

Construcción de una escala de depresión con el modelo de Rasch

Natalia Cadavid, Ana R. Delgado y Gerardo Prieto
Universidad de Salamanca

El objetivo de este trabajo es evaluar las propiedades psicométricas de un cuestionario compuesto por diferentes indicadores de depresión en población joven española con el objetivo de mejorar su calidad técnica. Participaron 310 sujetos de entre 18 y 24 años, con estudios de nivel básico, medio y superior. Los datos se analizaron con el modelo de Rasch. Se ha encontrado que el ajuste medio de los ítems al modelo es alto, y que también lo es la fiabilidad promedio de las estimaciones de los ítems (.97) y de las personas (.88). Tras eliminar los cuatro ítems que desajustan y los doce que muestran DIF asociado al sexo, la escala resultante permite una medición objetiva de la depresión. Con esta última escala se encuentran las previstas diferencias asociadas al sexo.

Construction of a depression scale with the Rasch model. This study examines the psychometric properties of a depression questionnaire. The goal was to improve the technical quality of traditional measures of depression in Spanish youth. 310 participants, aged 18-24 years, filled in the self-report questionnaire. The data were analysed by means of the Rasch model. Results show that model fit, average item reliability (.97), and average person reliability (.88) are high. After deleting four indicators showing misfit and 12 showing sex bias, the resulting scale measures clinical depression objectively. Using this scale, the expected sex-related differences are found.

Actualmente, los autoinformes de depresión son uno de los medios más utilizados para detectar sintomatología depresiva en la asistencia médica española. En su mayoría son medidas construidas con la Teoría Clásica de los Tests (TCT) que se sustentan básicamente en observaciones psiquiátricas y teorías psicológicas, dejando en segundo plano los hallazgos de la neurociencia afectiva, así como el empleo de metodologías psicométricas más eficientes para su evaluación y análisis.

Estas limitaciones nos llevaron a construir un nuevo cuestionario que integra diferentes fuentes empleadas en la evaluación de la depresión. Incluye ítems inspirados en los enunciados del Inventario de Beck para la depresión (BDI), del Cuestionario de ansiedad estado - rasgo (STAI), de la Escala de afectividad positiva y negativa (PANAS), así como de criterios del Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-IV) y otros indicadores derivados de hipótesis de la neurociencia afectiva. Las dos últimas fuentes se utilizaron para desarrollar nuevos ítems que describen aspectos de la depresión no cubiertos por las tres primeras escalas.

Inicialmente, el concepto de depresión se operacionalizó a partir de los indicadores clínicos del DSM-IV (American Psychiatry Association, 1995), puesto que el constructo de Episodio Depresivo Mayor es la base del diagnóstico del trastorno depresivo e incluye una visión global de la depresión clínica (Vázquez y Sanz, 1995).

Los planteamientos de la neurociencia afectiva se tuvieron en cuenta por dos motivos. El primero tiene que ver con que este enfoque comprende la depresión como un trastorno de la emoción en el que la alteración de sus mecanismos neurales se ven reflejados en expresiones disfuncionales de la conducta y regulación afectiva (Davidson et al., 2002a; Davidson, Pizzagalli, Nitschke, y Putnam, 2002b; Tucker, Luu, Frishkoff, Quirino, y Poulsen, 2003). Estas manifestaciones son las que esperamos se vean reflejadas en los contenidos de los ítems del cuestionario diseñado.

En segundo lugar, en la neurociencia afectiva se encuentra evidencia para plantear que la tristeza, depresión y ansiedad son fenómenos patofisiológicamente diferentes (Liotti, Mayberg, McGinnis, Brannan, y Jerabeck, 2002), por eso nos interesamos en verificar que los ítems de depresión reflejen estas diferencias a nivel comportamental. En otras palabras, que estos ítems evalúen la depresión y no otros constructos afectivos relacionados. La introducción de nuevos indicadores construidos desde esta perspectiva cubrirá nuevos aspectos no tenidos en cuenta en las escalas tradicionales, permitiendo así mantener vigentes los contenidos del cuestionario con los últimos hallazgos científicos sobre depresión.

Los ítems inspirados en el BDI, STAI y PANAS se incluyeron por dos razones: son autoescalas de uso común en contextos clínicos y de investigación, y existen versiones en español. Los ítems inspirados en enunciados del BDI se incluyeron porque se ha comprobado que miden el Episodio Depresivo Mayor (Bados, Solanas, y Andrés, 2005; Creamer, Foran, y Bell, 1995; Sanz y Vázquez, 1998; Skorikov y Vandervoort, 2003; Yin y Fan, 2000). Los ítems de contenido similar a los del STAI se emplearon para analizar las preguntas asociadas al concepto de ansiedad y evaluar su capacidad discriminativa entre indicadores de depresión y ansiedad, ya que la literatura encuentra que estas escalas presentan una alta co-

Fecha recepción: ??? • Fecha aceptación: ???

Correspondencia: Natalia Cadavid
Facultad de Psicología
Universidad de Salamanca
37005 Salamanca (Spain)
E-mail: ncadavid@usal.es

morbilidad entre sus constructos (Caci et al., 2003; Clark, Steer, y Beck, 1994; Spielberg, Gorsuch, y Lushene, 1994; Watson, Clark, y Tellegen, 1988). Los ítems inspirados en los contenidos del PANAS se utilizaron puesto que la literatura establece que la validez discriminante de los autoinformes de depresión se aumenta con la inclusión de síntomas indicativos de baja afectividad positiva (Sandín et al., 1999).

El método empleado para analizar las propiedades psicométricas de este nuevo cuestionario es el modelo de Rasch. Éste constituye un acercamiento alternativo a la TCT que resuelve algunas de sus desventajas metodológicas (Prieto y Delgado, 2003; Rasch, 1960; Wright y Stone, 1979). Existe un consenso de que este modelo es superior métricamente a la TCT, por las siguientes razones (Wilson, 2005):

1. En el caso de ajuste al modelo, la escala tiene propiedades de intervalo, lo que asegura que las diferencias en la escala son equivalentes a lo largo de todo el continuo del atributo medido.
2. La propiedad de objetividad específica garantiza que los parámetros de los ítems son independientes de la muestra de sujetos, y los parámetros de los sujetos, de la muestra de ítems.
3. Permite estimar el ajuste al modelo de los ítems y las personas. El grado de ajuste se cuantifica mediante medias de residuos (INFIT y OUTFIT), las cuales permiten asegurarse de la adecuación del instrumento para medir el constructo de interés.
4. Es posible obtener estimaciones de la fiabilidad de los parámetros a nivel individual y a nivel de grupo. Las estimaciones a nivel individual se refieren al error estándar (error típico de medida) del parámetro del ítem o del sujeto (Wright y Stone, 1979), que es igual a la desviación típica de las diferencias entre el valor verdadero (θ o β) y sus estimaciones empíricas (θ^* o β^*). A nivel de grupo se cuantifica la fiabilidad mediante un estadístico similar al coeficiente de fiabilidad clásico (cociente entre la varianza verdadera y la varianza observada de los parámetros). En el caso de los parámetros de las personas el estadístico se denomina *Person Separation Reliability*, en el caso de los ítems *Item Separation Reliability* (Schumacker, 2004).

Con posterioridad a la verificación del ajuste global, se realizó un análisis del Funcionamiento Diferencial de los Ítems (FDI) asociado al sexo, el cual pone a prueba uno de los aspectos de la validez de los ítems (Suárez, 2001). Esta técnica se emplea para purificar el cuestionario de posibles factores latentes desconocidos.

Para detectar el FDI se empleó la comparación entre los parámetros de localización de los ítems en la muestra de varones y la muestra de mujeres (Wright y Stone, 1979).

En definitiva, el propósito de emplear esta metodología consiste en diferenciar entre síntomas inespecíficos de la depresión y aquellos que permitirán un diagnóstico más preciso de esta psicopatología con una validez generalizada en ambos sexos.

Método

Participantes

La muestra se compone de 310 españoles (155 mujeres, 155 varones), entre los 18 y 24 años de edad. En cuanto al nivel edu-

cativo, los participantes se distribuyeron como sigue: 45 varones y 36 mujeres en el nivel básico, 71 varones y 50 mujeres en el medio, y 39 varones y 69 mujeres en el nivel superior. Los sujetos fueron entrevistados individualmente en diferentes lugares, con el fin de obtener una muestra que abarcara el mayor rango posible del constructo. Del total de la muestra, se utilizaron los datos de 304 sujetos para estimar los parámetros, ya que los seis restantes presentaban patrones de desajuste al modelo de Rasch.

Materiales

Se construyó un cuestionario sobre depresión (véase tabla 1), consistente en preguntas cerradas y dicotómicas, relacionadas con el episodio depresivo mayor, afectividad negativa y positiva. Las preguntas derivadas del BDI se formularon en estilo afirmativo, abarcando ambos géneros. Se utilizaron los enunciados que representaban el nivel más alto de perturbación psicoafectiva; también se incluyeron los cambios del BDI-II reportados por Arnau, Meagher, Norris y Bramson (2001) y Penley, Wiebe y Nwosu (2003). Varios de los enunciados derivados de este test fueron divididos en dos oraciones separadas, al considerar que una misma pregunta indagaba dos aspectos diferentes. Por ejemplo, la afirmación, *Me siento tan triste o tan desgraciado que no puedo soportarlo*, se dividió en dos afirmaciones independientes, así: *Me siento tan triste que no puedo soportarlo* y *Me siento una persona tan desgraciada que no puedo soportarlo*.

El formato del instrumento contiene un apartado de identificación, 35 afirmaciones, 28 términos y ocho preguntas. Las afirmaciones son enunciados que indagan la presencia de pensamientos, sentimientos y conductas relacionadas con depresión patológica. Los términos consisten en palabras aisladas que hacen referencia a estados afectivos positivos y negativos. Las preguntas averiguan la presencia de factores de riesgo y sucesos vitales asociados al episodio depresivo.

Procedimiento

La aplicación constó de dos partes, en la primera se realizó una exploración inicial del cuestionario con el fin de ajustar la redacción de los enunciados, en la segunda se efectuó la toma de datos. Esta última consistió en la autoaplicación individual del cuestionario en lugares públicos de la ciudad y en instituciones públicas y privadas de educación formal. Los sujetos participaron voluntariamente en ambos casos, respondiendo al cuestionario en un tiempo de cinco a diez minutos.

Los datos se analizaron con el programa QUEST (Adams y Khoo, 1996).

Resultados

La tabla 1 muestra para cada uno de los ítems del cuestionario su localización, el error estándar asociado y los estadísticos de ajuste.

Los resultados revelan que la mayor parte de los ítems se ajustan al modelo. Los valores promedio y las desviaciones típicas de ajuste sugieren el ajuste global de los ítems al modelo de Rasch (Infit= .99, DE= .14; Outfit= .93, DE= .37). El Infit oscila entre .78 y 1.44 y el outfit entre .23 y 1.82.

Tabla 1
Ítems del cuestionario con sus valores psicométricos correspondientes (en negrita los que componen la versión final)

Número del ítem	Contenido	Nivel de dificultad (en logits)	Error estándar	Infit	Outfit
1	Me siento tan triste que no puedo soportarlo	1.41	0.30	0.96	0.59
2	Estoy a dieta para adelgazar	0.07	0.18	1.08	1.42
3	He fracasado en las cosas que me he propuesto	-0.52	0.16	1.00	1.47
4	Siento que estoy siendo castigado/a	0.45	0.21	1.02	1.26
5	He perdido todo el interés por los demás	1.60	0.33	1.00	0.76
6	Las cosas que antes solían irritarme, me importan poco	-1.75	0.13	1.44	1.80
7	Estoy comiendo más que antes	-1.36	0.13	1.34	1.60
8	He perdido el interés por las actividades que antes realizaba con placer	0.07	0.18	0.95	0.90
9	He pensado en quitarme la vida	1.41	0.30	0.98	0.55
10	Me culpo por todo lo malo que sucede	0.45	0.21	0.92	0.88
11	Estoy demasiado cansado/a para hacer nada	-0.56	0.15	1.01	0.94
12	Me siento una persona totalmente fracasada	1.41	0.30	0.87	0.38
13	Tengo problemas para concentrarme	-2.48	0.13	0.89	0.87
14	Me detesto	1.84	0.36	0.96	0.46
15	Me suicidaría si tuviese la oportunidad	2.85	0.58	0.97	0.25
16	Me siento triste en situaciones que antes no me hacían sentirme así	-1.05	0.14	0.87	0.77
17	Antes era capaz de llorar, pero ahora no puedo, incluso aunque quisiera	0.37	0.20	1.13	1.23
18	Todo me produce insatisfacción	1.60	0.33	0.89	0.39
19	Me siento inútil y me culpo por ello	1.09	0.26	0.90	0.56
20	Ya me es imposible tomar decisiones	0.96	0.25	1.02	1.02
21	Estoy durmiendo menos que antes	-2.08	0.13	1.21	1.27
22	Me siento culpable constantemente	1.09	0.26	0.93	0.53
23	Presiento que todo me va a salir mal	-0.17	0.17	0.86	0.70
24	He intentado matarme	2.14	0.42	0.99	0.51
25	Me siento una persona tan desgraciada que no puedo soportarlo	2.14	0.42	0.90	0.23
26	Situaciones que antes no me producían miedo, ahora sí lo hacen	-1.15	0.14	1.06	1.00
27	Estoy durmiendo más que antes	-1.15	0.14	1.38	1.82
28	Todo me aburre	0.79	0.23	1.02	0.77
29	Siento que el futuro es desesperanzador y las cosas no mejorarán	-0.03	0.18	1.01	1.24
30	No puedo dejar de pensar cosas malas sobre mí	0.73	0.23	0.88	0.59
31	He perdido totalmente el interés por el sexo	1.98	0.39	0.98	0.64
32	Me siento más triste por la mañana que a otras horas del día	-0.20	0.17	0.98	0.83
33	Estoy comiendo menos que antes	-0.63	0.15	1.12	1.25
34	Todo me sale mal, no tengo suerte con nada	-0.03	0.18	0.86	0.73
35	Tardo más que antes en las actividades cotidianas	-0.12	0.17	0.96	0.90
36	Tranquilo/a	-1.16	0.14	0.86	0.77
37	Oprimido/a	0.04	0.18	0.95	0.76
38	Activo/a	-0.81	0.15	0.96	0.92
39	Entusiasmado/a	-1.35	0.13	0.89	0.89
40	Confiado/a	-1.53	0.13	0.94	0.92
41	Alterado/a	-1.43	0.13	0.94	0.84
42	Desasosegado/a	-0.66	0.15	1.23	1.59
43	Alegre	-0.15	0.17	0.78	0.54
44	Contrariado/a	-0.75	0.15	0.88	0.80
45	Satisfecho/a	-0.89	0.14	0.87	0.76
46	Angustiado/a	-0.52	0.16	0.79	0.62

Tabla 1
Ítems del cuestionario con sus valores psicométricos correspondientes (en negrita los que componen la versión final)

Número del ítem	Contenido	Nivel de dificultad (en logits)	Error estándar	Infit	Outfit
47	Inspirado/a	-2.37	0.13	1.03	1.01
48	Preocupado/a	-2.29	0.13	0.83	0.78
49	Relajado/a	-1.80	0.13	0.89	0.87
50	Seguro/a	-1.38	0.13	0.82	0.75
51	Nervioso/a	-2.01	0.13	0.86	0.80
52	Decidido/a	-1.59	0.13	0.95	0.88
53	Cómodo/a.	-0.85	0.14	0.89	0.97
54	Animado/a	-0.34	0.16	0.79	0.58
55	Sobreexcitado/a	-0.85	0.14	1.22	1.43
56	Motivado/a	-1.01	0.14	0.89	0.83
57	Aturdido/a	-0.06	0.18	0.89	0.71
58	Vacío/a	-0.12	0.17	0.81	0.60
59	Orgullosa/a	-1.77	0.13	0.94	0.94
60	Bien	-0.03	0.18	0.77	0.48
61	Tenso/a	-1.16	0.14	0.86	0.76
62	Agitado/a	-1.18	0.14	1.03	0.98
63	Solo/a	-0.12	0.17	0.90	0.73
Si presenta alguna de las siguientes enfermedades, márkelas con una X:					
64	Hipotiroidismo	3.95	1.00	1.00	0.34
65	Trastorno depresivo	3.26	0.71	0.99	0.57
66	Trastorno ansioso	1.72	0.34	1.01	0.59
67	Anemia	1.24	0.28	1.06	1.14
68	Alcoholismo	2.14	0.42	0.98	0.57
69	¿Alguno de sus padres o hermanos/as ha sufrido depresión?	-1.15	0.14	1.28	1.32
70	¿Se le ha muerto un ser querido en el último año?	-1.33	0.13	1.33	1.49
71	¿Se han separado o divorciado sus padres en el último año?	1.84	0.36	1.03	0.99
72	¿Ha sufrido un desengaño amoroso en los últimos meses?	-1.36	0.13	1.07	1.10
73	¿Ha consumido alcohol últimamente? ¿Con qué frecuencia lo ha hecho?	-3.40	0.14	1.08	1.53
74	Una vez al mes	-3.40	0.14	1.08	1.53
75	Una vez a la semana	-1.78	0.13	1.27	1.28
76	Todos los días	1.98	0.39	1.05	1.37
77	Varias veces en el mismo día	2.56	0.51	1.02	1.34
78	¿Ha consumido últimamente marihuana y/o hachís? ¿Con qué frecuencia lo ha hecho?	-1.26	0.13	1.10	1.07
79	Una vez al mes	-1.26	0.13	1.10	1.07
80	Una vez a la semana	-0.09	0.17	1.02	0.90
81	Todos los días	1.09	0.26	0.98	0.82
82	Varias veces en el mismo día	1.41	0.30	0.96	0.65
83	¿Ha consumido últimamente heroína, cocaína y/o éxtasis? ¿Con qué frecuencia lo ha hecho?	0.84	0.24	1.06	1.03
84	Una vez al mes	0.84	0.24	1.06	1.03
85	Una vez a la semana	2.56	0.51	1.01	1.23
86	Todos los días	3.26	0.71	1.02	1.74
87	Varias veces en el mismo día	3.26	0.71	1.02	1.74

Nota. N= 310, Nivel de probabilidad= 0.50

La localización de los ítems se distribuye entre 3.95 y -3.40 logits. Esto significa que los ítems permiten identificar un amplio rango de síntomas depresivos. Los ítems que puntuaron alto en la escala ($\beta > 2$) presentan los mayores errores estándar ($EE > .40$), debido al reducido número de participantes con depresión patológica en la muestra. No obstante, el índice de fiabilidad promedio (Item Separation Reliability) es apropiado (.97). El ajuste medio y las desviaciones típicas de los sujetos también son apropiados (Infit= 1.00, $DE= .17$; Outfit= .93, $DE= .37$). Un análisis pormenorizado de los ítems muestra el desajuste de 4 de ellos (ítems 6, 7, 27 y 70). Los dos últimos presentaron también FDI al igual que otros 10 ítems (n.^{os} 16, 26, 42, 48, 51, 53, 55, 59, 60 y 63). Todos se eliminaron del análisis de la última versión del cuestionario.

Los criterios diagnósticos y los factores de riesgo (ítems 64-87) también se excluyeron de la última versión de la escala, una vez observado que evalúan satisfactoriamente la presencia de síntomas depresivos, pero no constituyen indicadores esenciales para evaluar esta patología.

Los ítems que ajustaron al modelo y que constituyen indicadores específicos de la depresión se emplearon para realizar un análisis de las propiedades métricas de la última versión del cuestionario. Los resultados obtenidos en este segundo análisis muestran que los ítems tienen unos valores de ajuste adecuados: el Infit varía entre .74 y 1.37 y el Outfit entre .15 y 1.50. La fiabilidad global (Item Separation Reliability) es muy alta (.97). Sólo 5 ítems presentan errores estándar superiores a .40, debido a que hay pocos sujetos en la muestra con altos niveles de depresión.

La localización de los ítems oscila entre -2.76 y 3.33 logits, con una media de 0 ($DE= 1.45$). Estos resultados sugieren que este conjunto de ítems cumplen los requisitos necesarios para identificar una amplia gama de síntomas depresivos (figura 1).

Las estimaciones de los sujetos son fiables. El índice de fiabilidad global (Person Separation Reliability) es de .82. Los participantes presentaron un adecuado ajuste al modelo: Infit promedio= .99 ($DT= .17$) y Outfit promedio= .98 ($DT=.79$). Sólo el 13.49% de la muestra presentó valores de Infit y Outfit superiores a 1.50. El nivel medio de depresión de los sujetos fue de -2.40 ($DT= 1.27$), fluctuando entre 5.07 y 1.34 logits. Las medias de mujeres y varones fueron de -2.31 y -2.75 con desviaciones típicas de 1.37 y 1.34, respectivamente ($F(1, 302)= 8.07, p= .005; d= -0.32$), lo que indica una diferencia significativa, aunque pequeña, en depresión.

Discusión y conclusiones

El modelo de Rasch nos ha permitido evaluar, comparar y escalar conjuntamente ítems de cuestionarios tradicionales, permitiéndonos a la vez conocer las propiedades psicométricas de diferentes herramientas de diagnóstico clínico.

Concretamente, se ha encontrado que el cuestionario diseñado posee propiedades psicométricas apropiadas de fiabilidad y validez, ajustándose al modelo de Rasch y presentando un promedio en la severidad de los ítems superior al nivel presentado por los sujetos. Esto significa que el cuestionario es útil para medir niveles de depresión clínica.

Otro hallazgo importante hace referencia a tres indicadores compuestos, dos derivados del BDI y otro del STAI. En el cuestionario de este trabajo, todos ellos fueron divididos en dos ítems (n.^{os} 1 y 25, 18 y 28, 55 y 57). Se encontró que las subdivisiones de un mismo ítem no se ubican en la misma zona del continuo como se esperaba (véase figura 1). También se observó que mientras

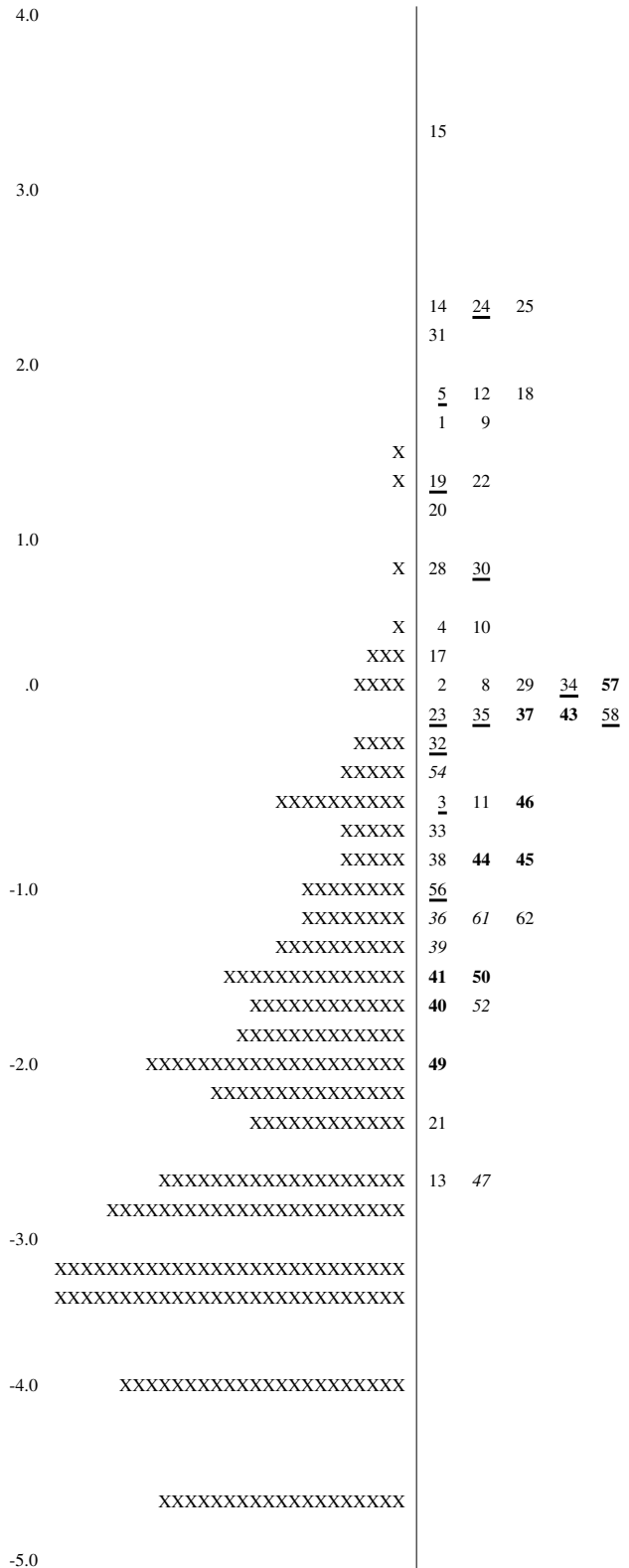


Figura 1. Distribución de los ítems del cuestionario y de los sujetos a lo largo del continuo de depresión. Los ítems de contenido similar al BDI están escritos en tipo regular; los derivados del STAI en negrita y los similares al PANAS se encuentran en cursiva. Los números subrayados representan ítems derivados del DSM-IV y de los planteamientos de la neurociencia afectiva

una subdivisión de un mismo ítem desajusta con el modelo de Rasch, la otra parte ajusta, siendo apropiada para evaluar sintomatología depresiva. En otras palabras, las subdivisiones difieren en su capacidad para medir depresión. Se evidencia así la poca utilidad de los ítems compuestos para realizar una medición precisa de la depresión, restringiendo la capacidad evaluativa de todo el cuestionario. La subdivisión de los ítems compuestos puede ser suficiente para ajustarlos (Stone, 2003).

El análisis individual de los indicadores muestra que algunos de ellos desajustan y/o presentan FDI. Los tres ítems que desajustan, con contenidos derivados del BDI, son inadecuados para medir sintomatología depresiva. Éstos son el ítem 6, síntoma de irritabilidad; se sugiere que desajusta por la alta correlación que mantiene con otras psicopatologías afectivas (Clark et al., 1994; Sandín et al., 1999; Sato et al., 2004). Éste parece ser un indicador general de afectividad negativa, no característico del cuadro clínico de interés. Los ítems 7 y 27 son derivados de modificaciones del BDI-II para permitir la bidireccionalidad de los indicadores de apetito y sueño, respectivamente (Arnau et al., 2001; Bonilla, Bernal, Santos, y Santos, 2004; Whisman, Pérez, y Ramel, 2000). Su desajuste parece indicarnos que estos cambios no poseen la suficiente capacidad discriminativa para indicar diferentes grados de depresión, incluso para detectar síntomas depresivos. Se puede concluir que el cuadro clínico depresivo clásico estaría marcado por un sueño y un apetito reducidos, siendo mejores predictores de sintomatología depresiva que sus opuestos.

Los doce ítems que presentan FDI parecen ser indicadores que reflejan más contenidos asociados con aspectos biológicos y sociales del género femenino y masculino que con la depresión en sí (Salokangas, Batear, Pacriev, Sohlman, y Lehtinen, 2002). Es decir, que éstos miden diferencialmente a varones y mujeres en atributos diferentes a la depresión.

Los ítems con FDI y aquellos que desajustan con el modelo de Rasch no son buenos indicadores para construir un cuestionario de sintomatología depresiva con validez generalizada. En cuanto a los ítems restantes, los derivados del BDI son los que miden un mayor

rango en la variable. Estos ítems presentan la mejor distribución a lo largo del continuo de depresión, evaluando un amplio espectro de anomalías características de esta psicopatología y comparan ubicación en el continuo de depresión (zona superior) con los indicadores derivados de los postulados de la neurociencia afectiva y el DSM-IV. Estos resultados muestran la adecuación de los ítems derivados del BDI para evaluar síntomas clínicos depresivos.

Los ítems derivados del PANAS y STAI sirven para identificar sujetos con bajos niveles de depresión. No obstante, hay ítems del BDI que miden estos mismos niveles, por tanto, los primeros no mejoran la medición de este cuestionario. Este hallazgo era esperado para los ítems de contenido similar al STAI, puesto que fueron diseñados para medir ansiedad. Es muy probable que ajusten con el modelo de Rasch por contribuir a la medición de un constructo común a la ansiedad y depresión, el de Afectividad negativa. Por otro lado, los ítems derivados de los contenidos de Afectividad positiva del PANAS muestran un resultado diferente al esperado, ya que se suponía que contribuían a la discriminación de sintomatología depresiva (Watson et al., 1988).

De todo esto se concluye que los indicadores derivados del BDI ya miden un amplio rango de síntomas depresivos, por lo que no parece indicado estas tres pruebas (o pruebas de contenido similar) simultáneamente.

Los ítems que ajustan al modelo de Rasch han servido además para mostrar que existe una diferencia en la severidad de síntomas depresivos que presentan los participantes de cada sexo. Esta diferencia concuerda con datos epidemiológicos que resaltan la mayor prevalencia de depresión en mujeres (Comer, 1995).

Futuras investigaciones deben enfocarse hacia la evaluación de otros aspectos metodológicos relacionados con la evaluación de la depresión por medio de autoinformes, para contribuir a mejorar sus propiedades psicométricas. Sería de un enorme valor determinar si escalas compuestas por indicadores dicotómicos son más eficientes que las compuestas por ítems politómicos, como sugiere Stone (2003). Finalmente, sería de interés evaluar la relevancia de incluir ítems comportamentales.

Referencias

- Adams, R.J., y Khoo, S. (1996). *Quest: The interactive test analysis system*. Victoria: ACER.
- American Psychiatry Association (1995). *DSM-IV Manual diagnóstico y estadístico de los trastornos mentales*. Barcelona: Masson.
- Arnau, R.C., Meagher, M.W., Norris, M.P., y Bramson, R. (2001). Psychometric evaluation of the Beck depression inventory - II with a primary care medical patients. *Health Psychology, 20*(2), 112-119.
- Bados, A., Solanas, A., y Andrés, R. (2005). Psychometric properties of the Spanish version of depression, anxiety and depression and stress scales (DASS). *Psicothema, 17*(4), 679-683.
- Bonilla, J., Bernal, G., Santos, A., y Santos, D. (2004). A revised Spanish version of the Beck Depression Inventory: Psychometric properties with a Puerto Rican sample of college students. *Journal of Clinical Psychology, 60*(1), 119-130.
- Caci, H., Baylé, F.J., Mattei, V., Dossios, C., Robert, F., y Boyer, P. (2003). How does the hospital and anxiety and depression scale measure anxiety and depression in healthy subjects? *Psychiatry Research, 118*, 89-99.
- Clark, D.A., Steer, R.A., y Beck, A.T. (1994). Common and specific dimensions of self-reported anxiety and depression: Implications for the cognitive and tripartite models. *Journal of Abnormal Psychology, 103*(4), 645-654.
- Comer, R.J. (1995). Mood disorders. En *Abnormal Psychology* (3rd. ed., pp. 273-300). Nueva York: Freeman and Company.
- Creamer, M., Foran, J., y Bell, R. (1995). The anxiety inventory in a non-clinical sample. *Behavioral Research Therapy, 33*(4), 477-485.
- Davidson, R.J., Lewis, D.A., Alloy, L.B., Amaral, D.G., Bus, G., Cohen, J.D., Drevets, W.C., Farra, M.J., Kagan, J., McClelland, J.L., Nolen-Hoeksema, S., y Peterson, B.S. (2002a). Neural and behavioral substrates of mood and mood regulation. *Biological Psychiatry, 52*, 478-502.
- Davidson, R.J., Pizzagalli, D., Nitschke, J.B., y Putnam, K. (2002b). Depression: Perspectives from affective neuroscience. *Annual Review of Psychology, 53*, 545-574.
- Liotti, M., Mayberg, H.S., McGinnis, S., Brannan, S.Z., y Jerabeck, P. (2002). Unmasking disease-specific cerebral blood flow abnormalities: Mood challenge in patients with remitted unipolar depression. *American Journal of Psychiatry, 159*, 1830-1840.
- Penley, J.A., Wiebe, J.S., y Nwosu, A. (2003). Psychometric properties of the Spanish Beck Depression Inventory-II in a medical sample. *Psychological Assessment, 15*(4), 569-577.
- Prieto, G., y Delgado, A.R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema, 15*, 94-100.

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Salokangas, R.K.R., Vaahtera, K., Pacriev, S., Sohlman, B., y Lehtinen, B. (2002). Gender differences in depressive symptoms: An artefact caused by measurement instruments? *Journal of affective disorders*, 68, 215-220.
- Sandín, B., Chorot, P., Lostao, L., Joiner, T.E., Santed, M.A., y Valiente, R.M. (1999). Escalas PANAS de afecto positivo y negativo: validación factorial y convergencia transcultural. *Psicothema*, 11(1), 37-51.
- Sanz, J., y Vázquez, C. (1998). Fiabilidad, validez y datos normativos del Inventario para la Depresión de Beck. *Psicothema*, 10(2), 303-318.
- Sato, T., Bottlender, R., Sievers, M., Schroler, A., Kleindienst, N., y Moller, H.J. (2004). Evaluating the inter-episode stability of depressive mixed states. *Journal of Affective Disorders*, 81, 103-113.
- Schumacker, R.E. (2004). Rasch measurement: The dichotomous model. En E.V. Smith y R.M. Smith (eds.): *Introduction to Rasch measurement* (pp. 226-257). Maple Grove, Mn: JAM Press.
- Skorikov, V.B., y Vandervoort, D.J. (2003). Relationships between the underlying constructs of the Beck Depression Inventory and the center for epidemiological studies depression scale. *Educational and Psychological Measurement*, 63(2), 319-335.
- Spielberg, C.D., Gorsuch, R.L., y Lushene, R.E. (1994). *Cuestionario de Ansiedad estado - rasgo STAI: adaptación española (4ª ed.)*. Madrid: TEA.
- Stone, M.H. (2003). Substantive scale construction. *Journal of Applied Measurement*, 4, 282-297.
- Suárez, J.C. (2001). Evaluación de la calidad métrica de los elementos del test: análisis de los ítems. En M.J. Navas (coordinador): *Métodos, diseños y técnicas de investigación psicológica* (pp. 320-327). Madrid: Universidad Nacional de Educación a Distancia.
- Tucker, D.M., Luu, P., Frishkoff, G., Quiring, J., y Poulsen, C. (2003). Frontal limbic to negative feedback in clinical depression. *Journal of American Psychology*, 112(4), 667-678.
- Vázquez, C., y Sanz, J. (1995). Trastornos del estado de ánimo: aspectos clínicos. En A. Belloch, B. Sandín y F. Ramos (eds.): *Manual de psicopatología*, vol. 2 (pp. 300-341). Madrid: Mc Graw Hill.
- Watson, D., Clark, L.A., y Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.
- Whisman, M.A., Pérez, J.E., y Ramel, R. (2000). Factor structure of the Beck Depression Inventory-second edition (BDI-II) in a students sample. *Journal of Clinical Psychology*, 56(4), 545-551.
- Wilson, M. (2005). *Constructing measures*. Mahwah, NJ: LEA.
- Wright, B.D., y Stone, M.H. (1979). *Best test design*. Chicago: MESA Press.
- Yin, P., y Fan, X. (2000). Assessing the reliability of Beck Depression Inventory scores: Reliability generalization across studies. *Educational and Psychological Measurement*, 60(2), 201-223.