De alguna forma, esto convierte al AF en semi-confirmatorio ya que, si la solución obtenida no concuerda demasiado con la matriz diana propuesta, entonces la hipótesis estructural es bastante cuestionable.

Los resultados se presentan en la tabla 3.

Los resultados de la tabla 3 son bastante claros. Los factores de preocupación (W) y de ansiedad facilitadora (F) están bien definidos, siendo los ítems bastante puros. En cambio, el factor de emocionalidad (E) está insuficientemente definido ya que sólo dos ítems (E4 y E7) demuestran ser buenos indicadores del mismo. En futuros trabajos, parece clara pues la necesidad de reforzar esta escala con nuevos ítems al tiempo que se eliminan otros.

La matriz de correlación entre factores se comporta de acuerdo con las predicciones hechas en la parte teórica. Preocupación y emocionalidad muestran entre sí una fuerte correlación positiva, en tanto que la ansiedad facilitadora correlaciona negativamente con los otros dos.

Estudios de Invariancia

Al final de su monografía sobre la ansiedad en escolares, Sarason y colaboradores (Sarason, Davidson, Lighthall, Waite y Ruebush, 1960, Cap. 9), se lamentaban por no haber tenido en cuenta desde el principio de su estudio las posibles diferencias sexuales en el constructo. Diferencias entre niños y niñas aparecían a lo largo de toda la investigación y se centraban en dos aspectos: (a) niveles claramente distintos en las puntuaciones totales en los diferentes tests y (b) predicciones diferenciales respecto a

diversos criterios de rendimiento académico. Estos dos hallazgos aparecen también sistemáticamente en investigaciones posteriores de otros grupos (Gaudry y Spielberger, 1971).

La revisión de la literatura indica que las diferencias sexuales en el constructo de ansiedad de prueba, se han tratado de explicar desde dos perspectivas, ambas basadas en el proceso de socialización. La primera, más directa, considera que las mujeres tienden a infravalorar su rendimiento, tienen más sentimientos negativos y se sienten menos confortables que los hombres en situaciones donde son evaluadas (ej. Benson et al., 1994). El segundo punto de vista, en cambio, considera que no hay diferencias reales en los niveles de ansiedad, pero que las mujeres están más dispuestas a admitir estos niveles, quizás porque desde edades muy tempranas se les anima a hacerlo, al ser este un rasgo 'femenino' en la cultura occidental (Dusek, 1980). De acuerdo con esta segunda teoría, por tanto, las diferencias entre puntuaciones, al menos en parte, podrían no reflejar diferencias reales en ansiedad, sino diferencias en la disposición a admitirla. Es desde esta segunda perspectiva, que ciertos autores (Maccoby y Jacklin, 1974) sugieren que algunos ítems de tests de ansiedad podrían tener un comportamiento diferencial en hombres y mujeres, lo que implicaría una estructura distinta en ambos grupos.

La evaluación e la invariancia de la solución en niños y niñas, se ha hecho siguiendo un procedimiento que consiste en rotar ambas soluciones a la solución más cercana posible que, a su vez, sea también interpretable desde un punto de vista teórico (véase por ejemplo Kiers, 1997). El procedimiento consiste en tres etapas.

1. Rotación ortogonal Procustes de uno de los patrones factoriales tomando como fijo al otro patrón (Cliff, 1966). Con esto se lleva a ambos patrones a la solución más cercana el sentido de los mínimos cuadrados.

2. Desde la posición anterior, rotación simultánea de ambos patrones factoriales a una posición de Estructura Simple mediante el procedimiento de la Matriz de Consensus

Tabla 3 Estructura del C.A.R. en tres dimensiones (parametrización factorial)			
a) Patrón rotado (rotación procusteana semiespecificada)			
	W	F	E
W1	0.436	0.118	0.086
W2	0.464	0.018	0.001
W3	0.417	0.077	0.069
W4	0.534	0.163	0.086
W5	0.622	-0.078	-0.094
W6	0.427	0.012	0.162
W7	0.548	0.025	0.167
W8	0.553	-0.094	-0.088
W9	0.486	-0.029	0.130
W10	0.465	0.145	0.162
W11	0.268	-0.256	0.397
W12	0.715	-0.251	-0.013
W13	0.540	-0.032	0.219
W14	0.681	-0.206	0.069
W15	0.583	-0.087	0.178
E1	0.303	-0.011	0.138
E2	0.327	0.039	0.353
E3	0.289	0.169	0.254
E4	-0.223	-0.017	0.943
E5	0.340	-0.052	0.406
E6	0.494	0.118	0.220
E7	-0.200	0.003	0.910
E8	0.409	0.136	0.248
E9	0.393	0.103	0.288
F1	-0.150	0.365	0.024
F2	0.058	0.440	0.054
F3	-0.048	0.576	0.252
F4	-0.007	0.669	-0.154
F5	-0.067	0.592	0.133
F6	0.052	0.425	0.035
b) Correlación entre factores			
	W	F	E
W	1.000	-0.296	0.616
F	-0.296	1.000	-0.168
E	0.616	-0.168	1.000

(Evans, 1971; Hakstian, 1973). Este procedimiento, a su vez, se desarrolla en tres pasos: 2-1) construir dicha Matriz de Consenso como el promedio de los dos patrones factoriales a comparar; 2-2) rotar dicha matriz mediante el criterio Varimax (Kaiser, 1958) utilizando 30 inicios aleatorios; y 2-3) transformar finalmente los dos patrones factoriales mediante la matriz de rotación Varimax obtenido en el paso 2-2.

3. En este punto, ambas soluciones se encuentran en una posición de máxima proximidad, a la vez que se ajustan a un criterio de estructura simple. Para compararlas, se calcula un índice de discrepancia: la raíz media-cuadrática residual (RMSR) y un índice de congruencia factorial: el de Burt-Tucker (1948).

Se desarrolló un programa informático en nuestro laboratorio para hacer operativo el procedimiento. Los resultados pueden resumirse como sigue: (a) La solución obtenida en ambos grupos era muy similar a la solución general descrita en el apartado anterior. (b) El índice de discrepancia entre las soluciones era $RMSR = 0.0512$. (c) los índices de congruencia eran: 0.9928 (F1), 0.9842 (F2) y 0.9250 (F3). En suma, resultados muy claros que indican que la estructura del CAR es virtualmente la misma en ambos sexos.

Precisión en la Medida y Validez Predictiva.

Se llevaron a cabo dos estudios de validez predictiva con referencia al mismo criterio: las calificaciones finales en la asignatura de matemáticas. La revisión de la literatura indica que esta asignatura es la que, consistentemente, muestra relaciones más claras con la ansiedad de tarea (ver p. ej. Hembree, 1988; Everson et al., 1993), de aquí la elección. De la muestra total, se disponía de las calificaciones en matemáticas de 1453 alumnos, siendo esta la muestra so-

bre la que se desarrollan los dos estudios que siguen.

El primer estudio es el que se utiliza tradicionalmente en la evaluación de la validez, y consiste en obtener las puntuaciones totales por escala mediante la suma simple de las respuestas a los ítems que definen dicha escala (puntuadas como cero y uno en la apropiada dirección). Después, estas puntuaciones se correlacionan con el criterio externo.

El segundo procedimiento es bastante más costoso. En primer lugar deben obtener las estimaciones del nivel de los sujetos en los factores a medir. A continuación son los estimadores los que se relacionan con el criterio.

Las relaciones teóricas entre los resultados obtenidos mediante ambos procedimientos son bien conocidas, y han sido sistematizadas por Drasgow y Miller (1982). En el caso de: (a) ítems factorialmente puros y estructuras claras y (b) escalas formadas por un buen número de reactivos, ambos métodos deben dar lugar a soluciones similares. Esto es así ya que, por una parte, en este caso las puntuaciones totales de la escala representan bastante adecuadamente al factor que intentan medir. Por otra, las combinaciones lineales de variables con pesos diferenciales tienden a dar resultados similares a las combinaciones con pesos uno y cero conforme aumenta el número de variables. Debe notarse, por último, que esta situación no se ha encontrado en el presente trabajo, por lo que son de esperar ciertas diferencias en las estimaciones de validez.

En este trabajo, como estimador de los niveles de los sujetos en cada uno de los tres rasgos o factores, se escogió la esperanza de la distribución posterior del vector de rasgos dado el vector de respuestas observado. El estimador descrito suele denominarse EAP (expected a posteriori) en la terminología anglosajona. Posee interesantes propiedades y la ventaja de ser no-iterativo. Sin

embargo, tiene la limitación de requerir la resolución de una integral múltiple que no puede obtenerse en forma cerrada, por lo que su evaluación suele hacerse por aproximación numérica, habitualmente mediante fórmulas de cuadratura. Para una descripción detallada del procedimiento aplicada al presente caso de un modelo TRI múltiple, puede consultarse Bock (1985).

Se desarrolló un programa de cálculo para obtener las estimaciones EAP de los 1453 sujetos. Algunos detalles técnicos son los siguientes: (a) para la estimación de la función de verosimilitud se utilizó la parametrización TRI del modelo (en lugar de la parametrización factorial presentada en la tabla 3), (b) La aproximación para la cuadratura en tres dimensiones se obtuvo mediante el producto de los nodos y pesos para la cuadratura unidimensional (véase Stroud y Sechrest, 1966) y (c) se utilizaron 6 puntos de cuadratura por dimensión (Bock, 1985 recomienda tan sólo 3 puntos por dimensión en el caso trifactorial, por lo que cabe esperar que las estimaciones obtenidas son suficientemente precisas).

Previamente a la evaluación de la validez, es conveniente tener una estimación de la precisión con la que las escalas, en un caso, o las puntuaciones factoriales en otro miden el rasgo que pretenden medir. En el caso de las puntuaciones totales, esto significa evaluar la fiabilidad de las tres escalas. Los resultados, utilizando el coeficiente alpha fueron: 0.81 para la escala de preocupación (15 ítems); 0.72 para la de emocionalidad (9 ítems) y 0.51 para la de ansiedad facilitadora (6 ítems). Las dos primeras estimaciones son razonables, habida cuenta de que se trata de escalas de personalidad con un reducido número de ítems.

En el caso de las estimaciones EAP, el concepto de fiabilidad como precisión global no tiene demasiado sentido, ya que, como es sabido, en modelos no-lineales de TRI la precisión con la que se estima el ras-

go latente es distinta en unos niveles que en otros. Para dar una idea de la precisión, en este caso, hemos optado por dar los valores mínimo, máximo y medio de los errores típicos de la distribución posterior para cada uno de los factores. Estos son: 0.067, 0.766 y 0.404 (factor de preocupación); 0.461, 0.820 y 0.680 (ansiedad facilitadora) y 0.239, 0.882 y 0.662 (emocionalidad). las correspondientes estimaciones de fiabilidad obtenidas según sugieren Bock y Mislevy (1982, fórmula 7) son: 0.995, 0.413 y 0.836 (preocupación), 0.787, 0.326 y 0.537 (ansiedad facilitadora) y 0.942, 0.222 y 0.562 (emocionalidad). Nótese que hay una cierta correspondencia entre los valores medios aquí citados y las estimaciones anteriores según el modelo clásico. Parece también claro que, en ambos casos la medida más precisa es la medida de la preocupación.

Los estimaciones de validez se presentan en la tabla 4.

Escala	ρ_{rr}	Factor	ρ_{rr}
W	-.325	FW	-.344
E	-.184	FE	.003
F	.221	FT	.176
N=1453			

En su revisión de 562 estudios empíricos, Hembree (1988) indica unas correlaciones promedio con diversos criterios de rendimiento académico de -0.26 para las escalas de preocupación, y de -0.19 para las de emocionalidad. Por otra parte, en el resumen presentado por Deffenbacher (1980), se estima que estas correlaciones son de -.30 y de -.10 respectivamente. Comparando esta información con los resultados obtenidos utilizando la suma simple de ítems, cabría concluir que el CAR se muestra en este estudio como un buen predictor de rendi-

miento, con coeficientes de validez estimados que se encuentran en el límite superior de lo que cabe esperar en este tipo de escalas. No hemos encontrado datos comparativos para la escala de ansiedad facilitadora del rendimiento.

La utilización de estimaciones EAP, en lugar de las puntuaciones directas por escala, no produce mejoras dramáticas, pero clarifica bastante las relaciones. El contraste más evidente es la ausencia de relación entre la escala de emocionalidad y el criterio. Esta discrepancia sugiere dos posibles explicaciones: (a) la ausencia de relación se debe a la falta de buenos indicadores en la escala (recordemos que sólo había dos) y (b) el rasgo de emocionalidad en realidad no influye sobre el rendimiento. De ser correcto el supuesto (b), entonces las relaciones observadas al utilizar puntuaciones directas, se deberían sobre todo al componente del rasgo de preocupación que tienen la mayor parte de ítems de emocionalidad al no ser factorialmente puros. En otras palabras, si se consiguiesen reactivos 'puros' de emocionalidad, con cargas cercanas a cero en el factor de preocupación, entonces estos ítems no estarían relacionados con el rendimiento. Los autores consideramos que este aspecto merece ser investigado bastante más a fondo.

Discusión

El Cuestionario de Ansiedad y Rendimiento demuestra, en líneas generales, unas características apropiadas desde el punto de vista psicométrico. La estructura es relativamente clara e invariante en niños y niñas y la validez es más que aceptable considerando el tipo de instrumento del que se trata. Es bien conocido que los tests de personalidad producen, casi siempre, predicciones más débiles que los tests de capacidad.

Cabe decir que algunos de los profesores de diversos centros donde se administró el test mostraron interés en el CAR, y es interesante hacer notar que la utilidad principal que estos profesionales le encontraban al cuestionario era la de identificar en el aula a aquellos alumnos con niveles exageradamente elevados de ansiedad.

En este punto, el lector podría cuestionarse si con un modesto coeficiente de validez de $-.34$ (en el mejor de los casos) es posible llegar a predicciones útiles desde un punto de vista aplicado. A este respecto, cabe decir que si se selecciona a los sujetos extremos (a partir del percentil 95) en el factor de preocupación, uno se encuentra que el porcentaje de fracasos en el criterio en este grupo es del 60 %, mientras que el mismo porcentaje en el grupo general está alrededor del 30%. Por tanto, la idea de identificar sujetos con niveles muy elevados de ansiedad que, se supone, interferirá negativamente en el rendimiento no parece tan descabellada.

Existen una serie de aspectos pendientes sobre los que debe aún trabajarse. La escala de emocionalidad necesita una revisión, y las posibles relaciones entre este rasgo y un criterio externo de rendimiento deben volver a evaluarse una vez reelaborada la escala. En la misma forma, aunque la escala de ansiedad facilitadora tiene una estructura clara, se muestra poco precisa en la estimación de este rasgo (hecho que se aprecia ya en las relativamente bajas saturaciones de los ítems). Mejorando la precisión, es de esperar que mejoraría también la capacidad predictiva. Todos estos aspectos deben ser objeto de futuras investigaciones. Por último, obviamente, si se pretende que el cuestionario tenga una utilidad aplicada más allá de la investigación, deben elaborarse unos baremos apropiados a la población escolar de nuestro país.

Referencias

- Aguilar, A. (1984). *CAR Cuestionario de ansiedad y rendimiento*. Doc. Mimeog. de uso exclusivo para la investigación. Tarragona: Seminario de Psicología.
- Alpert, R. y Haber, R.N. (1960). Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology*, 61, 207-215.
- Araki, N. (1992). Test anxiety in elementary school and junior high school students in Japan. *Anxiety, Stress and Coping*, 5, 205-215.
- Benson, J., Bandalos, D., y Hutchinson, S. (1994). Modeling test anxiety among men and women. *Anxiety, Stress, and Coping*, 7, 131-148.
- Birenbaum, M. (1990). Test anxiety components: comparison of different measures. *Anxiety Research*, 3, 149-150.
- Birenbaum, M., y Nasser, F. (1994). On the relationship between test anxiety and test performance. *Measurement and Evaluation in Counseling and Development*, 27, 293-301.
- Bock, R.D. (1985). Contributions of empirical Bayes and marginal maximum likelihood methods to the measurement of individual differences. En E.E. Roskam (ed.), *Measurement and Personality Assessment* (pp. 75-99). Amsterdam: Elsevier.
- Bock, R.D., y Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.
- Browne, M.W. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 207-212.
- Browne, M.W. (1991). *TARROT: Target Rotation Program*. Manuscrito no publicado.
- Burt, C. (1948). The factorial study of temperamental traits. *British Journal of Psychology*, 1, 178-203.
- Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, 31, 33-42.
- Collins, L.M., Cliff, N., McCormick, D.J., y Zatzkin, J.L. (1986). Factor recovery in binary data sets: a simulation. *Multivariate Behavioral Research*, 21, 377-391.
- Drasgow, F., y Miller, H.E. (1982). Psychometric and substantive issues in scale construction and validation. *Journal of Applied Psychology*, 67, 268-279.
- Deffenbacher, J.L. (1980). Worry and Emotionality in test anxiety. En I.G. Sarason (Ed.), *Test anxiety: theory, research and applications* (pp 111-128). Hillsdale: LEA.
- Dusek, J.B. (1980). The development of test anxiety in children. En I.G. Sarason (Ed.), *Test anxiety: theory, research and applications* (pp. 87-110). Hillsdale: LEA.
- Evans, G.T. (1971). Transformation of factor matrices to achieve congruence. *British Journal of Mathematical and Statistical Psychology*, 24, 22-48.
- Everson, H.T., Millsap, R.E., y Rodríguez, C.M. (1991). Isolating gender differences in test anxiety: a confirmatory factor analysis of the test anxiety inventory. *Educational and Psychological Measurement*, 51, 243-251.
- Everson, H.T., Tobias, S., Hartman, H., y Gourgey, A. (1993). Test anxiety and the curriculum: the subject matters. *Anxiety, Stress and Coping*, 6, 1-8.
- Fraser, C., y McDonald, R.P. (1988). NOHARM: least squares ítem factor analysis. *Multivariate Behavioral Research*, 23, 267-269.
- Gaudry, E., y Spielberger, Ch. D. (1971). *Anxiety and educational achievement*, New York: Wiley.
- Gessaroli, M.E., y De Champlain, A.F. (1996). Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of ítems. *Journal of Educational Measurement*, 33, 157-179.
- Hakstian, A.R. (1973). Procedures for the factor analytic treatment of measures obtained on different occasions. *British Journal of Mathematical and Statistical Psychology*, 26, 219-239.
- Hardy, L., y Hagtvet, K.A. (1996). Anxiety and performance: measurement and modelling issues. *Anxiety, Stress and Coping*, 9, 5-8.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58, 47-79.
- Kaiser, H.F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Kiers, H.A.L. (1997). Techniques for rotating two or more loading matrices to optimal agreement and simple structure: a comparison and some technical details. *Psychometrika*, 62, 545-568.
- King, N.J., Ollendick, T.H., y Gullone, E. (1991). Test anxiety in children and adolescents. *Australian Psychologist*, 26, 25-32.

- Liebert, R.M. y Morris, L.W. (1967) Cognitive and emotional components of test anxiety: a distinction and some initial data. *Psychological Reports*, 20, 975-978.
- Maccoby, E.E., y Jacklin, D. (1974). *The psychology of sex differences*. Stanford: Stanford University Press.
- Mandler, G., y Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, 47,166-173.
- McDonald, R.P. (1967). *Non linear factor analysis*. Psychometrika monograph n. 15.
- McDonald, R.P., y Ahlwat, K.S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82-99.
- McDonald, R.P., y Mok, M.C. (1995). Goodness of fit in ítem response models. *Multivariate Behavioral Research*, 30, 23-40.
- Morris, L.W., Davis, M.A., y Hutchings, C.H. (1981). Cognitive and emotional components of anxiety: literature review and a revised Worry-Emotionality scale. *Journal of Educational Psychology*, 73, 541-555.
- Pelechano, V. (1973). *Personalidad y parámetros: tres escuelas y un modelo*. Barcelona: Vicens Vives.
- Pelechano, V. (1975). Motivación y rendimiento académico. *Análisis y Modificación de conducta*, 1,83-110
- Sarason, S.B., Davidson, K.S., Lighthall, F.F., Waite, R.R., y Ruebush, B.K. (1960). *Anxiety in elementary school children*. New York: Wiley.
- Sarason, I.G. (1980). *Test anxiety: theory, research and applications*. Hillsdale: LEA.
- Sarason, I.G. (1984). Stress, anxiety and cognitive interference: reactions to tests. *Journal of Personality and Social Psychology*, 46, 929-938.
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, 4, 27-41.
- Stroud, A.H., y Secrest, D. (1966). *Gaussian quadrature formulas*. Englewood Cliffs: Prentice-Hall.
- Tanaka, J.S., y Huba, G.J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 38, 197-201.
- Williams, J.E. (1994). Anxiety measurement: construct validity and test performance. *Measurement and Evaluation in Counseling and Development*, 27, 302-307.

Aceptado el 15 de junio de 1998