

FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM EN EL CUESTIONARIO DE CONFIANZA PARA HABLAR EN PÚBLICO

M^º DOLORES HIDALGO MONTESINOS, CÁNDIDO JOSÉ INGLÉS SAURA
Y FRANCISCO XAVIER MÉNDEZ CARRILLO

Dpto. de Psicología Básica y Metodología. Facultad de Psicología. Universidad de Murcia
Apartado 4021 - 30080 Murcia (España)

Tel: (968) 363470 - Fax: (968) 364115 - E-mail: mdhidalg@um.es

Resumen

Este trabajo tiene como objetivo estudiar el posible funcionamiento diferencial de los ítems (DIF) que componen el Cuestionario de Confianza para Hablar en Público (PRCS; Paul, 1966). Las respuestas de 799 sujetos a los 30 ítems de la escala PRCS, de evaluación de la ansiedad a hablar en público, fueron analizadas en el marco de la teoría de respuesta al ítem bajo el modelo logístico de dos parámetros. Este estudio implementa dos procedimientos de detección del DIF basados en la Teoría de Respuesta al Ítem (estadístico de Lord y medidas exactas de área de Raju), para examinar si los ítems del PRCS funcionan de forma distinta en función del sexo.

Palabras clave: Funcionamiento diferencial del ítem, Miedo a hablar en público, PRCS, Sesgo del ítem, Teoría de respuesta al ítem.

Abstract

The purpose of this study was to determine the differential item functioning (DIF) of the Personal Report of Confidence as Speaker (PRCS; Paul, 1966). The responses of 799 subjects to 30 items of the PRCS speech anxiety scale were analyzed in an item response theory framework, using two parameter logistic model. This study used two different DIF detection procedures derived from item response theory (Lord's test and Raju's exact area measures) to evaluate the extent to which items in PRCS functioned differently for matched gender groups.

Key words: Differential item functioning, Item bias, Item Response Theory, PRCS, Speaking public fear.

Varios estudios epidemiológicos han puesto de manifiesto que la ansiedad y los miedos sociales son muy comunes en población adolescente comunitaria (Bell-Dolan, Last y Strauss, 1990). El miedo a hablar en público es una de las situaciones sociales más temidas durante la adolescencia (Argyle, 1981; Harris y Brown, 1982). En la investigación de Strauss y Last (1993) el 57% de los adolescentes consideró el miedo a hablar en público como la segunda situación 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 dificultad para hablar en público puede originar o agravar problemas en esta etapa evolutiva. En primer lugar, la escasa participación en el aula y la tendencia a no preguntar al profesor, pueden dar lugar al abandono de los estudios y/o a un bajo rendimiento escolar (Beidel, 1991; Francis & Radka, 1995). En segundo lugar, la baja tasa de reforzamiento social consecuencia de la evitación social puede originar depresión (Francis, Last y Strauss, 1992; Vaughn y Lancelotta, 1990).

Finalmente, los déficit de habilidades sociales concomitantes aumentan el riesgo de consumir drogas legales e ilegales (Clark y Kirisci, 1996; Clark y Sayette, 1993). En un estudio con 1035 sujetos de 12 a 17 años, Essau et al. (1999) encontraron que el 42% de los adolescentes diagnosticados de ansiedad social presentaron trastornos somatoformes, el 29.4% trastornos depresivos y el 23.5% abuso de sustancias. Por otro lado, el estudio realizado por Inglés, Hidalgo y Méndez (2001) con 538 adolescentes de 12 a 18 años, reveló que los sujetos con miedo a hablar en público presentaron mayores dificultades interpersonales en las áreas de aserción, relaciones heterosexuales, relaciones familiares y relaciones con amigos íntimos que el grupo control.

La conceptualización del miedo a hablar en público no está exenta de dificultades. Se han propuesto numerosas definiciones, aunque el mayor interés lo han mostrado los investigadores de orientación conductual (Méndez, Inglés e Hidalgo, 1999). Así, Fremouw y Breitenstein (1990) definen la ansiedad para hablar en público como un conjunto de reacciones fisiológicas y cognitivas no adaptativas ante sucesos ambientales que dan lugar a un comportamiento ineficaz. Bados (1990) distingue dos tipos de problemas: déficit de habilidades para intervenir en público y/o presencia de ansiedad social. Por su parte, Montorio, Fernández, Lázaro y López (1996) mantienen que ambos conceptos, déficit de habilidades y ansiedad social, deben entenderse como un continuo a lo largo del cual se sitúa cada persona.

La evaluación del miedo a hablar en público es un proceso complejo en el cual se emplean distintos procedimientos de evaluación como entrevistas, autoinformes, autorregistros, tests de observación conductual y registros psicofisiológicos. Sin embargo, las medidas de autoinforme son las más empleadas por su bajo coste y fácil aplicación. Desde el punto de vista de la medición psicológica, la objetividad de la medida realizada es un requisito imprescindible para cualquier autoinforme o escala psicológica, lo que implica estudiar no sólo la precisión de la misma sino también su validez. Uno de los problemas asociados con la medida psicológica, que más interés ha generado en los últimos 25 años, ha sido el posible sesgo de los tests. La investigación en este campo intenta determinar si características tales como raza y género, que pueden ser ajenas al rendimiento de los sujetos en una tarea cognitiva o psicológica, pueden tener un efecto sobre la medida de una variable psicológica. En instrumentos de medida que juegan cada vez más un papel prominente en tareas tales como el diagnóstico de la conducta disfuncional de miedo a hablar en público, la identificación de poblaciones de riesgo y la asignación a programas de tratamiento, es necesario explorar y comprender cómo las características demográficas de los sujetos (género, edad,...) y su bagaje cultural y social, intervienen artificialmente en el proceso de medida de variables psicológicas. En otras palabras, realizar estudios para comprobar la ausencia de lo que se denomina funcionamiento diferencial del ítem (DIF). Se dice que un ítem presenta DIF cuando dos grupos comparables de sujetos, es decir, con un nivel idéntico respecto a la variable o atributo medido por el test, lo responden o ejecutan de manera distinta.

Probablemente por su carácter pionero el Personal Report of Confidence as Speaker (PRCS; Paul, 1966) ha sido uno de los autoinformes más utilizados para evaluar el miedo a hablar en público. Este autoinforme fue diseñado para medir las reacciones tanto afectivas como conductuales en aquellas situaciones en las que se debe hablar ante un auditorio, es decir, para medir el miedo a hablar en público. Consta de 30 ítems, la mitad positivos (por ejemplo, "espero con ilusión la oportunidad de hablar en público") y la mitad negativos (por ejemplo, "siempre que me es posible, evito hablar en público"), con el fin de controlar el efecto de aquiescencia. La adaptación española de este instrumento fue realizada por Bados (1986), quien lo denominó Cuestionario de Confianza para Hablar en Público. En esta adaptación se omitió la referencia a la charla más reciente y sustituyó el formato de respuesta verdadero-falso por una escala Likert de seis puntos (1 = completamente de acuerdo; 6 = completamente en desacuerdo). Su consistencia interna es elevada en población americana (coeficiente alfa=0.91) (Daly, 1978; Paul, 1966) y española (coeficiente alfa=0.906) (Méndez et al., 1999).

Por otro lado, los hallazgos sobre diferencias sexuales en ajuste social son inconsistentes. Mientras que algunos estudios hallan mayores niveles de ansiedad social/miedo a hablar en público entre las adolescentes (Clark, Turner, Beidel, Donovan, Kirisci y Jacob, 1994; Inderbitzen, Walters y Bukowski, 1997; Inglés, Méndez e Hidalgo, 2000; La Greca y Lopez, 1998; Méndez, Inglés e Hidalgo, 2001; Olivares, García-López, Hidalgo, Turner y Beidel, 1999), otros estudios no encuentran diferencias estadísticamente significativas, tanto en muestras adolescentes anglo-parlantes (Connor, Dann y Twentyman, 1982; Essau et al., 1999; Furham y Gunter, 1983; Ollendick, Matson y Hessel, 1985; Phillips, Jones, Rieger y Snell, 1997), como hispano-parlantes (Inglés et al., 1999; Méndez et al., 1999; Méndez, Hidalgo e Inglés, en prensa). Con respecto a la fobia social el DSM-IV resalta la disparidad de resultados indicando que "los estudios epidemiológicos sugieren que la fobia social es más frecuente en las mujeres; sin embargo, cuando se estudian poblaciones clínicas, ambos sexos se muestran igualmente propensos, o bien se constata que el trastorno incide más frecuentemente en los varones" (APA, 1994, p. 424).

Teniendo en cuenta los hallazgos anteriormente comentados, interesa conocer si el Cuestionario de Confianza para Hablar en público refleja diferencias reales en función del género, en los comportamientos que trata de evaluar, o por el contrario, si estas diferencias son un artefacto provocado por un funcionamiento diferente de algunos de los ítems en función de la variable género (DIF).

El objetivo de este trabajo es evaluar la posible presencia de DIF en el Cuestionario de Confianza para Hablar en Público y distinguirlo del posible impacto (diferencias en el test debidas a diferencias reales en miedo a hablar en público).

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 comprendidas entre 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

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 advirtiendo la importancia de no dejar ningún ítem en blanco.

Análisis del DIF

La evaluación del DIF se abordó utilizando dos procedimientos diferentes basados en los modelos de respuesta al ítem: el estadístico de Lord (1980) y las medidas de área exactas de Raju (1988, 1990).

Estadístico de Lord.

Lord (1980) propone una prueba χ^2 para someter a comprobación estadística la hipótesis de igualdad de los parámetros de los ítems del grupo focal y de referencia. Esta prueba viene dada por:

$$CHI - LORD = V' S^{-1} V$$

donde V es el vector de diferencias entre los parámetros estimados para un ítem en el grupo de referencia y los parámetros estimados para ese mismo ítem en el grupo focal. Bajo el modelo logístico de 2-p, V' vendría dado por:

$$V' = (\hat{b}_R - \hat{b}_F, \hat{a}_R - \hat{a}_F)$$

y S^{-1} , es la inversa de la matriz de varianza-covarianza asintótica para los vectores de diferencias entre parámetros.

El estadístico propuesto por Lord, bajo la hipótesis nula de no DIF y en el modelo de 2-p, sigue una distribución χ^2 con dos grados de libertad ($gl=2$). Un ítem presenta DIF si el valor observado de CHI-LORD es mayor que el valor teórico asociado al χ^2 ($gl=2$) al nivel de significación establecido.

Medidas de Area Exactas

En Teoría de Respuesta al Ítem (TRI) se han propuesto un conjunto de medidas basadas en la estimación del área entre las Curvas Características del Ítem (CCI) del grupo focal y del grupo de referencia. Raju (1988, 1990) propone dos medidas de área exactas (con signo y absoluta) que permiten probar la significación estadística del área encerrada entre la CCI del grupo de referencia y la CCI del grupo focal. La expresión general de estas medidas, basadas en la integración continua viene dada por:

$$A_i = \int_{-\infty}^{\infty} f[P_R(\theta) - P_F(\theta)]d(\theta)$$

donde $P_R(\theta)$ y $P_F(\theta)$ representan las probabilidades de responder correctamente al ítem i en cada uno de los grupos referencia (R) y (F). La función f puede especificarse con signo (ESA) y en términos absolutos (EUA). La expresión general de la ecuación anterior adopta distintas formas según el modelo de TRI con el que se esté trabajando, y si se cumplen o no ciertas condiciones en los parámetros de los ítems. Para el modelo de 2-p Raju (1988, 1990) define el área con signo como:

$$ESA = (\hat{b}_{iF} - \hat{b}_{iR})$$

y el área absoluta como,

$$EUA = |\hat{b}_{iF} - \hat{b}_{iR}|$$

cuando los parámetros de discriminación del grupo de referencia y del grupo focal son iguales. Cuando los parámetros de discriminación en el grupo de referencia son distintos de los del grupo focal, el área absoluta (H) se obtiene según:

$$H = \frac{2(\hat{a}_{iF} - \hat{a}_{iR})}{D\hat{a}_{iF}\hat{a}_{iR}} \ln \left(1 + \exp \left(\frac{D\hat{a}_{iF}\hat{a}_{iR}(\hat{b}_{iF} - \hat{b}_{iR})}{\hat{a}_{iF} - \hat{a}_{iR}} \right) \right) - (\hat{b}_{iF} - \hat{b}_{iR})$$

donde D es una constante de valor 1.702.

Para probar la significación estadística de estas medidas de área, que se distribuyen normalmente, Raju (1990) propone una prueba Z. En un nivel de confianza prefijado por el investigador, si el valor Z observado es mayor o igual que el valor teórico correspondiente a la distribución normal tipificada, se rechaza la hipótesis nula de ausencia de DIF, en caso contrario se acepta la hipótesis nula.

Tanto el estadístico de Lord como las medidas exactas de área fueron calculadas utilizando el programa IRTDIF (Kim y Cohen, 1992) en el intervalo de habilidad de -4.0 a +4.0.

Previo a la aplicación de cualquiera de las medidas de anteriores, el primer paso a realizar es ajustar un modelo de TRI, por supuesto, seleccionando aquel que proporcione el mejor ajuste, donde se requiere la comprobación de la dimensionalidad del instrumento. Esta fase supone la estimación tanto de los parámetros del ítem como de los parámetros de habilidad, para el grupo de referencia y para el grupo focal por separado, obteniendo además la estimación de las varianzas y covarianzas de los parámetros de los ítems. En un segundo momento, las métricas de los grupos focal y de referencia deben igualarse, es decir, emplear algún procedimiento de equiparación de los parámetros de los ítems y sus matrices de varianza-covarianza que nos permitan transformar, mediante una transformación lineal, los parámetros de un grupo a la misma escala que los parámetros de otro grupo. En definitiva se trata de encontrar los valores A (pendiente) y K (intercepción) tal que los valores originales de un grupo (grupo focal) queden transformados a la misma métrica que los valores del otro grupo (grupo de referencia), las transformaciones serían:

$$a_F = a_F^* / A$$

$$Var(a_F) = Var(a_F^*) / A^2$$

$$b_F = Ab_F^* + K$$

$$Var(b_F) = A^2 Var(b_F^*)$$

donde * indica las estimaciones no transformadas.

Resultados

Análisis de la dimensionalidad

Se realizó un análisis factorial común de ejes principales para explorar la estructura subyacente al cuestionario, como paso previo al ajuste de un modelo de TRI. Los resultados del análisis factorial exploratorio apuntan la existencia de un solo factor que da cuenta del 21.06% de la varianza total (autovalor=6.317) frente al segundo factor que explica sólo un 4.25% de la varianza total (autovalor=1.276). Aunque el porcentaje de varianza explicado por el primer factor no es excesivamente alto (menos del 30% que se acepta comúnmente), hay que considerar que tanto la versión original del cuestionario (Paul, 1966) como la adaptación realizada por Méndez et al. (1999) con ítems de formato de respuesta polítonico, reportaron una estructura unidimensional.

Análisis descriptivo

La media total de la muestra fue de 16.33 y la desviación típica (DT) de 6.80. El rango de puntuaciones fue de 0 a 30. La media para el subgrupo de varones fue de 15.01 (DT=6.74) y para

el subgrupo de mujeres de 17.64 (DT=6.62). Se llevó a cabo una prueba t de diferencias entre medias, encontrándose diferencias significativas en función del sexo ($t_{97}=-5.558$; $p=0.000$), donde las mujeres presentan una mayor dificultad para hablar en público.

El valor de consistencia interna para el cuestionario (coeficiente alfa de Cronbach) fue de 0.882, para el subgrupo de varones fue de 0.876 y para el de mujeres de 0.879. El promedio de las correlaciones ítem-test fue de 0.463 (DT=0.114) para el grupo de varones, donde el valor mínimo fue de 0.199 y el máximo de 0.641. Para el grupo de mujeres el rango de correlaciones ítem-test se encontró entre 0.094 y 0.609, con un promedio de 0.467 y una desviación típica de 0.116.

Estimación de parámetros

De entre los modelos de TRI unidimensionales para ítems de formato dicotómico, se trabajó con el modelo de 2-p, dado que en principio no es posible asumir homogeneidad de los índices de discriminación de los ítems (igualdad de las correlaciones ítem-test para todo el cuestionario). La estimación de parámetros se realizó, por separado para el grupo de referencia y el grupo focal, vía el programa BILOG versión 3.04 (Mislevy y Bock, 1990), que implementa el algoritmo de estimación de MVMP (Máxima Verosimilitud Marginal a Posteriori).

La tabla 1 presenta los parámetros estimados (a y b), sus errores típicos (ET) y el valor del estadístico de ajuste del ítem (CHI-Q) y su probabilidad asociada (p) para cada uno de los ítems del Cuestionario y por separado para el grupo de varones (grupo de referencia) y el grupo de mujeres (grupo focal). De los 30 ítems, sólo el ítem 17 para el grupo de varones presenta un mal ajuste (CHI-Q=24.5, $p=0.0002$). El ajuste global del cuestionario en el grupo de varones fue CHI-GLOBAL₁₄₇=1329.2 ($p=0.6648$) y para el grupo de mujeres fue CHI-GLOBAL₂₁₁=179.7 ($p=0.9422$).

Igualación de parámetros

Las constantes de igualación A y K se obtuvieron implementando el procedimiento de equiparación de curvas características desarrollado por Stocking y Lord (1983). Para el cálculo de las mismas se utilizó el programa EQUATE versión 2.0 (Baker, 1993) que implementa dicho método. Los valores obtenidos fueron $A=1.0308$ y $K=0.4006$. Con estos valores se procedió a la transformación de los parámetros del grupo focal en la métrica del grupo de referencia. El promedio de los parámetros de discriminación transformados para el grupo focal fue 0.716 (DT=0.208) y para los parámetros de dificultad fue de -0.055 (DT=1.102).

Análisis del DIF

Estadístico de Lord.

La tabla 2 presenta los valores del estadístico de Lord para cada uno de los ítems del cuestionario de confianza para habla en público. De los treinta ítems cuatro reportaron valores de CHI-LORD estadísticamente significativos, indicando un posible funcionamiento diferencial de estos ítems.

En concreto, los ítems que presentaron parámetros diferentes para el grupo de varones y el de mujeres fueron el ítem 2 ($p=0.0132$), el 3 ($p=0.0351$), el 4 ($p=0.0368$) y el 20 ($p=0.0277$). Estos cuatro ítems resultaron estadísticamente significativos al 5%, lo que supone que el 13.33% de los ítems del Cuestionario presentaron DIF. Si trabajamos con un nivel de significación más restrictivo (1%) encontramos que ninguno de los ítems presentarían DIF.

Las figuras 1 a la 4 representan las curvas características para cada uno de los ítems en los que el estadístico de Lord detectó la presencia de DIF. Tal y como se puede observar en las mismas, los ítems 2, 3 y 20 presentan DIF uniforme, donde las CCI del grupo de varones y de mujeres no se cruzan, y el ítem 4 DIF no-uniforme. El ítem 2 y 3 favorece a las mujeres, la probabilidad de dar una respuesta positiva al ítem es mayor en las mujeres que en los varones.

Tabla 1. Parámetros estimados y errores típicos para el grupo de varones y para el grupo de mujeres.

Ítem	a _V (ET)	b _V (ET)	CHI-Q(V) (p)	a _M (ET)	b _M (ET)	CHI-Q(M) (p)
1	0.688 (0.089)	-0.711 (0.122)	2.9 (0.7138)	0.864 (0.106)	-1.087 (0.123)	8.3 (0.2171)
2	0.719 (0.094)	-0.045 (0.095)	3.8 (0.5766)	0.671 (0.090)	-0.905 (0.136)	2.3 (0.9412)
3	0.533 (0.077)	-0.273 (0.126)	4.5 (0.4783)	0.516 (0.080)	-1.195 (0.200)	2.8 (0.9455)
4	0.363 (0.065)	1.061 (0.243)	8.7 (0.1910)	0.569 (0.080)	0.723 (0.139)	4.8 (0.8513)
5	0.484 (0.072)	-0.841 (0.174)	10.5 (0.0622)	0.445 (0.072)	-1.155 (0.217)	5.5 (0.7062)
6	0.598 (0.090)	1.105 (0.170)	3.7 (0.7153)	0.777 (0.106)	0.660 (0.108)	12.0 (0.0611)
7	0.400 (0.068)	0.614 (0.172)	8.2 (0.2223)	0.500 (0.075)	0.470 (0.135)	2.8 (0.9720)
8	0.701 (0.096)	0.204 (0.098)	4.2 (0.5226)	0.708 (0.093)	-0.268 (0.099)	3.0 (0.8867)
9	0.951 (0.116)	-0.312 (0.083)	3.6 (0.6060)	1.105 (0.133)	-0.791 (0.091)	10.1 (0.1201)
10	0.940 (0.111)	0.217 (0.078)	2.7 (0.6043)	1.022 (0.124)	-0.258 (0.077)	0.7 (0.9936)
11	1.021 (0.116)	-0.092 (0.076)	6.6 (0.2512)	1.211 (0.141)	-0.469 (0.070)	6.7 (0.3534)
12	0.981 (0.121)	0.715 (0.096)	6.0 (0.1103)	0.937 (0.119)	0.106 (0.077)	8.0 (0.3351)
13	0.433 (0.090)	-3.348 (0.648)	1.1 (0.8932)	0.372 (0.093)	-4.759 (1.186)	4.0 (0.2584)
14	0.378 (0.076)	2.367 (0.460)	2.2 (0.9035)	0.630 (0.100)	1.511 (0.207)	3.2 (0.8714)
15	0.974 (0.128)	-0.529 (0.092)	6.0 (0.1980)	0.805 (0.103)	-0.836 (0.125)	11.4 (0.1215)
16	0.852 (0.108)	-0.273 (0.086)	3.0 (0.5689)	0.870 (0.113)	-0.710 (0.105)	1.3 (0.9726)
17	0.627 (0.075)	0.311 (0.114)	24.5 (0.0002)*	0.705 (0.083)	-0.290 (0.097)	12.7 (0.0794)
18	0.808 (0.102)	-0.282 (0.093)	2.0 (0.8523)	0.771 (0.100)	-0.732 (0.114)	4.3 (0.7469)
19	0.547 (0.086)	0.968 (0.164)	4.0 (0.6745)	0.617 (0.087)	0.623 (0.123)	9.7 (0.3759)
20	0.513 (0.074)	0.939 (0.167)	4.9 (0.5522)	0.595 (0.087)	0.895 (0.149)	7.1 (0.5229)
21	0.475 (0.070)	-0.012 (0.133)	1.2 (0.9480)	0.615 (0.086)	-0.394 (0.118)	5.9 (0.6633)
22	1.197 (0.155)	-0.141 (0.068)	6.3 (0.0947)	1.159 (0.150)	-0.664 (0.087)	6.5 (0.3671)
23	0.654 (0.089)	0.212 (0.103)	2.7 (0.7536)	0.705 (0.093)	-0.225 (0.102)	7.8 (0.4529)
24	0.574 (0.077)	-0.057 (0.114)	1.5 (0.9143)	0.548 (0.081)	-0.471 (0.135)	10.6 (0.2237)
25	1.023 (0.118)	-0.048 (0.074)	1.9 (0.7581)	0.833 (0.110)	-0.237 (0.087)	7.6 (0.3653)
26	0.922 (0.111)	-0.099 (0.080)	1.0 (0.9607)	0.709 (0.092)	-0.452 (0.107)	5.9 (0.5511)
27	0.711 (0.092)	-0.187 (0.099)	2.5 (0.7800)	0.934 (0.117)	-0.310 (0.083)	0.9 (0.9947)
28	0.693 (0.094)	-0.324 (0.104)	5.2 (0.3981)	0.573 (0.079)	-0.455 (0.128)	5.5 (0.7034)
29	0.817 (0.109)	0.361 (0.088)	2.4 (0.7901)	0.899 (0.113)	0.070 (0.079)	2.2 (0.9445)
30	0.552 (0.080)	-1.018 (0.167)	1.4 (0.9284)	0.473 (0.079)	-1.641 (0.272)	6.2 (0.4060)

Tabla 2. Valores del Estadístico de Lord

Item	CHI-LORD	Item	CHI-LORD
1	1.953	16	0.197
2	8.651*	17	2.057
3	6.698*	18	0.286
4	6.605*	19	0.731
5	0.769	20	7.172*
6	1.814	21	1.306
7	3.476	22	1.644
8	0.322	23	0.165
9	2.084	24	0.148
10	0.721	25	5.868
11	0.748	26	3.217
12	3.559	27	4.800
13	0.911	28	5.729
14	4.613	29	1.036
15	3.164	30	0.813

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

El ítem 20 favorece a los varones, siendo la probabilidad de dar una respuesta positiva mayor en los varones que en las mujeres a lo largo de todo el continuo de habilidad. Por último, en el ítem 4 la probabilidad de dar una respuesta positiva es mayor para los varones que para las mujeres en niveles bajos del rasgo (mayor confianza para hablar en público). Por el contrario, en los niveles altos del rasgo (mayor miedo a hablar en público), las mujeres tienen más probabilidad de dar una respuesta positiva al ítem que los varones.

Figura 1. Curvas Características estimadas del ítem 2 para el grupo de referencia (PR2) y el grupo focal (PF2).

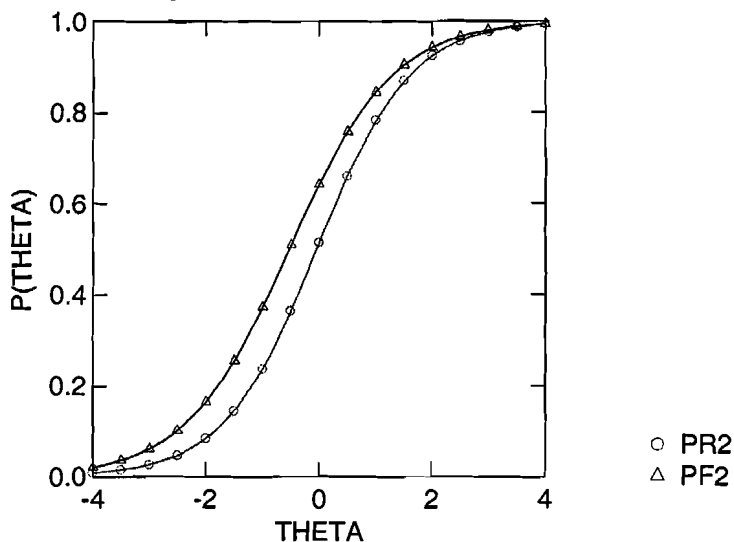


Figura 2. Curvas Características estimadas del ítem 3 para el grupo de referencia (PR3) y el grupo focal (PF3).

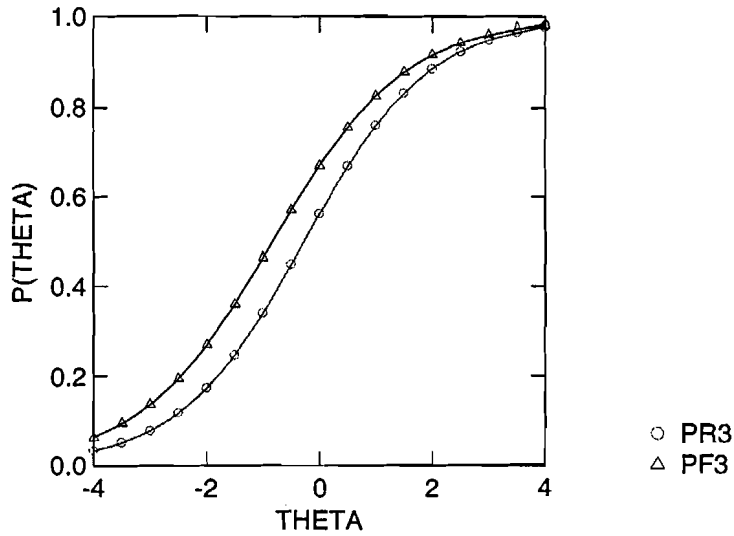


Figura 3. Curvas Características estimadas del ítem 4 para el grupo de referencia (PR4) y el grupo focal (PF4).

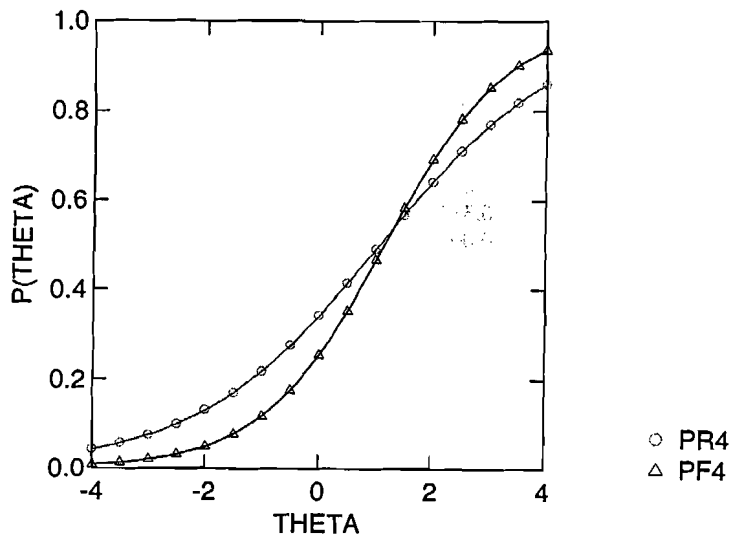


Figura 4. Curvas Características estimadas del ítem 20 para el grupo de referencia (PR20) y el grupo focal (PF20).

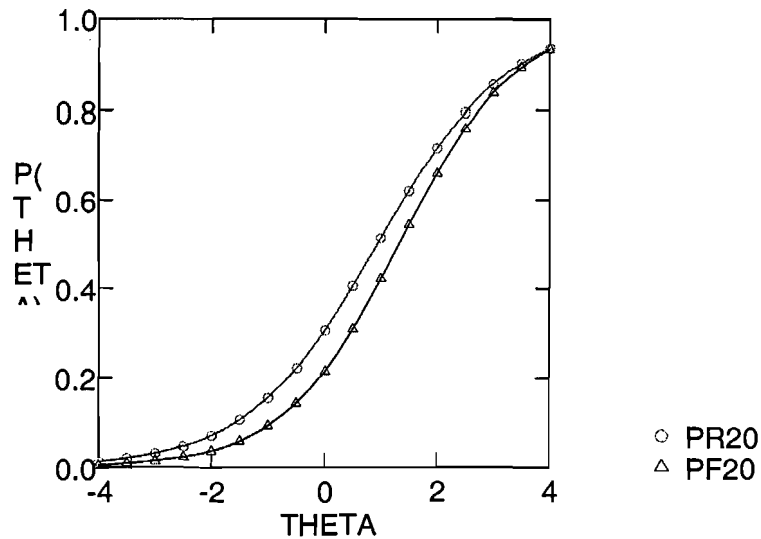


Tabla 3. Valores de las medidas de área exactas con signo.

Item	ESA	VAR(ESA)	Z(ESA)
1	-0.009	0.031	-0.053
2	-0.487	0.029	-2.869**
3	-0.558	0.058	-2.313*
4	0.085	0.080	0.300
5	0.051	0.080	0.179
6	-0.024	0.041	-0.120
7	0.271	0.049	1.222
8	-0.079	0.020	-0.562
9	-0.103	0.016	-0.819
10	-0.081	0.012	-0.732
11	0.009	0.011	0.084
12	-2.058	0.015	-1.659
13	-1.157	1.913	-0.836
14	-0.409	0.258	-0.805
15	0.068	0.025	0.427
16	-0.059	0.019	-0.425
17	-0.210	0.023	-1.384
18	-0.072	0.022	-0.477
19	0.075	0.043	0.362
20	0.384	0.051	1.691
21	0.007	0.033	0.038
22	-0.143	0.013	-1.268
23	-0.044	0.022	-0.298
24	-0.027	0.032	-0.152
25	0.204	0.014	1.755
26	0.034	0.018	0.249
27	0.268	0.017	2.044*
28	0.255	0.028	1.521
29	0.112	0.014	0.934
30	-0.273	0.106	-0.836

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Medidas exactas de área.

La tabla 3 presenta los valores de la medida exacta de área con signo (ESA), su varianza asociada (VAR(ESA)) y el valor Z observado para cada uno de los ítems del PRCS. Del total de ítems del cuestionario, un 10% de los mismos se detectaron como DIF (al nivel de confianza del 95%) y un 3.33% (al nivel de confianza del 99%). En concreto al 95%, estos ítems fueron el 2 ($p=0.0041$), 3 ($p=0.0207$) y 27 ($p=0.0409$), y al 99% sólo el ítem 2.

En la tabla 4 se presentan los valores de la medida exacta de área absoluta (EUA y H), sus varianzas asociadas (VAR(EUA) y VAR(H)), y el valor observado Z. Cuando se utilizó la medida exacta de área absoluta (Z(H)), un 20% de los ítems fueron detectados como DIF al nivel de significación del 5% y un 3.33% al nivel alfa del 0.01

Tabla 4. Valores de las medidas exactas de área absolutas

Ítem	EUA	VAR(EUA)	H	VAR(H)	Z(H)
1	0.211	0.013	0.211	0.036	1.105
2	0.488	0.011	-0.488	0.030	-2.838**
3	0.558	0.021	-0.558	0.059	-2.306*
4	0.772	0.068	0.772	0.188	1.781
5	0.208	0.046	-0.208	0.126	-0.585
6	0.281	0.025	0.281	0.068	1.082
7	0.423	0.039	0.423	0.107	1.294
8	0.080	0.007	-0.080	0.020	-0.567
9	0.132	0.003	0.132	0.009	1.390
10	0.086	0.004	0.086	0.010	0.846
11	0.105	0.005	0.105	0.015	0.851
12	0.207	0.005	-0.207	0.014	-1.753
13	1.164	0.726	-1.164	1.997	-0.824
14	0.893	0.129	0.893	0.354	1.500
15	0.214	0.008	-0.214	0.022	-1.451
16	0.059	0.007	-0.059	0.019	-0.424
17	0.220	0.009	0.220	0.025	1.392
18	0.102	0.014	-0.102	0.040	-0.514
19	0.144	0.020	0.144	0.054	0.619
20	0.396	0.010	0.396	0.028	2.379*
21	0.351	0.037	0.351	0.101	1.104
22	0.143	0.005	-0.143	0.013	-1.246
23	0.067	0.014	0.067	0.038	0.344
24	0.114	0.033	-0.114	0.090	-0.381
25	0.276	0.006	-0.276	0.016	-2.194*
26	0.302	0.012	-0.302	0.033	-1.655
27	0.339	0.011	0.339	0.029	1.990*
28	0.365	0.012	-0.365	0.034	-1.992*
29	0.120	0.005	0.120	0.014	0.995
30	0.380	0.068	-0.380	0.187	-0.877

Los ítems detectados con funcionamiento diferencial fueron el 2 ($p=0.0045$), 3 ($p=0.0211$), 20 ($p=0.0173$), 25 ($p=0.0283$), 27 ($p=0.0465$) y 28 ($p=0.0464$).

Las figuras 5 a la 7 representan las curvas características para cada uno de los ítems en los que bien la medida exacta de área con signo o absoluta detectaron, adicionalmente al estadístico de Lord, como sesgados. Para la primera de las medidas el ítem fue el número 27 y para la medida absoluta estos ítems fueron el 25 y el 28. Si observamos las curvas características que describen estos tres ítems tanto para el grupo de varones como para el de mujeres, comprobamos como el DIF se presenta no uniforme, donde las CCI del grupo focal y de referencia se cruzan en un

nivel de habilidad a lo largo del continuo. Así, para el ítem 25 la probabilidad de dar una respuesta positiva es mayor para los varones que para las mujeres en niveles altos del rasgo y similar para varones y mujeres en niveles del rasgo bajos (DIF no uniforme asimétrico). Por el contrario, para el ítem 27 la probabilidad de dar una respuesta positiva es mayor en varones en los niveles bajos del rasgo, y similar para ambos subgrupos en niveles altos de miedo a hablar en público. Por último, para el ítem 28, las chicas presentan una probabilidad más alta de dar una respuesta positiva en niveles altos del rasgo y similar a la de los chicos para niveles bajos del rasgo.

Figura 5. Curvas Características estimadas del ítem 25 para el grupo de referencia (PR25) y el grupo focal (PF25).

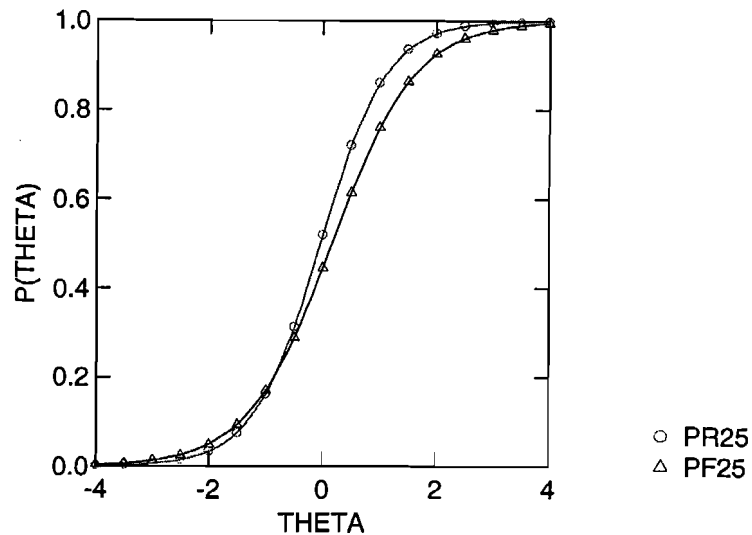


Figura 6. Curvas Características estimadas del ítem 27 para el grupo de referencia (PR27) y el grupo focal (PF27).

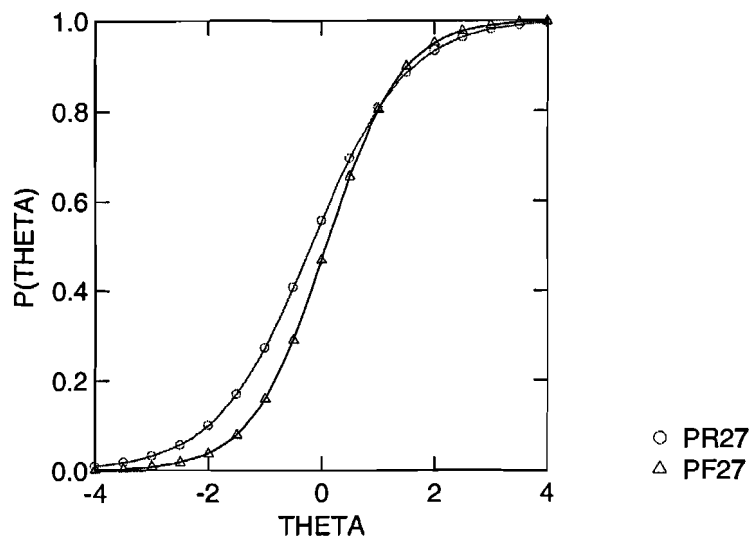
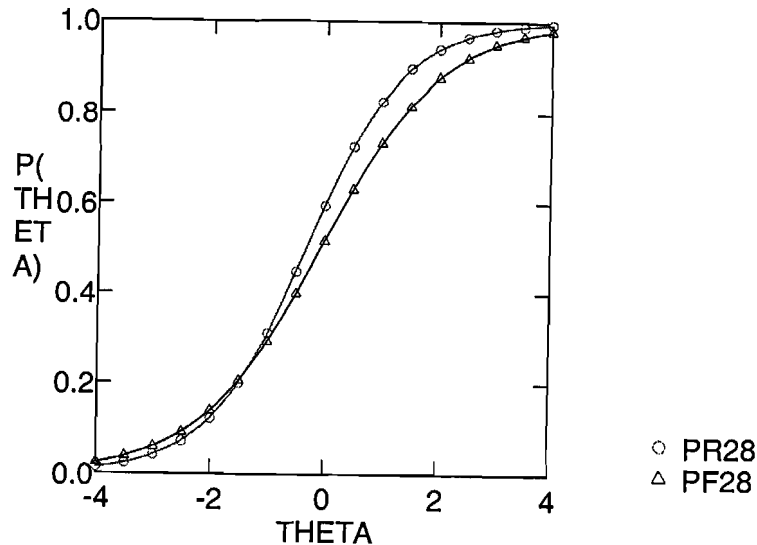


Figura 7. Curvas Características estimadas del ítem 28 para el grupo de referencia (PR28) y el grupo focal (PF28).



Acuerdo entre procedimientos

El acuerdo entre los tres procedimientos fue alto. El número de ítems que fueron detectados con DIF por los tres procedimientos empleados (Lord, Z(ESA) y Z(H)) fue de 2 y el número de ítems detectados como no DIF fue de 23 (al nivel de significación del 5%). El porcentaje de acuerdo entre los tres procedimientos fue del 83.33%. Por otro lado, si consideramos el acuerdo entre estadísticos considerados dos a dos, encontramos que el estadístico de Lord y Z(ESA) concordaron en el 90% de los ítems, mientras que el acuerdo del estadístico de Lord con Z(H) fue ligeramente menor (86.67%). Por último, tanto Z(ESA) como Z(H) presentaron un acuerdo del 90%. Si consideramos un nivel de significación del 0.01, encontramos que el acuerdo entre Z(ESA) y Z(H) fue del 100% y que el acuerdo entre estas medidas y el estadístico de Lord fue del 96.67%. Hay que tener en cuenta que tanto el estadístico de Lord como Z(H) presentan mayor capacidad para detectar el DIF no uniforme (Hidalgo y López, 1997), y que en general, el estadístico de Lord es de los tres procedimientos aquí empleados el que resulta más efectivo en la identificación del DIF (Cohen y Kim, 1993; Hidalgo y López, 1997).

Conclusiones

A pesar que en la actualidad existe un gran número de procedimientos cada vez más sofisticados para detectar el DIF en sus diversas formas (Gómez e Hidalgo, 1997; Hidalgo y Gómez, 1999; Millsap y Everson, 1993), es preciso tener en cuenta que estas técnicas sólo son apropiadas para detectar el sesgo potencial en un ítem y no ofrecen por sí mismas una explicación de las causas del DIF (Camilli, 1993). Por ello, una correcta interpretación del DIF surge de la conjunción entre los procedimientos estadísticos y las revisiones teóricas o juicios de expertos (Shepard, 1981; Tittle, 1982). Para afirmar que un ítem favorece a un determinado grupo sobre otro, hay que hacer referencia a las causas o razones por las que funciona de forma distinta y si esas diferencias son o no parte legítima del rasgo objeto de medición. En definitiva, examinar el contenido de los ítems en los cuales se detectó DIF, e intentar dar una explicación del mismo.

Para los ítems que presentaron un funcionamiento diferencial en esta muestra de sujetos adolescentes, encontramos que sólo dos de los ítems (el ítem 2 y el ítem 3) fueron detectados como DIF por las tres técnicas estadísticas utilizadas. El contenido del ítem 2 se refiere a "Me tiemblan las manos cuando trato de coger algo estando en la tarima" y el enunciado del ítem 3 a "Tengo un miedo constante de olvidar lo que voy a decir", ambos enunciados hacen referencia a aspectos negativos de la situación de hablar en público, pero un análisis más detenido del enunciado de estos ítems no nos proporciona ningún tipo de evidencia para afirmar el sesgo de los mismos; analizado el contenido de los mismos resulta difícil explicar porqué las chicas adolescentes muestran sistemáticamente una probabilidad mayor que los chicos de responder afirmativamente a esos dos ítems. En definitiva se requiere un análisis más detallado de estos ítems, aunque hay que tener en cuenta que los procedimientos de detección del DIF suelen producir falsos positivos, es decir, detectan ítems como DIF pese a ser ítems no sesgados. Otros dos ítems, el ítem 20 y el 27 fueron detectados con DIF por dos de los tres procedimientos. El ítem 20 le pregunta al sujeto sobre "Me siento a disgusto conmigo mismo(a) después de intentar dirigir la palabra o dar una charla a un grupo de personas" y el ítem 27 "Encuentro ligeramente agradable la perspectiva de hablar en público" (ítem positivo, que se ha invertido), en ambos ítems la probabilidad de dar una respuesta positiva fue mayor en los chicos que en las chicas. Del mismo modo que los otros dos ítems, no se encuentra una explicación teórica del posible sesgo de estos ítems. En resumen, los resultados encontrados parecen apoyar la idea de que este instrumento no funciona diferencialmente en función de la variable género, y que existe la posibilidad de que los ítems que han sido identificados como DIF, tanto por el estadístico de Lord como por las medidas de área, sean falsas identificaciones.

Referencias

- Argyle, M. (1981). The contribution of social interaction research to social skills training. En J. D. Wine y M. D. Snye (Eds.). *Social competence* (pp. 261-286). New York: Guildford Press.
- Asociación Psiquiátrica Americana-APA (1994). Manual diagnóstico y estadístico de los trastornos mentales (4ª ed.). Barcelona: Masson.
- 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 (España).
- Bados, A. (1990). Afrontamiento y prevención del estrés: Intervención sobre las dificultades para hablar en público. En J. M. Buceta y A. M. Bueno (Eds.). *Psicología y salud: Control de estrés y trastornos asociados* (pp. 63-84). Madrid: Dykinson.
- Baker, F.B. (1993). *EQUATE 2.0: A computer program for the characteristic curve method of IRT equating*. [Computer Program] Madison WI: University of Wisconsin. Laboratory of Experimental Design.
- Beidel, D. C. (1991). Social phobia and overanxious disorder in school-age children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30, 545-552.
- 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.
- Camilli, G. (1993). The case against item bias detection techniques based on internal criteria: Do item bias procedure obscure test fairness issues. En P. Holland y H. Wainer (Eds.), *Differential item functioning*. Hillsdale, NJ: LEA.
- Clark, D. B. y Kirisci, L. (1996). Posttraumatic stress disorder, depression, alcohol use and quality of life in adolescents. *Anxiety*, 2, 226-233.
- Clark, D. B. y Sayette, M. A. (1993). Anxiety and the development of alcoholism: Clinical and scientific issues. *American Journal of Addiction*, 2, 59-76.

- Clark, D. B., Turner, S. M., Beidel, D. C., Donovan, J. E., Kirisci, L. y Jacob, R. G. (1994). Reliability and validity of the Social Phobia and Anxiety Inventory for Adolescents. *Psychological Assessment*, 6, 135-140.
- Cohen, A.S. y Kim, S.H. (1993). A comparison of Lord's χ^2 and Raju's area measures in detection of DIF. *Applied Psychological Measurement*, 17, 32-52.
- Connor, J. M., Dann, L. N. y Twentyman, C. T. (1982). A self-report measure of assertiveness in young adolescents. *Journal of Clinical Psychology*, 38, 101-106.
- Daly, J. A. (1978). The assessment of social-communicative anxiety via self-reports: A comparison of measures. *Communication Monographs*, 45, 204-218.
- 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.
- Francis, G., Last, C. G. y Strauss, C. C. (1992). Avoidant disorder and social phobia in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 1086-1089.
- Francis, G. y Radka, D. F. (1995). Social anxiety in children and adolescents. In M. B. Stein (Ed.). *Social phobia: Clinical and research perspectives* (pp. 119-143). New York: Guilford Press.
- Fremouw, W. J. y Breitenstein, J. L. (1990). Speech anxiety. En H. Leitenberg (Ed.). *Handbook of social and evaluation anxiety*. Nueva York: Plenum Press.
- Furnham, A. y Gunter, B. (1983). Sex and personality differences in self-reported social skills among British adolescents. *Journal of Adolescence*, 6, 57-69.
- Gómez, J. e Hidalgo, M.D. (1997). Evaluación del funcionamiento diferencial en ítems dicotómicos: Una revisión metodológica. *Anuario de Psicología*, 74, 3-32.
- Harris, K. R. y Brown, R. D. (1982). Cognitive behavior modification and informed teacher treatment shy children. *Journal of Experimental Education*, 50, 137-143.
- Hidalgo, M.D. y Gómez, J. (1999). Técnicas de detección del funcionamiento diferencial en ítems politómicos. *Metodología de las Ciencias del Comportamiento*, 1, 39-60.
- Hidalgo, M.D. y López, J.A. (1997). Comparación entre las medidas de área, el estadístico de Lord y la regresión logística en la evaluación del funcionamiento diferencial de los ítems. *Psicothema*, 9, 417-431.
- Inderbitzen, H. M., Walters K. S. y Bukowski, A L (1997). The role of social anxiety in adolescent peer relations: Differences among sociometric status groups and rejected subgroups. *Journal of Clinical Child Psychology*, 26, 338-348.
- Inglés, C.J., Méndez, F.X. e Hidalgo, M.D. (2000). Cuestionario de Evaluación de las Dificultades Interpersonales en la Adolescencia. *Psicothema*, 12, 390-398.
- Inglés, C. J., Hidalgo, M. D. y Méndez, F. X. (2001, Marzo). *Dificultad interpersonal en la adolescencia: Relación con variables de ansiedad social y personalidad*. Comunicación presentada en el I Symposium sobre Psicología Clínica Infantil, Granada, España.
- Inglés, C. J., Méndez, F. X. e Hidalgo, M. D. (1999, Abril). *Estudio sobre la habilidad para hablar en público en población española preuniversitaria*. Comunicación presentada en el I Symposium sobre Habilidades Sociales: Técnicas y Areas de Aplicación, Granada, Spain.
- Kim, S.H. y Cohen, A.S. (1992). IRTDIF: A computer program for IRT differential item functioning analysis. *Applied Psychological Measurement*, 16, 158.
- La Greca, A. M. y Lopez, N. (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of Abnormal Child Psychology*, 26, 83-94.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Méndez, F. X., Hidalgo, M. D. e Inglés, C. J. (en prensa). The Matson Evaluation of Social Skills with Youngsters: Psychometric properties of the Spanish translation in the adolescent population. *European Journal of Psychological Assessment*.

- 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 medidas. *Psicothema*, 11, 65-74.
- Méndez, F. X., Inglés, C. J. e Hidalgo, M. D. (2001). *La versión española abreviada del Personal Report of Confidence as Speaker: Fiabilidad y validez en población adolescente*. Manuscrito remitido para su publicación.
- Millsap, R.E. y Everson, H.T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17, 297-334.
- Mislevy, R.J. y Bock, R.D. (1990). *PC-BILOG: Item analysis and test scoring with binary logistic models*. [Computer program]. Mooresville, IN: Scientific Software.
- Montorio, I., Fernández, M., Lázaro, S. y López, A. (1996). Dificultad para hablar en público en el ámbito universitario: Eficacia de un programa para su control. *Ansiedad y Estrés*, 2, 227-244.
- Olivares, J., García-López, L. J., Hidalgo, M. D., Turner, S. M. y Beidel, D. C. (1999). The Social Phobia and Anxiety Inventory: Reliability and validity in an adolescent Spanish population. *Journal of Psychopathology and Behavioral Assessment*, 21, 67-78.
- Ollendick, T. H., Matson, J. L. y Hesel, W. J. (1985). Fears in visually-impaired and normally-sighted youths. *Behaviour Research and Therapy*, 23, 375-378.
- Paul, G. L. (1966). *Insight vs desensitization in psychotherapy*. Stanford, CA: Stanford University Press.
- Phillips, G. C., Jones, G. E., Rieger, E. J. y Snell, J. B. (1997). Normative data for the Personal Report of Confidence as a Speaker. *Journal of Anxiety Disorders*, 11, 215-220.
- Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika*, 53, 492-502.
- Raju, N.S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14, 197-207.
- Shepard, L.A. (1981). Identifying bias in test items. *New directions for testing and measurement*, 11, 79-104.
- Stocking, M.L. y Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- Strauss, C. C. y Last, C. G. (1993). Social and simple phobias in children. *Journal of Anxiety Disorders*, 7, 141-152.
- Tittle, C.K. (1982). Use of judgmental methods in item bias studies. En R.A. Berk (Ed.) *Handbook of methods of detecting item bias*. Baltimore, MD: Johns Hopkins University Press.
- Vaughn, S. y Lancelotta, G. X. (1990). Teaching interpersonal social skills to poorly accepted students. Peer-pairing versus non-peer-pairing. *Journal of School Psychology*, 28, 181-188.