

## Análisis psicométrico del *Cuestionario de confianza para hablar en público*, usando la teoría de respuesta al ítem

María D. Hidalgo\*, José A. López-Pina\*, Cándido J. Inglés\*\* y F. Xavier Méndez\*\*

\*Departamento de Psicología Básica y Metodología

\*\*Departamento de Personalidad, Evaluación y Tratamiento Psicológicos  
Universidad de Murcia

**Resumen:** Las respuestas de 799 sujetos a los 30 ítems de la adaptación española del Personal Report of Confidence as Speaker (PRCS; Paul 1966; Bados, 1986), fueron analizadas en el marco de la Teoría de Respuesta al Ítem con el modelo generalizado de crédito parcial (Muraki, 1992). Los resultados indicaron que los ítems cumplieron los supuestos de dicho modelo. Además, se presenta información acerca de la precisión del cuestionario en los diferentes niveles de ansiedad a hablar en público. Por último, se presentan las funciones de información para cada ítem con la intención de valorar la aportación de cada ítem al test en su conjunto.

**Palabras clave:** Miedo a hablar en público; teoría de respuesta al ítem; modelo generalizado de crédito parcial.

**Title:** Psychometric analysis of *Personal Report of Confidence as Speaker* using item response theory.

**Abstract:** The responses of 799 subjects to 30 items, of the Personal Report of Confidence as Speaker (PRCS; Paul, 1966; Bados, 1986) speech anxiety scale were analyzed in an item response framework, using polytomous models. The generalized partial credit model (Muraki, 1992) was applied. The results indicated that the items satisfy the assumptions of generalized partial credit model. Moreover, measurement precision across latent trait levels are presented. Finally, item information function in order to examine the contribution of each item on test information function are showed.

**Key words:** Speech anxiety; item response theory; Generalized Partial Credit Model.

La ansiedad para hablar en público ha sido conceptualizada, desde el enfoque cognitivo-conductual, como un conjunto de reacciones fisiológicas y cognitivas no adaptativas ante sucesos ambientales que dan lugar a un comportamiento ineficaz (Fremouw y Breitenstein, 1990).

Varios estudios epidemiológicos han puesto de manifiesto que la ansiedad y los miedos sociales son muy comunes en población adolescente comunitaria (Bados, 2001; Bell-Dolan, Last y Strauss, 1990; Inglés, Méndez e Hidalgo, 2000; Méndez, Inglés e Hidalgo, 2002), siendo el miedo a hablar en público una de las situaciones sociales más temidas durante la adolescencia. En la investigación de Strauss y Last (1993) el 57% de los adolescentes consideró el miedo a hablar en público como la segunda si-

tuación social más temida. Essau, Conradt y Petermann (1999) hallaron que el 31.1% de los adolescentes tuvo miedo a hacer una tarea delante de otros y el 19.7% informó experimentar un elevado grado de miedo a hablar en público. Inglés, Méndez e Hidalgo (1999) encontraron que el 45.75% de los adolescentes evitaban hablar en público.

La mayoría de autoinformes para evaluar el miedo a hablar en público proceden de la cultura anglosajona y fueron creados para población universitaria y adulta. Así, el *Cuestionario de Confianza para Hablar en Público* (Personal Report of Confidence as Speaker; PRCS) fue diseñado para medir las reacciones tanto afectivas como conductuales en aquellas situaciones en las que se debe hablar ante una audiencia. La versión original del PRCS (Gilkinson, 1942) se compuso de 104 ítems y fue usado profusamente entre 1940 y 1960, especialmente en las investigaciones relativas a la ansiedad para hablar en público, pero resultaba un instrumento demasiado largo para la práctica clínica. Aunque, posteriormente se desarrollaron varias

\* **Dirección para correspondencia:** María Dolores Hidalgo Montesinos. Departamento de Psicología Básica y Metodología. Facultad de Psicología. Universidad de Murcia. Apdo. 4021, 30008 Murcia (España). E-mail: mdhidalg@um.es

versiones abreviadas, la más utilizada ha sido la publicada por Paul (1966), bien como instrumento de tamización (Jones, Phillips y Rieger, 1995; Nickell, Witherspoon y Long, 1989; Paul, 1966), bien en el contexto de la investigación sobre la efectividad de varios tratamientos clínicos dirigidos a disminuir el miedo a hablar en público (Cradock, Cotler y Jason, 1978; García-López, 2000; Gatchel, 1979; Schuler, Gilner, Austrin y Davenport, 1982; Trussel, 1978; Weissberg, 1977). La versión abreviada de Paul (1966) consiste de 30 ítems, la mitad positivos y la mitad negativos, con el fin de controlar el efecto de aquiescencia. Los ítems son valorados mediante un formato de respuesta dicotómico (*verdadero-falso*). La adaptación española realizada por Bados (1986) sustituyó el formato de respuesta verdadero-falso de la escala original por una escala graduada de seis puntos (1 = *completamente de acuerdo*; 6 = *completamente en desacuerdo*) para cada uno de los 30 ítems, además omitió la referencia a la charla más reciente. Este instrumento ha demostrado una alta consistencia interna ( $\alpha$  de Cronbach de 0,91); y una adecuada fiabilidad test-retest (0.61), para un intervalo de seis meses (Paul, 1966). Méndez, Inglés e Hidalgo (1999) analizaron la consistencia interna y la estructura factorial del PRCS en una muestra de adolescentes de 16 a 18 años. La consistencia interna ( $\alpha$  de Cronbach) fue 0,91. La estructura factorial encontrada resultó ser unidimensional.

El objetivo del presente trabajo es mostrar la utilidad de la Teoría de Respuesta al Ítem (TRI) para el estudio y la evaluación de las propiedades psicométricas del PRCS.

### Consideraciones sobre la TRI

La TRI es una de las actuales teorías psicométricas que ha sido ampliamente utilizada en la construcción y evaluación de tests cognitivos y que miden habilidades superiores. Sin embargo, en los últimos años se ha demostrado su gran potencial en el campo de la evaluación psicológica y más concretamente en el de la evaluación clínica y de la personalidad. Así, numerosos son

los trabajos que analizan tests, autoinformes y escalas de personalidad bajo este nuevo prisma (Childs, Dahlstrom, Kemp y Pauter, 2000; Cooke, Kosson y Michie, 2001; Cooke y Michie, 1999; Krueger y Finger, 2001; Lange y Houran, 1999; Reise y Henson, 2000; Rouse, Finger y Butcher, 1999). Además, se han desarrollado un buen número de trabajos de divulgación de la TRI en estos ámbitos profesionales en los que se venía aplicando la TCT (Cella y Chang, 2000; Embretson y Prenovost, 1999; Embretson y Reise, 2000; Harvey y Hammer, 1999; Hays, Morales y Reise, 2000; Yen y Edwardson, 1999).

¿Qué características de esta aproximación psicométrica la hacen especialmente útil para la construcción y evaluación de tests psicológicos? En primer lugar, el ajuste de un modelo de TRI proporciona estimaciones invariantes de las propiedades psicométricas de los ítems, así como de las características psicológicas de los sujetos, es decir, que los parámetros que caracterizan al ítem no dependen de la muestra particular de sujetos utilizada y que los parámetros que caracterizan al sujeto no dependen de la muestra particular de ítems utilizada (Hambleton y Swaminathan, 1985; Lord, 1980). En segundo lugar, la TRI proporciona una descripción detallada del comportamiento individual de los ítems a lo largo de todos los niveles del rasgo, cuya representación gráfica se denomina Curva Característica del Ítem (CCI). En tercer lugar, permite obtener índices de precisión (fiabilidad) no sólo del test sino también del ítem y lo más importante, la precisión puede ser evaluada en cada uno de los niveles del rasgo psicológico. Para dicha finalidad, la TRI proporciona los conceptos de Función de Información del Ítem (FII) y Función de Información del Test (FIT) que serán las herramientas de trabajo para la selección de ítems y construcción de tests. Por último, esta teoría facilita la construcción de escalas y tests adaptados al sujeto, tanto para un nivel dado del rasgo que pretendemos medir como para todo el continuo del mismo. Asimismo, permite la construcción de tests adaptativos computerizados (Reise y Henson, 2000).

La TRI proporciona una amplia gama de modelos que permite trabajar con tests tanto unifactoriales como multifactoriales y con distintos formatos de respuesta (p.e., dicotómico, politómico o continuo); en Mellenbergh (1994) y Van der Linden y Hambleton (1997) encontramos una exposición detallada de estos modelos. En general, esta teoría asume que la relación entre la probabilidad de dar una respuesta positiva o correcta a un ítem y el nivel de habilidad latente ( $\theta$ ) o nivel del rasgo de un grupo de sujetos es una función de regresión no-lineal cuya representación gráfica se denomina CCI. Esta CCI se construye en todo el continuo del rasgo medido cuya escala se encuentra, normalmente, en el rango de  $-3$  a  $+3$ , con media cero y desviación típica de 1. Así, valores negativos y cercanos a  $-3$  indicarían niveles bajos en el rasgo, valores cercanos a cero bien positivos o bien negativos, indicarían niveles intermedios en el rasgo y por último, valores positivos y cercanos al extremo superior de la escala ( $+3$ ) indican que los evaluados tienen niveles altos en el rasgo. Atendiendo a las diferentes formas de la CCI encontramos distintos modelos de TRI. Los modelos más básicos son los referidos a ítems de respuesta dicotómica, de estos el modelo de 2-p (Birnbaum, 1968) ha sido uno de los más usados. La expresión general para este modelo es:

$$P(u_i = 1/\theta) = \frac{1}{1 + e^{-1.7 a_i(\theta - b_i)}}$$

donde la probabilidad (P) de que un sujeto con un nivel dado del rasgo latente ( $\theta$ ) dé una respuesta positiva a un ítem  $i$  ( $u_i=1$ ) es función de dos parámetros:  $a_i$  el parámetro de discriminación, que es proporcional a la pendiente de la CCI y  $b_i$  el parámetro de dificultad, también denominado de localización, que indica el nivel del rasgo requerido para que un sujeto tenga una probabilidad de 0.5 o mayor de dar una respuesta positiva al ítem. Cuando el parámetro de discriminación de un ítem es bajo, la probabilidad de una respuesta al mismo a lo largo del continuo del rasgo no varía, por lo que la capacidad del ítem para diferenciar a los sujetos es mucho menor que cuando el valor de discrimi-

nación del ítem es alto, es este caso el ítem permite diferenciar a los sujetos con niveles del rasgo cercanos al parámetro de localización ( $b$ ). Por otro lado, los ítems con valores altos de  $b$  serán respondidos positivamente sólo por aquellos sujetos con niveles altos en el rasgo, mientras que para aquellos ítems con valores bajos de  $b$  se espera que la mayoría de los sujetos den una respuesta afirmativa y que sólo unos pocos, precisamente aquellos con niveles muy bajos en el rasgo medido, den una respuesta negativa.

Para aquellos ítems cuyo formato de respuesta original es politómico, tales como los de respuesta graduada, en los últimos años se han desarrollado un extenso número de modelos. En términos muy generales estos modelos se dividen en dos clases. Por un lado estarían aquellos modelos que asumen cierto orden en las opciones de respuestas, lo más usual en tests de actitudes y personalidad. Por otro los que asumen que las respuestas son nominales, es decir, no es posible asumir a priori ningún orden entre las respuestas. Entre los primeros modelos encontramos, entre otros, el modelo de respuesta graduada de Samejima (1969) y el modelo de crédito parcial generalizado de Muraki (1992). Este último presenta una gran flexibilidad y potencial cuando trabajamos con ítems politómicos. El modelo se formula a partir del modelo logístico de 2-p, y se asume que para un ítem dado, la probabilidad de elegir una respuesta o categoría ( $b$ ) frente a la categoría inferior ( $b-1$ ) viene dada por:

$$P_{ih/h-1,h}(\theta) = \frac{P_{ih}(\theta)}{P_{i,h-1}(\theta) + P_{ih}(\theta)}$$

De este modo, la probabilidad de elegir como respuesta al ítem una categoría dada vendría dada por:

$$P_{ih}(\theta) = \frac{\exp\left[\sum_{v=1}^h Da_i(\theta - b_{iv})\right]}{\sum_{c=1}^{m_i} \exp\left[\sum_{v=1}^c Da_i(\theta - b_{iv})\right]}$$

donde  $m_i$  es el número de categorías de respuesta de un ítem,  $D$  es una constante de valor 1.7 que coloca la escala de  $\theta$  en la misma mé-

trica que el modelo de ogiva normal,  $a_i$  es el parámetro de inclinación del ítem,  $b_{ib}$  es el parámetro de categoría referido al ítem que se puede descomponer en dos parámetros  $b_i$  que es un parámetro de localización del ítem y  $d_b$  que es un parámetro de la categoría en sí misma, que se interpreta como la popularidad relativa de la categoría  $b$  en comparación con otras categorías dentro del mismo ítem, así  $b_{ib} = b_i - d_b$ . El parámetro  $a$  indica el grado en el que las respuestas a las categorías varían entre los ítems conforme el nivel en el rasgo cambia, este concepto es en cierto modo análogo al concepto de coeficiente de discriminación del ítem en el contexto del Modelo Clásico de Tests. Los  $b_{ib}$  parámetros que se estiman, el número total de categorías de respuesta al ítem menos uno, son aquellos niveles en la escala de  $\theta$  donde las curvas características de las categorías de respuesta interseccionan. Para comprender mejor estos parámetros consideremos un ítem con tres categorías de respuesta ("Nunca", "A veces", "Siempre"), cuyo parámetro  $a$  es mayor de cero, entonces si para un sujeto dado su nivel en el rasgo que mide el ítem es igual al parámetro  $b_{ib}$  de la categoría "A veces", el sujeto tiene la misma probabilidad de elegir esta categoría que de elegir la inmediatamente inferior, si por el contrario el nivel en el rasgo para este sujeto es mayor que el parámetro  $b_{ib}$  de la categoría "A veces", entonces el sujeto tiene mayor probabilidad de elegir la categoría "A veces" que la categoría "Nunca", y si el nivel en el rasgo es inferior que el parámetro de la categoría, entonces el sujeto tendrá una mayor probabilidad de elegir la categoría "Nunca" que la categoría "A veces".

## Método

### Sujetos

Tras realizar un muestreo aleatorio por conglomerados (representados por cada una de las zonas geográficas de la Región de Murcia), se seleccionó una muestra de 799 adolescentes, 396 varones y 403 mujeres, con edades com-

prendidas entre los 16 y 18 años (media=17.20; desviación típica=0.58). Todos los sujetos eran alumnos del Curso de Orientación Universitaria (C.O.U.) o del 1º Curso del Segundo Grado de Formación Profesional.

### Procedimiento

Para la aplicación de la prueba se llevó a cabo una entrevista con los jefes de los departamentos de orientación y/o con los jefes de estudios de los centros para explicar los objetivos de investigación, presentar el instrumento de evaluación que se iba a aplicar y solicitar su colaboración.

Una vez obtenido el correspondiente permiso, los sujetos completaron voluntariamente el cuestionario de forma colectiva en el aula durante la hora asignada a la actividad de tutoría. El entrevistador procedió a la entrega de los ejemplares que incluían las instrucciones y los ítems de la prueba; a continuación, leyó en voz alta las instrucciones, aclarando cualquier duda que surgiera, pero procurando no influir en la respuesta de los sujetos y advirtiéndoles la importancia de no dejar ningún ítem en blanco.

### Análisis preliminares

Los 30 ítems del PRCS fueron analizados bajo el Modelo Generalizado de Crédito Parcial (MGCP) propuesto por Muraki (1992). Los resultados de este primer análisis indicaron que el ítem 30 presentaba una correlación negativa con el rasgo medido por el resto de ítems y se procedió a su eliminación. De este modo se procedió a ajustar el MGCP a los 29 ítems restantes. Los resultados obtenidos revelaron que 10 de esos ítems presentaban un mal ajuste. El estudio del perfil de respuesta a cada ítem indicó que la distribución de frecuencias para algunos de estos ítems estaba sesgada, de tal modo que en las categorías de respuestas más extremas, bien superiores o inferiores, el porcentaje de sujetos que la habían seleccionado era pequeño (ver Tabla 1) con respecto a las categorías centrales. De este modo se decidió agrupar categorías, colapsando las mismas a tres, por lo

que las categorías de respuesta 1 y 2 se agrupan ahora en la categoría 1, las categorías de respuesta 3 y 4 en la categoría 2 y las 5 y 6 en la 3. Estos datos fueron de nuevo analizados bajo el Modelo Generalizado, siendo estos los resultados que se presentan en el siguiente apartado. Todo los análisis se realizaron con el programa PARSCALE v. 3 (Muraki y Bock, 1991).

**Tabla 1:** Distribución de frecuencias para cada uno de los 29 primeros ítems del PRCS.

Ítems	Categorías de respuesta					
	1	2	3	4	5	6
1	211	182	178	154	47	27
2	132	168	186	101	138	74
3	150	196	162	118	118	55
4	63	83	140	243	196	74
5	165	198	171	99	107	59
6	42	72	138	196	194	157
7	65	112	153	162	165	142
8	85	126	199	151	150	88
9	174	201	153	119	84	68
10	98	144	170	181	123	83
11	101	156	217	148	129	48
12	40	98	179	184	181	117
13	411	240	88	28	17	15
14	39	55	69	155	224	257
15	153	197	192	115	90	52
16	129	165	211	170	89	35
17	79	138	187	169	132	94
18	192	170	140	132	107	58
19	56	87	134	139	168	215
20	61	69	132	160	185	192
21	163	130	144	143	97	122
22	125	164	212	166	95	37
23	74	134	198	196	136	61
24	128	144	172	136	141	78
25	75	144	216	152	138	74
26	123	127	207	148	125	69
27	96	155	205	192	96	55
28	96	163	210	169	177	44
29	98	103	160	176	158	104

## Resultados

### Parámetros de los Ítems

En la Tabla 2 se presentan los parámetros estimados y sus errores típicos, bajo el modelo generalizado de Muraki (1992), para cada uno de los 29 ítems del PRCS. Los ítems se evalúan teniendo en cuenta su parámetro de discriminación ( $a$ ), su parámetro de localización ( $b$ ) y los parámetros umbrales de las categorías ( $d_k$ ), estos últimos se interpretan como la dificultad asociada a una categoría dada en comparación con otras categorías, o lo que es lo mismo la distancia (desviación) de cada categoría umbral de la localización del ítem ( $b$ ).

En general los ítems presentan valores de discriminación medios, siendo el promedio de todos ellos igual a 0.588 (D.T.=0.228). El ítem más discriminativo es el ítem 22 que hace referencia a “Mi mente está clara cuando me encuentro delante de un auditorio” ( $a=1.121$ ), junto a este los ítems 10 y 11 presentaron un valor de discriminación cercano a 1 (0.903) y (0.926), además tanto el ítem 12 y el ítem 16 presentaron valores de discriminación de 0.836 y 0.838, respectivamente. Los ítems menos discriminativos fueron el ítem 13 ( $a=0.251$ ) y el ítem 14 ( $a=0.255$ ), además los ítems 3, 4, 5, 7, 19 y 21 también presentaron valores del parámetro bajos.

Si el parámetro de discriminación es grande las CCI son muy discriminativas en un rango pequeño de valores de habilidad, pero poco discriminativos en el resto de valores de habilidad. Si el parámetro de discriminación es pequeño, los ítems discriminan de forma gradual en un amplio rango de valores del rasgo, las CCI son más achatadas. Así, con las CCI podemos comparar el funcionamiento de los ítems y determinar el rango del rasgo en el que el ítem es más discriminativo.

**Tabla 2:** Parámetro de discriminación (a), parámetro de localización (b) y parámetros de las categorías (d) para cada uno de los 29 ítems del PRCs

Item	Categoría			
	a	b	d <sub>2</sub>	d <sub>3</sub>
1	0.535 (0.042)	1.199 (0.099)	1.090 (0.088)	-1.090 (0.148)
2	0.470 (0.043)	0.278 (0.082)	0.458 (0.108)	-0.458 (0.119)
3	0.392 (0.040)	0.633 (0.109)	0.471 (0.125)	-0.471 (0.150)
4	0.305 (0.036)	-0.655 (0.118)	1.494 (0.191)	-1.494 (0.157)
5	0.320 (0.036)	0.831 (0.130)	0.408 (0.151)	-0.408 (0.185)
6	0.579 (0.042)	-0.784 (0.079)	0.865 (0.116)	-0.865 (0.084)
7	0.338 (0.033)	-0.560 (0.107)	0.763 (0.167)	-0.763 (0.143)
8	0.692 (0.050)	-0.083 (0.065)	0.746 (0.081)	-0.746 (0.078)
9	0.698 (0.049)	0.584 (0.070)	0.479 (0.073)	-0.479 (0.093)
10	0.903 (0.062)	0.083 (0.059)	0.702 (0.062)	-0.702 (0.065)
11	0.926 (0.062)	0.199 (0.060)	0.761 (0.059)	-0.761 (0.067)
12	0.836 (0.056)	-0.446 (0.061)	0.815 (0.078)	-0.815 (0.062)
13	0.251 (0.048)	3.854 (0.714)	-0.301 (0.239)	0.301 (0.471)
14	0.255 (0.038)	-2.075 (0.292)	0.315 (0.286)	-0.315 (0.190)
15	0.792 (0.051)	0.540 (0.064)	0.633 (0.065)	-0.633 (0.084)
16	0.838 (0.055)	0.484 (0.064)	0.911 (0.062)	-0.911 (0.082)
17	0.488 (0.034)	-0.036 (0.078)	0.882 (0.109)	-0.882 (0.108)
18	0.603 (0.049)	0.548 (0.076)	0.447 (0.084)	-0.447 (0.103)
19	0.370 (0.040)	-0.941 (0.121)	0.500 (0.168)	-0.500 (0.130)
20	0.435 (0.035)	-0.902 (0.102)	0.660 (0.147)	-0.660 (0.110)
21	0.307 (0.034)	0.317 (0.133)	0.462 (0.162)	-0.462 (0.175)
22	1.121 (0.074)	0.390 (0.056)	0.810 (0.049)	-0.810 (0.064)
23	0.562	0.034	1.040	-1.040

	(0.040)	(0.075)	(0.096)	(0.098)
24	0.479 (0.041)	0.168 (0.084)	0.586 (0.107)	-0.586 (0.114)
25	0.769 (0.054)	0.013 (0.065)	0.799 (0.072)	-0.799 (0.074)
26	0.796 (0.053)	0.144 (0.063)	0.746 (0.068)	-0.746 (0.074)
27	0.643 (0.041)	0.323 (0.073)	1.026 (0.080)	-1.026 (0.095)
28	0.606 (0.041)	0.315 (0.076)	0.959 (0.084)	-0.959 (0.098)
29	0.757 (0.054)	-0.170 (0.063)	0.673 (0.076)	-0.673 (0.071)

En cuanto a los valores del parámetro de localización se encontraron entre el valor mínimo de  $-2.075$  del ítem 14 “Me gusta observar las reacciones del auditorio ante mi charla” y el máximo de  $3.854$  del ítem 13 “Prefiero tener notas a mano por si olvido mi charla”, con un promedio de  $0.147$  (D.T.= $0.967$ ). En la figura 1 aparecen las Curvas Características de las Categorías del Ítem (CCCI) para el ítem 14 (ítem positivo); en este ítem podemos observar que independientemente del miedo para hablar en público, la categoría de respuesta más probable fue la 3, donde la mayoría de los adolescentes no le gusta observar las reacciones del auditorio ante la charla, sólo aquellos sujetos con niveles muy bajos en el rasgo, es decir, con poco miedo a hablar en público, manifiestan estar de acuerdo con el ítem. Por el contrario, las CCCIs para el ítem 13 (Figura 2) muestra que la categoría con más probabilidad de ser seleccionada fue la primera. En este caso los sujetos con niveles bajos y medios-altos en el rasgo manifiestan no necesitar notas a mano, siendo aquellos adolescentes con niveles muy altos de miedo los que puntúan más alto en el ítem. En la Tabla 1 vemos que la proporción de sujetos que seleccionaron las categorías de respuesta 1 y 2 del ítem 13 frente a las restantes fue muy alta.

Figura 1: Curvas Características de las Categorías para el ítem 14.

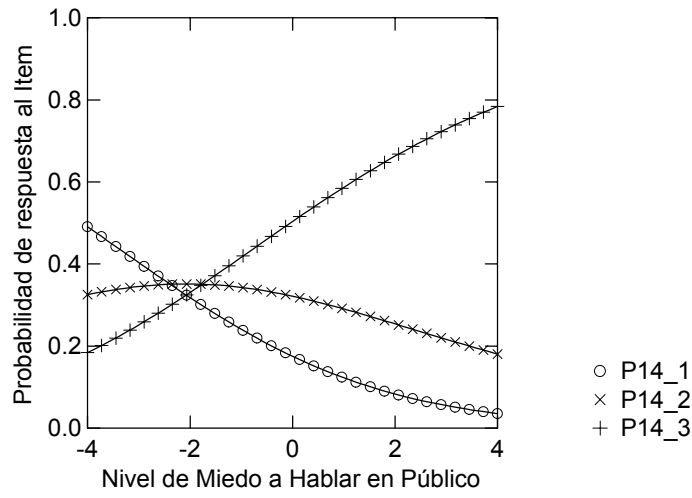
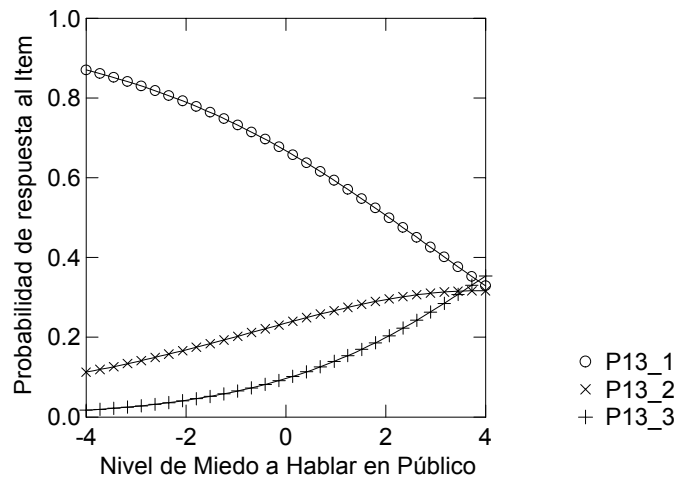


Figura 2: Curvas Características de las Categorías para el ítem 13.



Algunos ítems con valores de  $b$  bajos fueron el ítem 19 ( $b=-0.941$ ) y el ítem 20 ( $b=-0.902$ ). Sus CCCIs se presentan en las figuras 3 y 4, tal y como se puede observar son ítems de dificultad media baja, al ser valores cercanos al promedio, indica que la probabilidad de seleccionar la categoría 1 frente a las categorías 2 y 3 es mayor en los niveles bajos del rasgo que en

los altos, por el contrario en niveles altos del rasgo la probabilidad de seleccionar la categoría 3 es mayor que seleccionar la 1 y 2. En la figura 4, podemos observar que las CCCIs del ítem 20 son muy similares a las del ítem 19, aunque la pendiente de las mismas es mayor que en la figura 3, indicando que el ítem 20 es ligeramente más discriminativo que el ítem 19.

Figura 3: Curvas Características de las Categorías para el ítem 19.

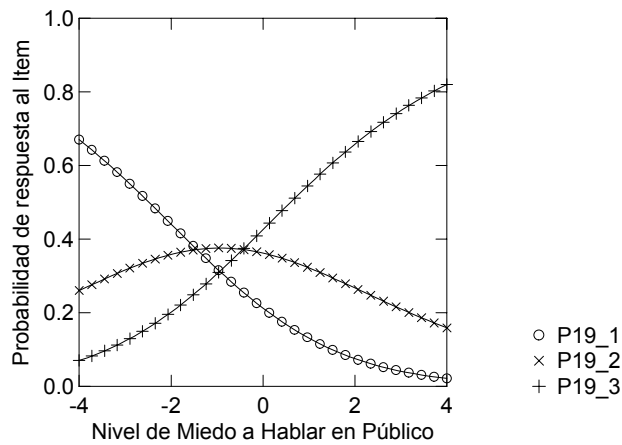
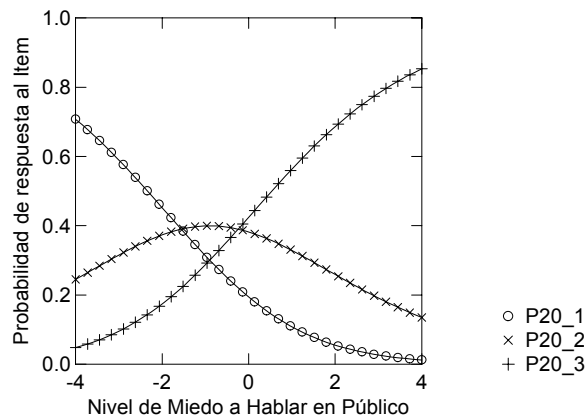


Figura 4: Curvas Características de las Categorías para el ítem 20.





Si trabajamos con el Modelo Clásico de Tests, se observa que los ítems 5 y 18 tienen la misma media, 2.952 para el ítem 5 y 2.957 para el ítem 18, sin embargo cuando se representan sus CCCIs estas son muy diferentes. En las figuras 5 y 6 aparecen representadas gráficamente las CCCIs de estos dos ítems, se observa que el ítem 18 es más discriminativo que el ítem 5, siendo las curvas de las categorías 1 y 3 más inclinadas que las respectivas del ítem 5, esto indica que el ítem 18 discrimina mejor entre la primera y la tercera categoría, que el ítem 5. Además, ambas figuras muestran que la CCCI de la primera y la tercera categoría dominan sobre la categoría intermedia, y que la CCCI de la categoría intermedia para el ítem 5 es más achatada que la del ítem 18, indicando que hay menos adolescentes que seleccionaron esta categoría

en el ítem 5 que en el ítem 18, pese a ser dos ítems con el mismo promedio.

### Estadísticos de Ajuste del Ítem

En la Tabla 3 se presentan los estadísticos  $G^2$  para cada ítem, sus grados de libertad asociados (g.l.) y el valor de probabilidad correspondiente ( $p$ ). De los 29 ítems ocho presentan valores de  $p \leq 0.01$ , indicando el posible mal ajuste de los mismos al modelo; estos ítems fueron el 1, 9, 17, 20, 21, 25, 26 y 28. Sin embargo, hay que tener en cuenta que el estadístico  $G^2$ , que se distribuye según una distribución ji-cuadrado, es sensible al número de intervalos en el que se divide el continuo de habilidad, así cuando el número de intervalos resulta elevado, los valores de este estadístico pueden incrementarse artificialmente (Muraki, 1997).

Figura 5: Curvas Características de las Categorías para el ítem 18.

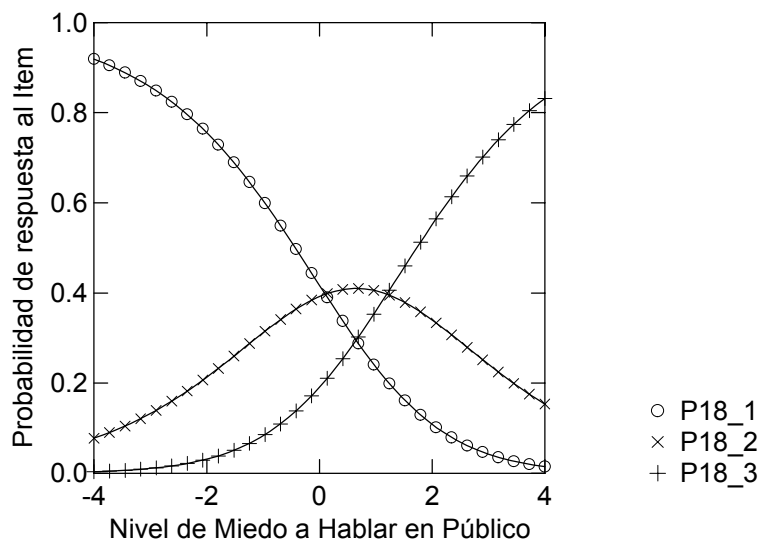
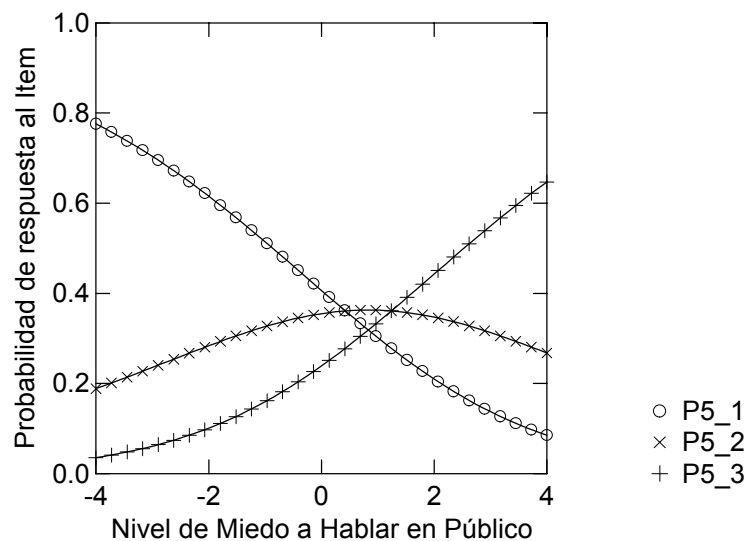


Figura 6: Curvas Características de las Categorías para el ítem 5.



El estadístico de bondad de ajuste global de modelo fue 1426.101 (g.l.=1080,  $p=0.000$ ); la razón entre  $G^2/g.l.$  fue igual a 1.32, indicando un buen ajuste global del modelo a los datos.

#### Función de Información de los Items y del Test

La Función de Información de un ítem (FII) muestra los niveles en el continuo del rasgo donde el ítem es más preciso, es decir, indica en qué niveles del rasgo el ítem es más informativo y por lo tanto más adecuado para estimar el nivel del rasgo. Además, permite combinar la información que proporcionan tanto los parámetros de localización como los parámetros de discriminación.

En las figuras 7 a la 12 se presentan la Función de Información del Ítem para cada uno de los 29 ítems del cuestionario.

Los ítems más informativos en niveles medios-altos del rasgo fueron el 1, 15, 16, 18, 22, 27 y 28, tal y como se puede observar en las figuras 7 a la 12. El ítem 1 (“Espero con ilusión una oportunidad de hablar en público”) fue el de máxima información en niveles muy altos de miedo a hablar en público. Los mejores ítems, considerando todo el continuo del rasgo, fueron el ítem 10 y el 22, la cantidad de información para el ítem 22 fue alta en valores altos del rasgo puesto que el parámetro de localización de este ítem fue ligeramente superior a la media y el parámetro de discriminación también fue elevado. El ítem siguiente de más información (ítem 10) mostró una FII más simétrica que el ítem 22, aportando más precisión en niveles bajos del rasgo.

Los ítems menos informativos, considerando todo el continuo del rasgo, fueron ítems tales como el 4, el ítem 5, el ítem 13, el ítem 14, y

el ítem 21, cuyas funciones de información fueron relativamente achatadas. Estos ítems son precisamente los menos discriminativos (parámetros  $a$  bajos). Los ítems 13 y 14 (figura 9) son los que presentaron la FII más plana, en estos ítems se combina el valor bajo de su parámetro de discriminación con parámetros para las categorías de mayor amplitud. Siendo ítems más informativos en un rango mínimo de valores del rasgo, en este caso solamente en los extremos de la distribución, bien en niveles muy bajos de ansiedad (ítem 14) o bien en niveles muy altos de ansiedad (ítem 13).

También se calculó la función de información del test para todos los ítems (FIT), este gráfico muestra los niveles del rasgo donde la escala completa es más discriminativa. Además, proporciona una medida global del error típico de medido (SEM) en cada nivel del rasgo, dado que la raíz cuadrada de la inversa de la función de información del test proporciona dicho estadístico. En otros términos, la FIT indica para cada nivel del rasgo latente, la cantidad de imprecisión esperada en la estimación de la puntuación para un sujeto. Por ejemplo, si tenemos un cuestionario para medir ansiedad social y este presenta un SEM bajo en el extremo superior del continuo del rasgo y por el contrario presenta un SEM alto en el extremo inferior del continuo, el clínico que utilice dicho cuestionario es conocedor que para puntuaciones altas en la escala éstas son estimaciones bastantes precisas del nivel de ansiedad social del sujeto, y por lo tanto el uso de este instrumento resulta apropiado para la detección de sujetos con niveles altos en el rasgo. Por otro lado, cuando dicho cuestionario se aplica a sujetos con niveles bajos en el rasgo, la estimación del nivel en el rasgo resultará imprecisa, el instrumento no será capaz de detectar diferencias entre sujetos con niveles medios en el rasgo y sujetos con niveles en el rasgo por debajo de la media. De esta forma, la FIT proporciona una guía para orientar la aplicación de un ins-

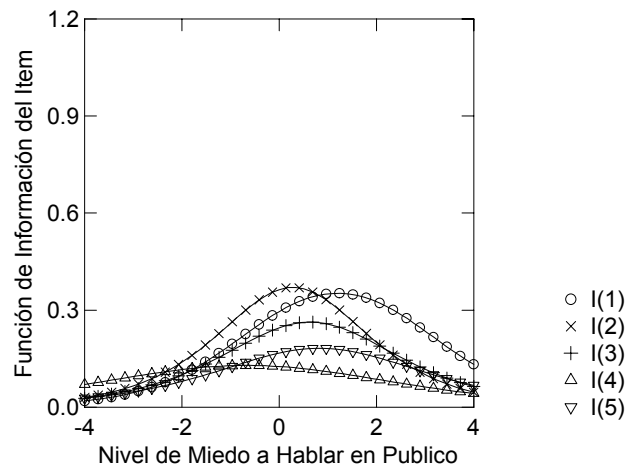
trumento dado, y decidir en qué situaciones y para que propósitos será más adecuado.

La FIT en la figura 13 muestra que el conjunto de 29 ítem producen una cantidad de información razonable entre  $-3$  y  $+3$ , aunque es más informativo en niveles intermedios de miedo a hablar en público y superiores e inferiores a la media del grupo en dos desviaciones típicas.

**Tabla 3:** Estadísticos de Ajuste del Ítem.

Item	G <sup>2</sup>	g.l.	p
1	69.84	45	0.010
2	25.72	38	0.936
3	57.47	42	0.056
4	36.14	38	0.556
5	22.78	44	0.997
6	31.28	33	0.553
7	60.26	39	0.016
8	31.69	35	0.629
9	67.72	38	0.002
10	27.63	34	0.772
11	53.59	35	0.023
12	27.12	33	0.755
13	56.74	39	0.033
14	32.17	30	0.359
15	56.04	38	0.030
16	39.63	36	0.311
17	76.99	38	0.000
18	51.79	38	0.067
19	49.93	37	0.076
20	57.98	35	0.009
21	67.27	43	0.010
22	41.13	34	0.186
23	36.73	38	0.528
24	46.73	38	0.156
25	62.27	36	0.004
26	62.73	35	0.003
27	55.52	38	0.033
28	87.08	38	0.000
29	34.15	35	0.509

**Figura 7:** Función de Información para los ítems 1, 2, 3, 4 y 5.



**Figura 8:** Función de Información para los ítems 6, 7, 8, 9 y 10.

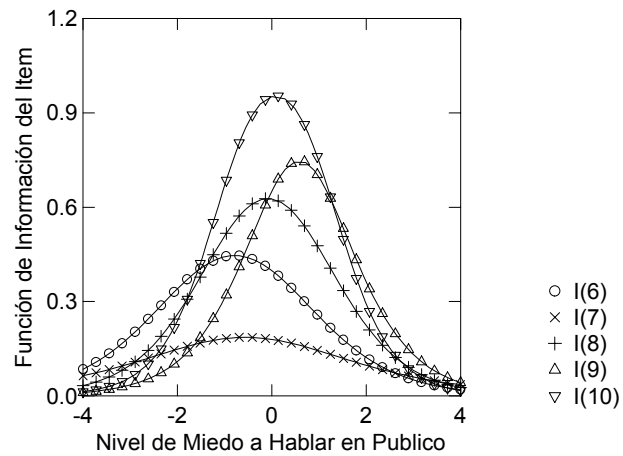


Figura 9: Función de Información para los ítems 11, 12, 13, 14 y 15.

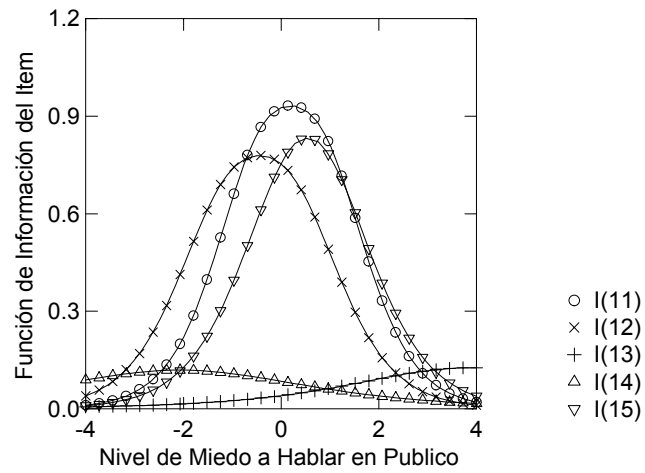


Figura 10: Función de Información para los ítems 16, 17, 18, 19 y 20.

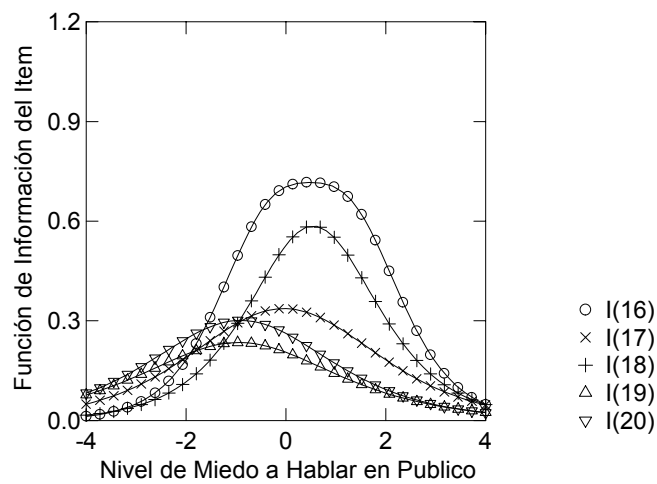


Figura 11: Función de Información para los ítems 21, 22, 23, 24 y 25.

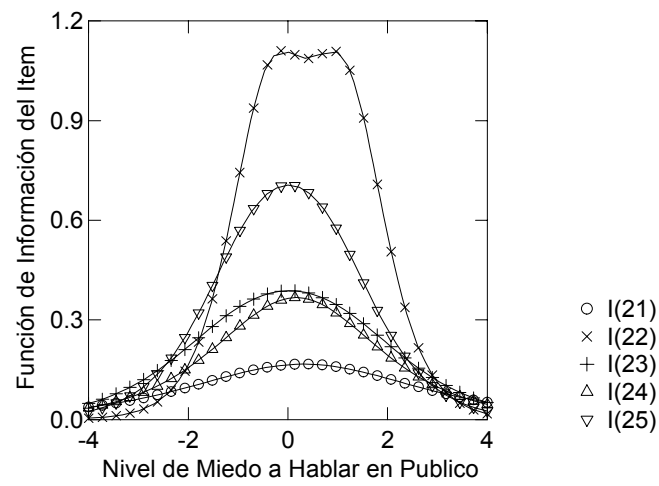


Figura 12: Función de Información para los ítems 26, 27, 28 y 29.

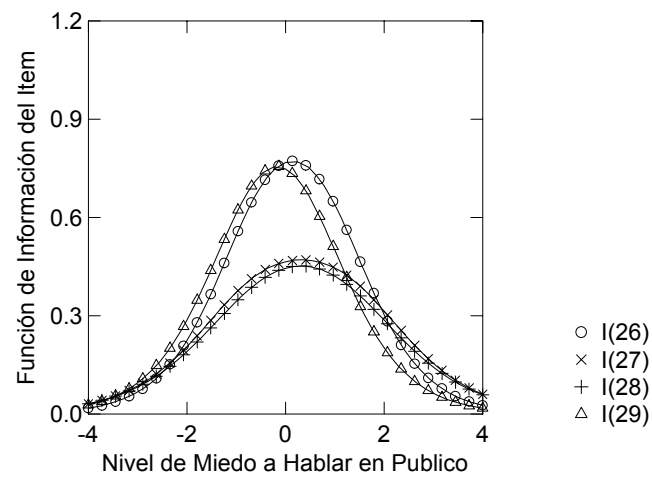
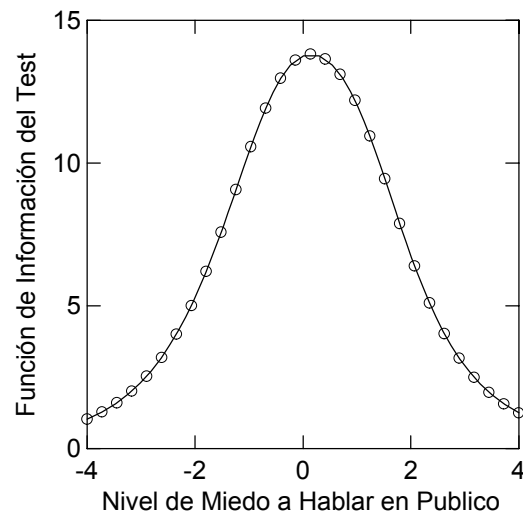


Figura 13: Función de Información del Test



## Conclusiones

El Cuestionario de Confianza para Hablar en Público mostró un ajuste adecuado al Modelo de Crédito Parcial Generalizado, sólo 8 de los 29 ítems presentaban un mal ajuste al modelo, de los cuales seis, en concreto los ítems 1, 9, 20, 25, 26 y 28, fueron ítems que resultaron informativos. Tanto el ítem 1 como el ítem 28 fueron informativos en niveles medios-altos del rasgo; y los ítems 9, 20, 26 y 25 resultaron útiles para la toma de decisiones tanto en niveles medios del rasgo como en niveles medios-bajos o medios-altos, sus FIIs muestran valores apropiados para considerar la permanencia de estos ítems en el cuestionario. Por otro lado, el ítem 17 y el 21 presentaron mal ajuste siendo poco informativos, además el análisis de su contenido informó de la inadecuación relativa del mismo. Así, el enunciado del ítem 17

(“Aunque no me gusta hablar en público no le tengo particular temor”) incluye una doble negación, teniendo además en cuenta que este es un ítem positivo, es decir que a mayor puntuación en este ítem, mayor grado de desacuerdo con la sentencia, mayor miedo a hablar en público; todo parece indicar que el ítem presenta problemas de ajuste por la redacción/traducción, y se tendría que reformular para su inclusión en sucesivas aplicaciones del cuestionario. En cuanto al ítem 21 (“Me gusta preparar una charla”), consideramos también que habría que reformularlo, el contenido del mismo hace referencia a una actividad que bien puede no resultar una actividad habitual para un adolescente, si el término “charla” se sustituyera por otros tales como “exposición” o “intervención”, el ítem sería más propio para la evaluación del miedo a hablar en público en adolescentes.

Por otro lado, el análisis tanto de las funciones de información de los ítems como del test, reveló que el test de 29 ítems fue apropiado para diferenciar a sujetos con niveles intermedios-altos de miedo a hablar en público, la función de información del test fue más elevada en ese rango de puntuaciones, indicando una mayor precisión en la estimación del miedo a hablar en público para estos adolescentes.

En general, los resultados de este trabajo señalan la utilidad de la TRI para el análisis de instrumentos de evaluación psicológica, así como su potencial en la construcción y adaptación de cuestionarios. Por último, otra ventaja derivada de utilizar la TRI se refiere a la posibilidad de construir tests adaptados al sujeto, es decir, los ítems se pueden seleccionar para cada sujeto en función de su nivel en el rasgo, y en este caso encontramos que tests con un número menor de ítems son más precisos que tests

con un tamaño de ítems fijo mayor, la regla psicométrica clásica que a mayor número de ítems mayor fiabilidad resulta obsoleta. El uso de tests adaptativos presupone que sujetos con diferentes niveles del rasgo se les administra diferentes ítems, pero sus puntuaciones se interpretan del mismo modo y por lo tanto se pueden comparar directamente. Los modelos de medida desarrollados en el marco de la Teoría de Respuesta al Ítem ofrecen la posibilidad de construir este tipo de tests, sin embargo, estos modelos en su versión politómica son poco conocidos por los profesionales de la psicología y, en general, han sido implementados en menor número de tests. En este trabajo se ajusta uno de los modelos más prometedores, Modelo de Crédito Parcial Generalizado, en el ámbito de la evaluación de la personalidad y el diagnóstico clínico.

## Referencias

- Bados, A. (1986). *Análisis de componentes de un tratamiento cognitivo-somático-conductual del miedo a hablar en público*. Tesis doctoral no publicada, Universidad de Barcelona, Barcelona.
- Bados, A. (2001). *Fobia social*. Madrid: Síntesis.
- Bell-Dolan, D. J., Last, C. G. y Strauss, C. C. (1990). Symptoms of anxiety disorders in normal children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 29, 759-765.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F.M. Lord y M. Novick, *Statistical theories of mental scores*. Reading, MA: Addison-Wesley.
- Cella, D. y Chang, C.H. (2000). A discussion of item response theory and its applications in health status assessment. *Medical Care*, 38 (Suppl. 9), II66-II72.
- Childs, R.A., Dahlstrom, W.G., Kemp, S.M. y Pauter, A.T. (2000). Item response theory in personality assessment: A demonstration using the MMPI-2 Depression Scale. *Assessment*, 7, 37-54.
- Cooke, D.J. y Michie, C. (1999). Psychopathy Across Cultures: North America and Scotland Compared. *Journal of Abnormal Psychology*, 108, 58-68.
- Cooke, D.J., Kosson, D. y Michie, C. (2001). Psychopathy and ethnicity: Structural, item, and test generalizability of the Psychopathy Checklist-Revised (PCL-R) in Caucasian and African American Participants. *Psychological Assessment*, 13, 531-542.
- Cradock, C., Cotler, S. y Jason, L. A. (1978). Primary prevention: Immunization of children for speech anxiety. *Cognitive Therapy and Research*, 2, 389-396.
- Embretson, S.E. y Prenovost, L.K. (1999). Item response theory in assessment research. En P.C. Kendall, J.N. Butcher, y G.N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (2<sup>nd</sup> ed.). New York, NY: John Wiley & Sons.
- Embretson, S.E. y Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ, US: Lawrence Erlbaum Associates.
- Essau, C. A., Conradt, J. y Petermann, F. (1999). Frequency and comorbidity of social phobia and social fears in adolescents. *Behaviour Research and Therapy*, 37, 831-843.
- Fremouw, W.J. y Breitenstein, J.L. (1990). Speech anxiety. En H. Leitenberg (Ed.), *Handbook of social and evaluation anxiety* (pp. 412-431). Nueva York: Plenum Press.
- García-López, L.J. (2000). *Un estudio de la eficacia entre tres modalidades de tratamiento para población adolescente con fobia social*. Tesis doctoral no publicada, Universidad de Murcia, Murcia.
- Gatchel, R. J. (1979). Comparison of heart rate biofeedback, false, biofeedback, and systematic desensitization in reducing speech anxiety: Short and long-term effectiveness. *Journal of Consulting and Clinical Psychology*, 47, 620-622.
- Gilkinson, H. (1942). Social fears as reported by students in college speech classes. *Speech Monographs*, 9, 141-160.
- Hambleton, R.K. y Swaminathan, J. (1985). *Item response theory: principles and applications*. Boston, MA: Kluwer-Nijhoff Publishing.



- Harvey, R.J. y Hammer, A.L. (1999). Item response theory. *Counseling Psychologist*, 27, 353-383.
- Hays, R.D., Morales, L.S. y Reise, S.P. (2000). Item response theory and health outcomes measurement in the 21<sup>st</sup> century. *Medical Care*, 38 (suppl.9), II28-II42.
- Inglés, C.J., Méndez, F.X. e Hidalgo, M.D. (1999, Abril). *Estudio sobre la habilidad para hablar en público en población española interuniversitaria*. Comunicación presentada en el I Symposium sobre Habilidades Sociales: Técnicas y Áreas de Aplicación, Granada, España.
- Inglés, C. J., Méndez, F. X. e Hidalgo, M. D. (2000). Cuestionario de Evaluación de Dificultades Interpersonales en la Adolescencia. *Psicothema*, 12, 390-398.
- Jones, G., Phillips, G. y Rieger, E. (1995). Effects of cardiac awareness, public speaking anxiety, and physiological "priming" on state-trait anxiety (state) and feelings of apprehension. *Psychophysiology*, 32, 43.
- Krueger, R.F. y Finger, M.S. (2001). Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. *Psychological Assessment*, 13, 140-151.
- Lange, R. y Houran, J. (1999). Scaling MacDonald's AT-20 using item-response theory. *Personality and Individual Differences*, 26, 467-475.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mellenbergh, G.J. (1994). Generalized linear item response theory. *Psychological Bulletin*, 115, 300-307.
- Méndez, F. X., Inglés, C. J. e Hidalgo, M. D. (1999). Propiedades psicométricas del Cuestionario de Confianza para Hablar en Público: Estudio con una muestra de alumnos de enseñanzas medias. *Psicothema*, 11, 65-74.
- Méndez, F.X., Inglés, C.J. e Hidalgo, M.D. (2002). Estrés en las relaciones interpersonales: Un estudio descriptivo en la adolescencia. *Ansiedad y Estrés*, 8, 23-31.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Muraki, E. (1997). A generalized Partial Credit Model. En W.J. Van der Linden y R.K. Hambleton (1997). *Handbook of modern item response theory*. New York: Springer Verlag.
- Muraki, E. y Bock, R.D. (1991). *PARSCALE: Parameter Scaling of Rating Data*. [Computer Program]. Chicago, IL: Scientific Software, Inc.
- Nickell, E. B., Witherspoon, A. D. y Long, C. K. (1989). Decreasing speech anxiety of female prisoners. *Psychological Reports*, 65, 1351-1357.
- Paul, G. L. (1966). *Insight vs desensitization in psychotherapy*. Stanford, CA: Stanford University Press.
- Reise, S.P. y Henson, J.M. (2000). Computerization and adaptive administration of the NEO PI-R. *Assessment*, 7, 347-364.
- Rouse, S.V., Finger, M.S. y Butcher, J.N. (1999). Advances in Clinical Personality Measurement: An Item Response Theory Scale of the MMPI-2 PSY-5 Scales. *Journal of Personality Assessment*, 72, 282-307.
- Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometric Monograph*, N° 17.
- Schuler, K., Gilner, F., Austrin, H. y Davenport, D. G. (1982). Contribution of the education phase to stress-inoculation training. *Psychological Reports*, 51, 611-617.
- Strauss, C. C. y Last, C. G. (1993). Social and simple phobias in children. *Journal of Anxiety Disorders*, 7, 141-152.
- Trussell, R. (1978). Use of graduated behavior rehearsal, feedback, and systematic desensitization for speech anxiety. *Journal of Counseling Psychology*, 25, 14-20.
- Van der Linden, W.J. y Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer Verlag.
- Weissberg, M. (1977). A comparison of direct and vicarious treatments of speech anxiety: Desensitization, desensitization with coping imagery and cognitive modification. *Behavior Therapy*, 8, 606-620.
- Yen, M. y Edwardson, S.R. (1999). Item response theory approach in scale development. *Nursing Research*, 48, 234-238.

(Artículo recibido: 17-6-2002, aceptado: 12-7-2002)

