

EVALUACIÓN DE LAS HABILIDADES SOCIALES EN ADULTOS CON DISCAPACIDAD INTELECTUAL: CALIBRACIÓN DE LA ESCALA SPSS-R MEDIANTE EL MODELO DE RASCH

Víctor B. Arias¹, Laura E. Gómez², Benito Arias³,
Henar Rodríguez³ y Noemí Silva⁴

¹Universidad de Talca (Chile); ²Universidad de Oviedo; ³Universidad de Valladolid;

⁴Centro San Juan de Dios de Valladolid (España)

Resumen

El objetivo del presente estudio ha sido la calibración de la "Escala de actuación social revisada" (*Social Performance Schedule Scale Revised*, SPSS-R), en una muestra de 181 personas adultas con discapacidad intelectual, mediante el modelo de escalas de clasificación de Rasch-Andrich, a fin de clarificar su estructura, dimensionalidad y propiedades psicométricas. Fue necesaria una reducción drástica de la escala, a fin de obtener una versión con una estructura factorial clara y consistente con el constructo evaluado (habilidades sociales). La escala resultante, de 24 ítems, se configuró en dos dimensiones bien definidas. Sus propiedades psicométricas (fiabilidad y validez) resultaron adecuadas y presentó un correcto ajuste de ítems y personas al modelo de Rasch. Conforme a su dificultad, los ítems se concentraron en la zona media de las dimensiones evaluadas, lo que aconseja en futuras revisiones incluir ítems en rangos altos y bajos de la variable latente.

PALABRAS CLAVE: *habilidades sociales, discapacidad intelectual, evaluación, SPSS-R, teoría de la respuesta al ítem, modelo de Rasch.*

Abstract

This study aims at calibrating the Social Performance Schedule Scale Revised (SPSS-R) in a sample of 181 adults with intellectual disability using the Rasch-Andrich Rating Scale Model with the purpose of elucidating its structure, dimensionality, and psychometric properties. In order to obtain a version with a clear factorial structure and coherent with the analysed construct (social skills), it was necessary to proceed with a drastic reduction of the scale. The resulting scale, composed of 24 items, showed the emergence of two distinct dimensions. Its psychometric properties (reliability and validity) turned out to be adequate, and it showed correct items and individuals fit to the Rasch model. According to their difficulty, the items concentrated in the mean zone of the analysed dimensions, which makes it advisable to include, in future studies, items in the high and low range of the latent variable.

KEY WORDS: *social skills, intellectual disability, assessment, SPSS-R, item response theory, Rasch model.*

Introducción

Quizá no haya una definición de habilidades sociales que sea universalmente aceptada, pero es posible mencionar aspectos centrales en los que parece haber acuerdo en la comunidad científica. Caballo (1993) definió las habilidades sociales como conductas que se producen en contextos interpersonales, y que son reflejo de sentimientos, actitudes, derechos y opiniones de la persona. Dichos comportamientos son adecuados al contexto en el que se producen, son respetuosos con los derechos y opiniones de los demás y, el efecto de los comportamientos adecuados es resolver problemas inmediatos y minimizar el impacto de potenciales problemas futuros. Siguiendo a otros autores, podemos añadir que las habilidades sociales son conductas orientadas a la solución de problemas y a la evitación del conflicto interpersonal innecesario (Matson y Swiezy, 1994), están muy relacionadas con el contexto inmediato en el que se producen y reflejan la capacidad de adaptación de la persona a diferentes situaciones y retos (La Greca, Stone y Bell, 1982); son comportamientos observables y mensurables, y promueven la aceptación social del individuo y su calidad de vida (Bellack, 1983).

Las habilidades sociales están, por otra parte, estrechamente relacionadas con la calidad de vida de las personas. Actualmente el constructo de calidad de vida es esencial en el ámbito de la discapacidad intelectual, tanto para su investigación como para el ejercicio profesional, el desarrollo de programas de apoyos individualizados, la mejora de la calidad del servicio y el avance en políticas de inclusión (Gómez, Verdugo, Arias y Arias, 2010; Gómez, Verdugo, Arias e Iruña, 2011). Posiblemente el modelo de calidad de vida en el ámbito de la discapacidad más aceptado por su relevancia científica y aplicada sea el de Schalock y Verdugo (Schalock y Verdugo, 2002; Schalock y Verdugo, 2007; Gómez, Verdugo y Arias, 2010; Schalock y Verdugo, 2012a, 2012b; Schalock, Bonham, y Verdugo, 2008; Schalock, Keith, Verdugo, y Gómez, 2010). Entre las dimensiones de calidad de vida contenidas en el modelo de Schalock y Verdugo, al menos dos se refieren de forma directa a aspectos de la interacción entre individuo y ambiente social (las dimensiones *relaciones interpersonales* e *inclusión social*). Así, de la importancia de las habilidades sociales para la promoción de la calidad de vida de la persona (Bellack, 1983), y de su influencia sobre diversos aspectos de la conducta adaptativa (Jacobson, Mullick y Rojahn, 2007) se sigue la importancia de su intervención para la mejora de los resultados personales y la calidad de vida del individuo. De la relevancia de la intervención se deriva, por lógica, la necesidad de sistemas de evaluación adecuados y, por tanto, la necesidad de contar con medios de evaluación de las habilidades sociales que ofrezcan garantías de validez y fiabilidad.

La evaluación de las habilidades sociales tiene una considerable trayectoria histórica, tanto en población general como en poblaciones específicas (Nangle, Hansen, Erdley y Norton, 2010). En el caso de las personas con discapacidad intelectual, los medios más comunes para medir habilidades sociales han sido los registros observacionales de conducta, las técnicas de representación de papeles (*role-playing*) y las escalas de clasificación (Bielecki y Swender, 2004). En la tabla 1 se presenta un resumen de las escalas de clasificación existentes para la evaluación

de habilidades sociales en personas con discapacidad intelectual. En el caso de las escalas de clasificación, los métodos de evaluación en personas con discapacidad intelectual suelen requerir la participación de observadores externos que conozcan bien a la persona evaluada, normalmente familiares y profesionales, debido tanto a la heterogeneidad en las capacidades que se observa en esta población, como también a la dificultad que se ha observado en algunas personas con discapacidad intelectual para auto reportar información de manera fiable, ya sea por dificultades relativas a la capacidad lectora o a la comprensión verbal, o por cierta tendencia a la aquiescencia y a elegir sistemáticamente opciones de respuesta extremas.

Tabla 1

Resumen de las escalas de clasificación desarrolladas con el objetivo de evaluar habilidades sociales en personas con discapacidad intelectual

Instrumento	Autores	Descripción	Población objetivo
"Escala de actuación social-revisada" (<i>Social Performance Survey Schedule-Revised, SPSS-R</i>)	Matson, Helsel, Bellack y Senatore (1983)	Adaptación para personas con discapacidad intelectual de la escala SPSS (Love y Cautela, 1978). Consta de 57 ítems agrupados en 4 dimensiones. Gran parte de sus propiedades psicométricas son desconocidas	Adultos con discapacidad intelectual leve o moderada
"Evaluación de Matson de las habilidades sociales en jóvenes" (<i>Matson Evaluation of Social Skills in Youngsters, MESSY</i>)	Matson, Rotari y Helsel (1983)	Evalúa dos dimensiones: habilidades sociales apropiadas y asertividad inapropiada. Existen dos versiones, una con 62 ítems realizada mediante autoinforme, y otra versión de 64 ítems, realizada mediante heteroinforme con entrevista semiestructurada	Niños entre 4 y 18 años
"Evaluación de la competencia social" (<i>Assessment of Social Competence, ASC</i>)	Meyer et al. (1985)	Diseñada para la medida de las competencias sociales, incluyendo habilidades de la vida diaria y problemas de comportamiento. Consta de 252 ítems, agrupados en 11 dimensiones	Niños y jóvenes con discapacidad intelectual grave y profunda (7-28 años)
"Sistema de calificación de habilidades sociales" (<i>Social Skills Rating System, SSRS</i>)	Gresham y Elliot (1990)	Evalúa tres dimensiones: Colaboración, Aserividad y Autocontrol. En la versión SSRS-CH aparece una cuarta dimensión, (Empatía). La escala se presenta en tres versiones, para profesores (SSRS-T, 3 subescalas de 10 ítems), para padres (SSRS-P, 4 subescalas de 10 ítems) y para niños (SSRS-CH, 4 subescalas de 10 ítems)	Niños y adolescentes con discapacidad intelectual leve o moderada
"Evaluación de Matson de las habilidades sociales para personas con retraso mental grave" (<i>Matson Evaluation of Social Skills for Individuals with Severe Mental Retardation, MESSIER</i>)	Matson (1995)	Consta de 85 ítems, agrupados en 6 dimensiones, 3 positivas y 3 negativas (Verbal, No verbal y General). Se realiza mediante heteroinforme, a través de entrevista semiestructurada	Adultos con discapacidad intelectual grave y profunda

Por otra parte, en general las escalas de clasificación son, frente a otras modalidades de evaluación, las técnicas más sencillas de aplicar. Si bien sus resultados pueden no alcanzar la profundidad y detalle de otras técnicas como la entrevista conductual o la representación de papeles (*role-playing*), resultan aptas en diversas circunstancias, sobre todo para objetivos de cribado (*screening*), puesto que requieren menor entrenamiento del aplicador que otras técnicas, y resultan más económicas en términos de tiempo, especialmente cuando el número de personas a evaluar es elevado. No es extraño entonces que las escalas de clasificación sean uno de los métodos de evaluación más utilizados por clínicos y profesionales (Lecavalier y Butter, 2010). Pese a ello, para la evaluación de habilidades sociales en adultos con discapacidad intelectual, hasta donde sabemos, existen tan sólo dos escalas validadas para su medida (tabla 1), la "Evaluación de Matson de las habilidades sociales para personas con retraso mental grave" (*Matson Evaluation of Social Skills for Individuals with Severe Mental Retardation*, MESSIER; Matson, 1995) y la "Escala de actuación social-revisada" (*Social Performance Survey Schedule-Revised*, SPSS-R; Matson, Hensel, Bellack y Senatore, 1983).

La escala SPSS-R, objeto de este estudio, fue construida a partir de la "Escala de actuación social" (*Social Performance Survey Schedule*, SPSS; Lowe y Cautela, 1978). Matson *et al.* (1983) redujeron la escala SPSS original de 100 ítems a una versión simplificada de 57 ítems, reteniendo aquellos indicadores cuya correlación ítem-total corregida superó el valor de 0,30 después de aplicada a una muestra de 207 adultos con discapacidad intelectual moderada y ligera. Los ítems de la escala final reflejan diversas conductas de interacción social, que deben ser valoradas por personas que conozcan suficientemente a la persona evaluada, en una escala de respuesta de cinco opciones de frecuencia. Según los autores, los ítems se organizaron en cuatro dimensiones de habilidades sociales, según los resultados de un análisis de componentes principales: habilidades adecuadas (26 ítems), habilidades pobres de comunicación (12 ítems), asertividad inadecuada (9 ítems) y comportamiento sociopático (10 ítems). Sin embargo, se observan ciertos aspectos mejorables en el procedimiento de análisis, tales como: a) en el estudio original no se reportaron índices de fiabilidad ni de consistencia interna, a excepción de la mencionada correlación ítem-total corregida; b) se utilizó como método de extracción un análisis de componentes principales, que si bien es apto para fines exploratorios (Abad, Olea, Ponsoda y García, 2010), no debe ser catalogado como técnica de análisis factorial, tal como hacen los autores, a causa de las diferentes asunciones respecto del AFE que este método hace sobre la varianza; c) se utilizó como heurístico de retención de factores la regla de Kaiser-Guttman (retener tantos factores como valores propios superiores a 1 se observen), cuando hay evidencia de que dicha estrategia es susceptible de llevar a la sobre-factorización bajo ciertas circunstancias (Preacher y McCallum, 2003); d) cada dimensión recibió aquellos ítems cuyo peso factorial en la matriz de configuración fuera mayor en relación a su peso en el resto de factores; sin embargo, no se reportaron los pesos factoriales de los ítems en el resto de dimensiones, con lo que es imposible determinar si hubo cargas cruzadas (es decir, ítems por alguna razón multidimensionales, o bien no relacionados con el constructo que se pretende

medir); y e) se retuvieron ítems con cargas factoriales menores a 0,3 en el factor principal, cuando se considera que valores inferiores a 0,3 posiblemente sean indicativos de que el indicador no pertenece a la dimensión.

Como se ha mencionado, existe una evidente escasez de instrumentos de medida de habilidades sociales adecuados para personas adultas con discapacidad intelectual y, hasta donde sabemos, sólo uno de ellos (la escala SPSS-R) está orientado a la evaluación de adultos con un nivel de funcionamiento elevado. Por otra parte, algunos de esos instrumentos, cual es el caso de la escala objetivo de este estudio, no han sido sometidos a un proceso de validación suficientemente riguroso, y la mayoría de sus propiedades psicométricas son desconocidas (Lecavalier y Butter, 2010). Con el objeto de contribuir a suplir dichas carencias, nos hemos planteado mediante el presente estudio validar y calibrar mediante el modelo de Rasch la escala SPSS-R sobre una muestra de adultos con discapacidad intelectual ligera. Las ventajas de la teoría de respuesta a los ítems (TRI), y del modelo de Rasch en particular, respecto a la teoría clásica de los test (TCT) han sido profusamente difundidas (Andrich, 1988; Ayala, 2009; Bond y Fox, 2001; Crocker y Algina, 2008; Embretson y Hershberger, 1999; Embretson y McCollam, 2000; Embretson y Reise, 2000; Fidalgo, 2005; Hambleton, Swaminathan y Rogers, 1991; Prieto y Delgado, 1999, 2000; Prieto y Dias, 2004; Wright y Stone, 1979). Las más relevantes son: invarianza de los parámetros en distintas muestras; posibilidad de realizar estimaciones del grado de precisión con la que cada test (y cada ítem individual) mide los diferentes niveles de habilidad de los sujetos examinados; independencia de la estimación de *theta* (θ) respecto a la prueba utilizada; medición conjunta (los parámetros de las personas y de los ítems se expresan en las mismas unidades intervalares [*logits*] y se localizan en el mismo continuo); objetividad específica (la diferencia entre dos personas en un atributo no depende de los ítems específicos con los que sea estimada); propiedades de intervalo (la interpretación de las diferencias en la escala es la misma a lo largo del atributo medido, de modo que a diferencias iguales entre un sujeto y un ítem le corresponden probabilidades idénticas de dar una respuesta correcta); especificidad del error típico de medida y, finalmente, facilidad para la personalización de las pruebas.

El objetivo de la presente investigación ha sido analizar las propiedades psicométricas de la escala SPSS-R en una muestra de adultos con discapacidad intelectual, desde modelos encuadrados en la TRI, concretamente mediante el Modelo de escalas de clasificación de Rasch-Andrich (*Rasch Rating Scale Model*, *RSM*; Andrich, 1978, 1988; Rasch, 1960, 1977; Wright y Masters, 1982).

Método

Participantes

Se ha utilizado una muestra incidental de 181 adultos con discapacidad intelectual ligera (67,2% hombres y 32,8% mujeres), usuarios de un mismo servicio ocupacional radicado en España, con una media de edad de 37,91 años (*DT*= 10,86). La media de edad de las mujeres fue de 38,64 años (*DT*= 10,90) y la

de los hombres de 36,42 años ($DT= 10,71$). La distribución por edad presentó valores moderados tanto en asimetría ($g1= 0,183$; $EE= 0,183$) como en curtosis ($g2= -0,734$; $EE= 0,363$).

Instrumento

En el presente estudio se utilizó la escala SPSS-R (Matson *et al.*, 1983) para la evaluación de las habilidades sociales en personas con discapacidad intelectual de grado moderado a ligero. Se tradujo la escala original al español, según las directrices para la adaptación y traducción propuestas por la Comisión Internacional de Test (*International Test Commission [ITC]*, 2006; Muñiz, Elosua y Hambleton, 2013; van de Vyver y Hambleton, 1996). Se realizaron tres traducciones independientes del instrumento desde el inglés al español. Se instó a los traductores para que reflejaran los matices del idioma, el contenido de los ítems, las opciones de respuesta y las instrucciones. El objetivo de la traducción fue que los ítems fueran equivalentes palabra a palabra (lingüísticamente), concepto a concepto (semánticamente), y que las expresiones utilizadas fueran aceptables y culturalmente relevantes.

Las traducciones se compararon con el objetivo de reflejar las discrepancias existentes entre ellas y poner de manifiesto las palabras más ambiguas. Trabajando desde el cuestionario original al mismo tiempo que desde la versión del primer traductor (TA) y la versión del segundo y tercero (TB, TC), se realizó una síntesis produciendo una traducción común (T-ABC). No se consideró necesario realizar la traducción inversa puesto que la directa alcanzó suficiente consenso entre los traductores.

La escala de respuesta del instrumento original constaba de cinco opciones de frecuencia (nada, poco, algo, bastante y mucho). Sin embargo, para el presente estudio se consideró variar la escala a una de adecuación de cuatro puntos (muy inadecuado, inadecuado, adecuado, muy adecuado), considerando que varias de las conductas descritas por los indicadores pueden ser interpretadas como déficit en habilidades sociales cuando se presentan de forma pronunciada en cualquiera de los dos polos de frecuencia (p. ej., "mantiene el contacto visual al hablar"). De esta forma, se instruyó a los evaluadores para que valoraran el grado de adecuación en que la persona presentaba cada una de las conductas valoradas (p. ej., para el ítem "inicia contactos y conversaciones con los demás", se instó a los evaluadores a que juzgasen, no tanto la frecuencia de la conducta, sino el grado de adecuación con que la persona la manifiesta, en el entendimiento de que tanto un defecto como un exceso en la frecuencia de dicha conducta puede considerarse inadecuado a las necesidades del individuo y a los condicionantes contextuales).

Procedimiento

Las escalas fueron aplicadas por dos observadores que acreditaron capacidad para evaluar con garantía suficiente, en tanto que ambos eran profesionales psicólogos, especializados en el ámbito de la discapacidad intelectual, y habían tenido la oportunidad de observar a las personas evaluadas durante más de seis

meses. En todos los casos se utilizó un sistema de codificación a fin de preservar la confidencialidad de los resultados.

Análisis estadísticos

En la primera fase del análisis de datos se realizó un análisis factorial exploratorio, implementado mediante el programa FACTOR 9.2 (Lorenzo-Seva y Ferrando, 2006) al objeto de determinar la estructura dimensional de la escala.

En una segunda fase se procedió a la depuración de la escala original, a través de los resultados del análisis factorial exploratorio anterior y de un procedimiento de evaluación inter-jueces sobre las características de observabilidad y sensibilidad de los ítems. Posteriormente, se realizó de nuevo el análisis factorial exploratorio sobre el conjunto de ítems resultante de la evaluación de expertos.

En una tercera fase se utilizó el Modelo de escalas de clasificación de Rasch (*Rasch Rating Scale Model*, RSM; Andrich, 1978, 1988; Rasch, 1960, 1977; Wright y Masters, 1982) implementado en el programa WINSTEPS, v. 3.81.0 (Linacre, 2008, 2014; Linacre y Wright, 1999).

Resultados

Se observaron cuatro casos (2,20%) con ausencia de información en alguna de las variables. Analizado el patrón de datos perdidos, la estructura resultante fue de tipo perdidos completamente al azar (MCAR o *Missing Completely at Random*, Test de Little= 0,968), con lo que se procedió a la eliminación de aquellos casos con algún dato ausente. Los subsiguientes análisis se realizaron sobre una muestra de 177 sujetos.

Inicialmente se eliminó un ítem (“Se minusvalora”) por presentar un índice de homogeneidad corregido inferior a 0,20. Sobre los 56 ítems restantes se aplicaron los mismos análisis que realizaron los autores de la escala en el estudio de validación original (Matson *et al.*, 1983), esto es, un análisis de componentes principales con rotación varimax reteniendo cuatro factores. Los resultados no llevaron a conclusiones claras: a) 17 ítems no se agruparon en los factores descritos en el estudio original; b) 12 ítems presentaron cargas cruzadas en más de un factor (es decir, cargas mayores a 0,30 en más de un factor, y/o diferencias menores de 0,15 entre dos cargas factoriales) y c) 1 ítem presentó carga menor de 0,30 en todos los factores. Habida cuenta de estos resultados, se procedió a depurar la escala original. En primer lugar, se sometieron los 57 ítems a valoración por parte de cuatro jueces expertos. Dichos jueces fueron en todos los casos licenciados en psicología, y con más de cinco años de experiencia en la práctica profesional directa en el ámbito de la discapacidad intelectual. Se pidió a los expertos que valoraran los ítems en cuanto a su sensibilidad (la conducta descrita no depende exclusivamente de rasgos o características personales, sino también de aspectos contextuales, de modo que dichas conductas sean modificables y mejorables mediante la prestación de apoyos individualizados (p. ej., programas de entrenamiento en habilidades sociales) y observabilidad (grado en que el ítem refleja conductas que pueden ser reconocidas por terceras personas a través de la

observación). Se pidió a los jueces que valorasen ambas dimensiones en cada ítem de acuerdo a una escala tipo Likert de cuatro puntos, de (1) nada (sensible, observable) a (4) muy (sensible, observable). Como criterios de retención de los ítems se consideraron: a) su media en ambas dimensiones es superior a 3 y su desviación típica inferior a 1, y b) ambas dimensiones muestran una puntuación de 3 o más en al menos tres de los cuatro expertos. Evaluado el conjunto de ítems, finalmente 24 de ellos cumplieron los criterios de sensibilidad y observabilidad arriba descritos.

Se llevó a cabo un nuevo análisis factorial exploratorio sobre los 24 ítems resultantes de la evaluación por expertos. Se realizaron los cálculos sobre la matriz de correlaciones policóricas, dada la naturaleza ordinal de los datos de entrada. Los datos resultaron adecuados al procedimiento de análisis utilizado ($KMO= 0,84$, Índice de Bartlett= 1784,6 ($gl= 276$; $p= 0,000$)), determinante de la matriz $p= 0,000$). Como método de extracción, se utilizó el método de mínimos cuadrados no ponderados (*unweighted least squares*, ULS) y la factorización de ejes principales (*principal axis factoring*, PAF). Ambos procedimientos de extracción llegaron a resultados prácticamente idénticos (se mostrarán en adelante los resultados obtenidos mediante ULS). Como método de rotación, se optó por Oblimin Directo ($\delta= 0$), dado que en otras medidas de habilidades sociales se informa de la existencia de relación entre factores (Nangle *et al.*, 2010). La regla de Kaiser Guttman (retener tantos factores como valores propios superiores a 1.00 se observen), sugirió la retención de 5 factores. Sin embargo, el tercer factor explicó el 6% de la varianza, con un valor propio de 1,19, y contuvo sólo dos ítems. Dado que existe evidencia de que el método K-1 suele llevar a la sobrefactorización (Preacher y McCallum, 2003), se optó por utilizar un análisis paralelo optimizado (Timmerman y Lorenzo-Seva, 2011) como criterio de decisión, ya que se ha mostrado repetidamente como el procedimiento más fiable para la toma de decisiones sobre cuántos factores es recomendable retener (Abad *et al.*, 2010). El análisis paralelo arrojó una solución óptima de dos factores.

Tras repetir análisis limitando el número de factores a dos, se obtuvieron dos factores claramente definidos, que explicaron el 45,5% de la varianza. Todas las saturaciones factoriales superaron el valor recomendado de 0,3 en su factor, y no se observaron cargas cruzadas (tabla 2). La correlación entre los factores fue de 0,217.

Para denominar a los factores, se conservaron dos de los nombres utilizados en la escala SPSS-R original. El primero, Habilidades adecuadas (en adelante HA), está compuesto por 11 ítems, la mayoría redactados en términos directos, que agrupan conductas relacionadas con destrezas comunicativas, prosocialidad y habilidades de escucha. El segundo factor, Asertividad inadecuada (en adelante AI), está compuesto por 13 ítems que describen conductas relacionadas con el respeto a los derechos de los demás, la tendencia a reaccionar con más agresividad de lo que la situación requiere, y a conductas susceptibles de deteriorar las relaciones sociales del individuo.

Tabla 2
Saturaciones factoriales de los ítems

Ítem	Factor 1 (HA)	Factor 2 (AI)
HA01 Mantiene el contacto visual al hablar	0,776	-0,026
HA02 Inicia contactos y conversaciones con los demás	0,633	0,032
HA03 Pide a los demás que le cuenten como están y lo que les pasa	0,694	0,059
HA04 Muestra interés por lo que el otro dice	0,553	0,160
HA05 Hace preguntas cuando habla con otros	0,808	-0,104
HA06 Se mantiene en contacto con sus amigos	0,586	0,126
HA07 Elogia la ropa, el peinado, etc. de los demás	0,472	0,125
HA08 Pregunta si puede ser de ayuda	0,630	0,152
HA09 Defiende sus derechos	0,583	-0,120
HA10 Mantiene el contacto visual cuando escucha	0,734	-0,074
HA11 Explica las cosas con excesivo detalle	0,631	-0,034
AI01 Muestra impaciencia al escuchar a los otros	0,009	0,589
AI02 Cuida las pertenencias de los demás como si fueran suyas	0,225	0,378
AI03 Interrumpe a los demás	-0,008	0,567
AI04 Hace comentarios embarazosos	0,010	0,671
AI05 Amenaza verbal o físicamente	-0,011	0,754
AI06 Se mete en discusiones	0,096	0,524
AI07 Ordena que la gente haga algo en lugar de pedirlo	0,061	0,582
AI08 Culpa de los problemas a los demás	0,056	0,596
AI09 Se queja	0,018	0,621
AI10 Se enfada con facilidad	0,032	0,650
AI11 Es agresivo cuando está en desacuerdo con alguien	0,039	0,683
AI12 Coge o utiliza cosas de otros sin su permiso	0,249	0,384
AI13 Reacciona con más enfado de lo que la situación requiere	-0,009	0,718

Notas: HA= Habilidades adecuadas; IA= Asertividad inadecuada. Se han marcado en negrita las saturaciones factoriales superiores a un valor absoluto de 0,30.

Evaluación del ajuste de los ítems y las personas

En las tablas 3 y 4 se ofrecen los resultados del análisis de Rasch (localización de los ítems, errores estándar, valores de ajuste interno (*infit*) y ajuste externo (*outfit*), coeficientes punto-biserial y discriminación estimada de los ítems), para la escala HA y la escala AI. Sólo uno de los ítems mostró un valor de ajuste interno mayor de 1,50 (ítem 12 de la escala AI, "Coge o utiliza cosas de otros sin su permiso", ajuste interno= 1,54.); no obstante, se consideró no excluir el ítem del análisis, ya que la desviación del límite de 1,50 es pequeña y su valor de ajuste externo se encontró dentro del rango considerado adecuado (ajuste externo= 1,37).

Tabla 3
Estimaciones de los parámetros de los ítems (“Escala de habilidades adecuadas”)

Ítem	Medida	Error estándar	Ajuste interno		Ajuste externo		Biser-punt.		Exact obs%	Esperada exp%	Discr
			MNSQ	ZEMP	MNSQ	ZEMP	Corr.	Esp.			
HA01 Mantiene el contacto visual al hablar	0,46	0,14	0,95	-0,5	0,93	-0,5	0,73	0,65	65,3	66,6	1,03
HA02 Inicia contactos y conversaciones con los demás	-0,95	0,17	1,03	0,3	0,98	0,0	0,65	0,63	73,9	75,0	0,94
HA03 Pide a los demás que le cuenten como están y lo que les pasa	0,21	0,15	0,85	-1,4	0,83	-1,3	0,68	0,65	79,3	67,6	1,19
HA04 Muestra interés por lo que el otro dice	0,42	0,15	1,13	1,2	1,22	1,6	0,62	0,65	67,6	66,7	0,83
HA05 Hace preguntas cuando habla con otros	-0,65	0,16	0,90	-0,9	0,79	-1,5	0,72	0,63	70,5	72,7	1,13
HA06 Se mantiene en contacto con sus amigos	-1,19	0,17	0,98	-1,0	0,86	-0,8	0,60	0,63	80,1	76,8	1,11
HA07 Elogia la ropa, el peinado, etc. de los demás	0,97	0,14	1,25	2,3	1,25	1,9	0,55	0,67	60,2	64,2	0,69
HA08 Pregunta si puede ser de ayuda	-0,47	0,16	1,08	0,7	1,01	0,1	0,66	0,64	70,5	71,4	0,95
HA09 Defiende sus derechos	0,21	0,15	1,03	0,3	1,04	0,3	0,58	0,65	68,8	67,6	0,96
HA10 Mantiene el contacto visual cuando escucha	0,71	0,14	0,89	-1,1	0,92	-0,6	0,70	0,66	62,5	65,6	1,09
HA11 Explica las cosas con excesivo detalle	0,27	0,15	0,96	-0,4	0,87	-0,9	0,62	0,65	68,8	66,9	1,10
Ítems	M	0,15	1,00	0,0	0,97	-0,2	--	--	69,3	69,2	--
	DT	0,67	0,01	0,11	1,0	0,14	1,0	--	5,3	3,9	--
Personas	M	-0,40	0,98	0,0	0,97	-0,1	--	--	48,6	47,5	--
	DT	1,93	0,15	0,58	1,1	0,64	1,1	--	15,8	5,2	--
Separación de ítems	4,17	Fiabilidad de ítems	0,95	RMSE	0,16	RS-MC	-1,0				
Separación de personas	2,47	Fiabilidad de personas	0,86	RMSE	0,75	RS-MC	0,97				Alfa de Cronbach

Nota: RMSE= Raíz cuadrática media de los errores estándar; RS-MC= Correlación entre medida y puntuaciones originales.

Tabla 4
Estimaciones de los parámetros de los ítems (escala Asertividad Inadecuada, AI)

Ítem	Medida	Error estándar	Ajuste interno		Ajuste externo		Corr. Biser-punt.		Exact obs%	Esperada exp%	Discr
			MNSQ	ZEMP	MNSQ	ZEMP	Corr.	Esp.			
AI01 Muestra impaciencia al escuchar a los otros	1,04	0,14	1,16	1,5	1,14	1,1	0,63	0,64	59,3	65,9	0,81
AI02 Cuida las pertenencias de los demás como si fueran suyas	-1,08	0,18	1,37	2,5	1,35	1,8	0,47	0,55	74,6	79,2	0,74
AI03 Interrumpe a los demás	1,06	0,14	1,12	1,1	1,17	1,3	0,62	0,64	61,0	65,9	0,85
AI04 Hace comentarios embarazosos	-0,12	0,16	0,86	-1,2	0,70	-2,0	0,62	0,59	77,4	73,1	1,18
AI05 Amenaza verbal o físicamente	0,20	0,15	0,77	-2,2	0,75	-1,7	0,72	0,60	74,0	70,8	1,19
AI06 Se mete en discusiones	-0,64	0,17	0,97	-0,2	0,87	-0,7	0,52	0,57	78,5	76,2	1,04
AI07 Ordena que la gente haga algo en lugar de pedirlo	0,17	0,15	1,07	0,6	1,05	0,4	0,62	0,60	71,2	71,0	0,93
AI08 Culpa de los problemas a los demás	0,73	0,15	1,06	0,6	1,00	0,0	0,61	0,63	72,9	67,6	0,93
AI09 Se queja	-0,25	0,16	0,85	-1,2	0,76	-1,5	0,57	0,58	77,4	73,6	1,17
AI10 Se enfada con facilidad	-0,17	0,16	0,87	-1,1	0,73	-1,8	0,60	0,58	75,1	73,3	1,16
AI11 Es agresivo cuando está en desacuerdo con alguien	-0,17	0,16	0,78	-1,9	0,81	-1,2	0,60	0,58	77,4	73,3	1,17
AI12 Coge o utiliza cosas de otros sin su permiso	-1,18	0,18	1,54	3,3	1,37	1,9	0,51	0,55	78,0	80,0	0,70
AI13 Reacciona con más enfado de lo que la situación requiere	0,41	0,15	0,77	-2,3	0,73	-2,0	0,65	0,61	72,3	69,3	1,21
Ítems	M	0,16	1,01	0,0	0,96	-0,3	--	--	73,0	72,3	--
	DT	0,01	0,23	1,7	0,23	1,4	--	--	5,9	4,3	--
Personas	M	0,64	0,94	-0,2	0,96	-0,1	--	--	48,6	47,5	--
	DT	0,17	0,63	1,3	0,72	1,3	--	--	15,8	5,2	--
Separación de ítems	3,92	Fiabilidad de ítems	0,94	RMSE	0,17	RS-MC	-1,0	--	--	--	--
Separación de personas	2,23	Fiabilidad de personas	0,83	RMSE	0,73	RS-MC	0,96	Alfa de Cronbach	0,87	--	0,87

Nota: RMSE= Raíz cuadrática media de los errores estándar; RS-MC= Correlación entre medida y puntuaciones originales.

La polaridad de los ítems indica que todas las correlaciones punto-biserials resultaron positivas y superiores al valor recomendado de 0,20, con un rango de 0,58 a 0,73 en la escala HA, y de 0,47 a 0,72 en la escala AI. En consecuencia, todos los ítems cumplen el requisito crucial en el análisis Rasch de estar alineados en la misma dirección en la variable latente.

El índice de separación de los ítems fue de 4,17 (HA) y 3,92 (AI), indicando que discriminan entre diferentes niveles del rasgo entre los sujetos. La fiabilidad global (*item separation reliability*) (0,95 en HA y 0,94 en AI) indica que los ítems configuran variables bien definidas y que la fiabilidad de la ubicación de los ítems en cada escala es buena, aportando asimismo evidencia del cumplimiento del supuesto de independencia local. Una fiabilidad baja significaría que la muestra no es suficientemente grande como para localizar con precisión los ítems en la variable latente. Con esta muestra, las dificultades de los ítems se estiman con suficiente exactitud.

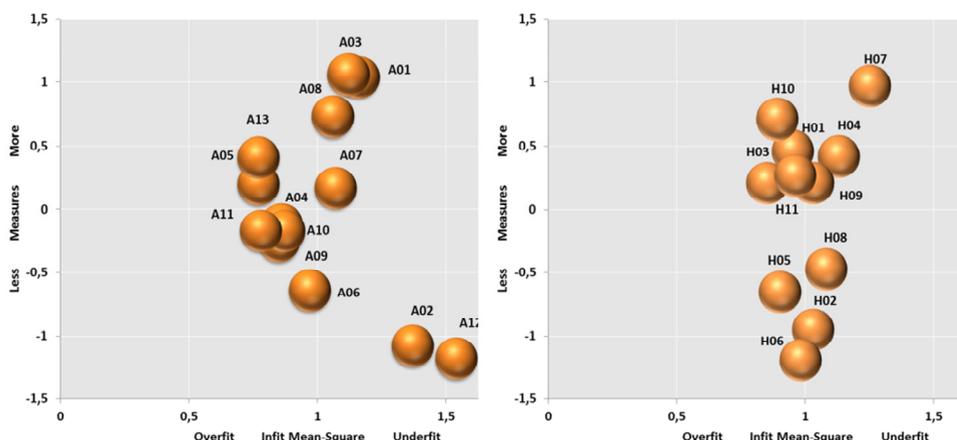
Las estimaciones de los sujetos fueron fiables. La separación es de 2,47 (HA) y 2,23 (AI). Este índice evalúa en qué medida el test discrimina a la muestra en suficientes niveles para nuestro propósito. Equivale, aproximadamente, a un valor KR-20 o alfa de Cronbach de 0,88 (HA) y 0,87 (AI), y denota que ambas escalas, en la muestra estudiada, discriminan al menos entre dos niveles en cada una de las variables latentes estudiadas (es decir, nivel alto y bajo). El índice de fiabilidad promedio de las personas (*person separation reliability*) es apropiado en ambos casos (0,86 en HA, 0,83 en AI).

Los valores RS-MC (correlaciones de las puntuaciones originales respecto de la medida, o *raw score-to-measure correlation*) son las correlaciones de Pearson entre las puntuaciones directas y las medidas, incluyendo las puntuaciones extremas. Se espera que, cuando los datos son completos, sean cercanos a 1,0 para las personas y a -1,0 para los ítems (lo que se cumple en nuestro caso para ambas escalas).

El ajuste medio y las desviaciones típicas de los ítems son apropiados. Para la escala HA fueron: *Ajuste interno*= 1,00; *DT*= 0,11; *ajuste externo*= 0,97; *DT*= 0,14 y para la escala AI: *Ajuste interno*= 1,01; *DT*= 0,23; *ajuste externo*= 0,96; *DT*= 0,23. El ajuste medio y las desviaciones típicas de los sujetos son asimismo apropiados. En este caso, para la escala HA fueron: *Ajuste interno*= 0,98; *DT*= 0,58; *ajuste externo*= 0,97; *DT*= 0,64 y para la escala AI: *Ajuste interno*= 0,94; *DT*= 0,63; *ajuste externo*= 0,96; *DT*= 0,72. Estos resultados sugieren que ambos conjuntos de ítems cumplen en principio los requisitos necesarios para identificar los constructos evaluados.

Una representación gráfica del ajuste mediante ajuste interno y externo para los ítems de ambas escalas puede verse en la figura 1. El ítem más fácil de la escala HA (es decir, con más probabilidades de adhesión) es el 6 ('Se mantiene en contacto con sus amigos'), en tanto que el de adhesión más difícil es el número 7 ('Elogia la ropa, el peinado, etc., de los demás'). Todos los errores estándar son parejos y razonablemente reducidos, como denota el diámetro de la burbuja que representa cada ítem. Finalmente, todos los ítems se sitúan en la zona de 0,5 a 1,5, o zona de ajuste aceptable y productivo para la medida (Linacre, 2008).

Figura 1
Distribución de los ítems en función de los valores de ajuste interno y externo



Nota: El diámetro de las burbujas representa el error estándar.

En cuanto a los sujetos, también se constata un adecuado ajuste al modelo: el ajuste interno promedio ha sido 0,98 ($DT= 0,58$) para la escala HA y 0,94 ($DT= 0,63$) para la escala AI y el ajuste externo promedio fue de 0,97 ($DT= 0,64$) para HA y de 0,96 ($DT= 0,72$) para la AI. Cabe señalar que únicamente 22 sujetos (12,4%) en la escala HA y 24 sujetos (13,5%) en la escala AI presentaron valores superiores a 1,50 en ajuste interno y externo. En consecuencia, en ambas escalas la proporción de personas que presentan un buen ajuste es superior al 85%.

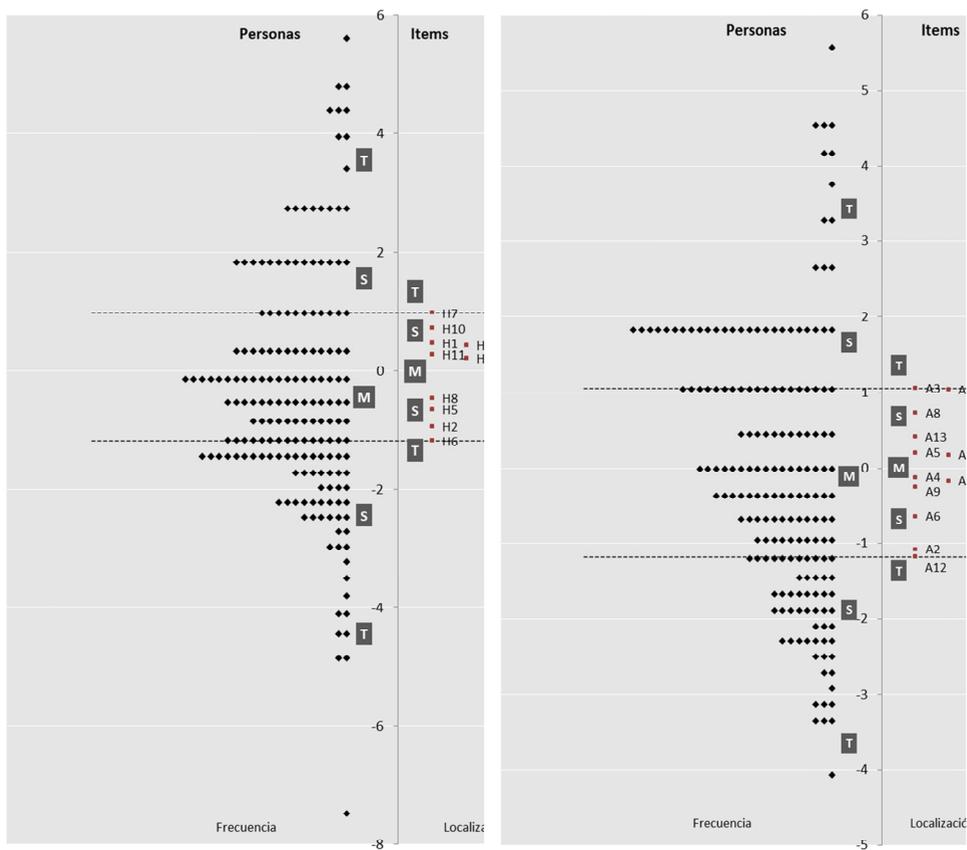
Objetividad específica

El análisis de la objetividad específica se ha llevado a cabo dividiendo la muestra original en dos submuestras aleatorias de ($n_1= 91$; $n_2= 90$) y realizando a continuación un análisis de regresión lineal simple entre los parámetros de dificultad de los ítems obtenidos en cada una de ellas (Prieto y Delgado, 2003). Se utilizó el primer conjunto como predictor y el segundo como criterio. En la subescala AI la correlación entre ambos conjuntos de parámetros ha sido 0,920, la intercepción 0,002, la pendiente 0,976 y el coeficiente de determinación 0,934. En la subescala HA se ha obtenido una correlación entre ambos conjuntos de parámetros de 0,960, una intercepción de 0,001, una pendiente de 0,945 y un coeficiente de determinación de 0,922. Puesto que los valores que denotarían un ajuste perfecto entre los datos y el modelo serían 1, 0, 1 y 1 respectivamente, concluimos que se cumple satisfactoriamente el requisito de invarianza de los parámetros de los ítems y que los datos presentan un buen ajuste global al modelo.

Adecuación del nivel de dificultad de los ítems para la muestra

Los mapas de personas e ítems, también conocidos como ‘Mapas de Wright’, ilustran gráficamente cómo los ítems progresivamente mayores en nivel de dificultad se van solapando con los niveles de las personas en el rasgo latente (habilidades apropiadas y asertividad inadecuada) evaluado. Dado que el modelo de Rasch utiliza la misma medida (*logit*) ambas métricas pueden compararse para determinar si la dificultad de los ítems es o no apropiada para la muestra de personas. Si la muestra fuera apropiada, debería existir un solapamiento considerable en el mapa entre los parámetros de dificultad de los ítems y los niveles del rasgo latente de las personas. A este alineamiento entre ítems y personas se le denomina alineamiento (*targeting*) en el argot del análisis Rasch.

Figura 2
Mapa de Wright (personas-ítems)



Notas: M= Media; S= Desviación típica; T= Dos desviaciones típicas. Las líneas punteadas delimitan la zona de alineamiento (*targeting*) entre personas e ítems.

En la figura 2 se muestran el mapa conjunto de ítems y personas ordenados desde los niveles más altos a los más bajos. Por consiguiente, las personas con niveles elevados en cada rasgo evaluado, así como los ítems de adhesión más difícil, se encuentran en la parte superior del mapa. En ambas escalas podemos comprobar cómo el rango de los parámetros de dificultad de los ítems se solapa parcialmente con el rango de los parámetros del rasgo latente de las personas, lo que indica que todos los ítems han evaluado a los sujetos con distintos niveles en los respectivos rasgos latentes. Con todo, es preciso hacer las consideraciones siguientes. En primer lugar, en el caso de la escala Habilidades Adecuadas, la media de las personas ($M= -0,44$; $DT= 2,00$) es moderadamente inferior a la de los ítems ($M= 0,00$; $DT= 0,67$). En segundo lugar, la amplitud de las personas (de $-7,48$ a $5,60$ *logits*) es muy superior a la de los ítems (de $-1,19$ a $0,97$ *logits*). En tercer lugar, un total de 31 sujetos (17,5%) puntúan por encima del rango de dificultad de los ítems, en tanto que 59 (33,3%) lo hacen por debajo de dicho rango. La zona de alineamiento entre la dificultad de los ítems y la presencia del rasgo latente en los sujetos agrupa a 87 personas (49,1%). En lo que respecta a la escala de Asertividad inadecuada, los resultados son similares: la media de las personas ($M= -0,11$; $DT= 1,77$) se encuentra muy cercana a la de los ítems ($M= 0,00$; $DT= 0,68$). La amplitud de las personas (de $-4,08$ a $5,56$ *logits*) es de nuevo muy superior a la de los ítems (de $-1,18$ a $1,06$ *logits*). 37 sujetos (20,9%) puntúan por encima del rango de dificultad de los ítems, en tanto que 56 (31,3%) lo hacen por debajo de dicho rango. La zona de alineamiento entre la dificultad de los ítems y la presencia del rasgo latente en los sujetos agrupa a 84 personas (47,4%).

Evaluación de la dimensionalidad

Una de las asunciones subyacentes al Modelo de Rasch es que cada escala analizada sea unidimensional. Comprobamos este requisito mediante análisis factorial exploratorio (cuyos resultados se detallan en un apartado anterior), los estadísticos de ajuste interno y externo y mediante el análisis de componentes principales de los residuos estandarizados de Rasch.

En el análisis de los modelos de Rasch, la falta de unidimensionalidad se refleja en unos índices de ajuste pobres. Como más arriba queda dicho, habitualmente se utilizan dos índices de ajuste: ajuste interno y ajuste externo, ambos con un rango teórico desde 0 a $+\infty$, que determinan en qué medida cada uno de los ítems representa la única dimensión subyacente. Mientras que el ajuste interno está afectado por patrones inesperados de respuesta de los sujetos ubicados cerca de la posición del ítem en la escala, el ajuste externo es más sensible a patrones inesperados de respuesta de los sujetos situados lejos de la ubicación del ítem. Puesto que los índices de ajuste se calculan dividiendo el valor de χ^2 por los grados de libertad, los valores de 1 son los ideales, lo que sugiere que la varianza observada equivale a la varianza esperada. Valores de ajuste interno y externo de $1+x$ indican $(100*x)\%$ más variación entre los patrones observado y predicho por el modelo de lo que se esperaría si los datos y el modelo ajustaran perfectamente. Por ejemplo, el ajuste interno de 1,25 obtenido en el ítem 7 de la escala HA ('Elogia la ropa, el peinado, etc. de los demás') indica que en dicho ítem

tenemos un 25% más de variación en los datos observados que la predicha por el modelo. Se asume que los ítems con valores de ajuste mayores que 1,00 presentan un infra-ajuste, lo que sugiere la presencia de ruido no modelado o de otras fuentes de varianza en los datos (Linacre, 2008). Un ítem con un estadístico de ajuste grande generalmente indica que no pertenece al constructo único que está siendo medido. Por el contrario, los ítems con valores de ajuste menores que 1,00 presentan un sobre-ajuste (lo que sugiere que existe menos variación en los datos observados que en el modelo y, en consecuencia, el modelo predice los datos demasiado bien, ocasionando estadísticos sumarios inflados). Como se puede apreciar en las Tablas 3 y 4, en el caso de la escala HA los valores de ajuste interno tienen un rango de 0,85 (ítem 3, 'Pide a los demás que le cuenten como están y lo que les pasa') a 1,25 (ítem 7, 'Elogia la ropa, el peinado, etc., de los demás'); en el caso de la escala AI, el rango de valores de ajuste interno oscilan entre 0,77 (ítem 13, 'Reacciona con más enfado de lo que la situación requiere') y 1,54 (ítem 12, 'Coge o utiliza cosas de otros sin permiso').

En la última columna de las tablas 3 y 4 se señalan los índices de discriminación. Cuando son inferiores a 1,00 indican infra-discriminación, lo que implica una débil diferenciación desde un nivel al siguiente (Linacre, 2008). En nuestro caso, sólo el ítem 7 de la escala HA ("Elogia la ropa, el peinado, etc. de los demás") presenta un índice de discriminación inferior a 0,70.

Además del examen de los valores de ajuste descritos en los párrafos anteriores, llevamos a cabo un análisis de componentes principales de los residuos de Rasch para determinar la unidimensionalidad de las escalas.

El análisis de componentes principales descompone la matriz de correlaciones entre los ítems basándose en los residuos estandarizados (es decir, diferencias entre los valores observados y los predichos por el Modelo de Rasch) para determinar si existen o no otras dimensiones potenciales.

El primer factor del análisis corresponde a la Dimensión Rasch. Se considera adecuada una varianza igual o superior al 60%. La segunda dimensión (o primer contraste de los residuos) indica si existen patrones en las diferencias dentro de los residuos suficientemente grandes como para sugerir que es plausible la existencia de más de una dimensión.

Si la varianza de la dimensión de Rasch fuera escasa, y a la vez fuera significativa en los sucesivos contrastes, la escala podría ser multidimensional. Suele adoptarse la regla de que la segunda dimensión deberá tener al menos tres ítems (de acuerdo con el valor propio [*eigenvalue*]) para poder ser considerada como una posible segunda dimensión, y debería representar al menos un 5% de la varianza no explicada (Linacre, 2008).

El análisis de componentes principales sobre el total de los ítems de ambas escalas muestra que el primer contraste presenta un valor propio de 5,1, muy superior al valor de 3,0 necesario para considerar la existencia de más de una dimensión. Realizado el mismo análisis sobre las escalas por separado, observamos un valor propio para el primer contraste de 1,9 (explicando el 9,4% de la varianza de los datos no modelados) para la escala HA, y de 2,2 (9,7% de la varianza de los datos no modelados) para la escala AI, en ambos por debajo del valor de 3,0. En consecuencia con lo dicho, ambas escalas podrían considerarse unidimensionales

(o 'suficientemente unidimensionales' si se quiere ser más preciso). El análisis de componentes principales sobre la escala HA muestra que el 46,8% de la varianza queda explicada por los datos; este porcentaje es prácticamente idéntico a la varianza explicada por el modelo (47,1%). En el caso de la escala AI, el porcentaje de varianza explicada por los datos asciende al 43,3% (43,4% explicada por el modelo). En el caso de las dos escalas, sus porcentajes de varianza explicada por los datos son moderadamente inferiores al valor recomendado del 60%.

Función de las categorías de respuesta

Se revisaron a continuación las clasificaciones de cada ítem a fin de determinar si las categorías de respuesta funcionaban según lo esperado. En primer lugar, todas las frecuencias de las cuatro categorías utilizadas (*muy inadecuado*, *inadecuado*, *adecuado*, *muy adecuado*) excedieron el mínimo de 10 recomendado por Linacre (1999). En ambas escalas la más frecuente fue *adecuado* ($n= 1113$ en HA, $n= 1478$ en AI) seguida de *inadecuado* ($n= 609$ en HA, $n= 625$ en AI), *muy inadecuado* ($n= 164$ en HA, $n= 131$ en AI) y *muy adecuado* ($n= 61$ en HA, $n= 67$ en AI). Los valores de ajuste interno fueron próximos al valor esperado de 1,00 en todas las categorías (0,95, 0,99, 1,07 y 0,96 para la escala HA; 1,06, 0,94, 0,98 y 1,14 en la escala AI). Los valores de ajuste externo también fueron próximos a 1,00 en las cuatro categorías (0,95, 0,93, 1,10 y 0,88 en HA; 1,05, 0,88, 0,98 y 1,01 en AI), lo que indica que en todos los casos la categoría proporciona más información (es decir, varianza sistemática) que ruido (es decir, varianza de error) en el proceso de medición (Linacre, 1999).

En segundo lugar, se constató que todas las medidas promedio para las categorías avanzaran monotónicamente, y no hubiera ninguna categoría especialmente ruidosa. Así, las medidas promedio (-3,15, -1,52, 0,36 y 3,88 en HA; -2,63, -1,47, 0,53 y 3,29 en AI) y las estimaciones de los umbrales (-3,68, -1,30 y 4,98 en HA; -3,64, -1,40 y 5,04 en AI) presentan un incremento que corre parejo con el incremento de las etiquetas de las categorías, lo que sugiere que la categorización de la escala de clasificación ha sido satisfactoria (figura 3). La secuencia, por tanto, es la siguiente: $\tau_1 < \tau_2 < \tau_3 < \tau_4$, siendo:

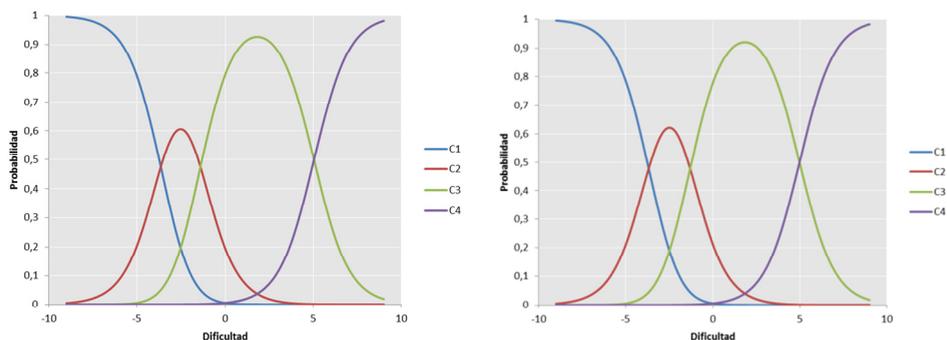
$$\sum_{m=1}^4 \tau_m = 0 : -3,68 - 1,30 + 4,98 = 0 \text{ (HA);}$$

$$\sum_{m=1}^4 \tau_m = 0 : -3,64 - 1,40 + 5,04 = 0 \text{ (AI)}$$

Esta secuencia de valores está indicando que los parámetros de umbral de Rasch-Andrich están ordenados. Por ello, desde la categoría *muy inadecuado*, la transición más probable que se produce es a la categoría *adecuado*, y así sucesivamente. Junto a los valores de estos parámetros de umbral se muestra el error estándar de los pasos del ítem, observándose que los valores son relativamente bajos (0,10, 0,06 y 0,15 en HA; 0,10, 0,06 y 0,14 en AI).

Observando el gráfico de las curvas características de las categorías de respuestas (CCCR) podrá apreciarse con más claridad cuál es la categoría de respuesta más probable a lo largo del continuo. Esta curva relaciona la probabilidad de respuesta a un ítem con su nivel en el constructo medido con el test, siendo útil en la evaluación de las propiedades de los ítems. Como puede observarse en la figura 3, los puntos de intersección entre las categorías de respuestas coinciden con los parámetros de umbral de la medida (τ). A su vez, estos puntos definen en el continuo las regiones de respuestas más probables. Como también se puede apreciar en la figura 3, en ambas escalas se observa gran distancia entre los parámetros umbral de las categorías *inadecuado* y *adecuado*. Este hecho podría ser indicativo de la necesidad de incorporar una nueva categoría de respuesta que contribuyese a compensar dicho “salto” de información (se discutirá este resultado con más detalle en la sección de discusión).

Figura 3
Curvas de probabilidad de las categorías de respuesta



Discusión

Se ha dedicado este estudio a la calibración de la escala SPSS-R, utilizando el Modelo de escalas de clasificación de Rasch-Andrich. Hasta donde sabemos, la SPSS-R no ha sido -al menos en España- calibrada y validada psicométricamente mediante procedimientos encuadrados en la teoría de respuesta a los ítems. Ese marco metodológico dota a los instrumentos de evaluación de una serie de ventajas que no ofrece la metodología de análisis tradicional (es decir, teoría clásica de los test), como son la invarianza de los parámetros, la estimación del grado de precisión de los ítems y del test, la independencia de la estimación con respecto a la prueba, la medición conjunta de ítems y personas, la objetividad específica, las propiedades de intervalo y especificidad del error típico de medida, y la capacidad de personalización de las pruebas. Consideramos adecuado este enfoque, toda vez que la bondad psicométrica de la escala original no está suficientemente fundamentada (Lecavalier y Butter, 2010), y el análisis mediante modelos de TRI permite un conocimiento considerablemente más profundo de las propiedades psicométricas de los ítems y de la escala.

Los resultados han puesto de manifiesto que los datos considerados globalmente cumplen con las exigencias de los estadísticos de ajuste del RSM. Se encontró en términos generales un buen ajuste de las personas y los ítems al modelo: los ítems de cada una de las dos dimensiones de la SPSS-R permiten identificar un rango relativamente amplio de conductas, y tanto el índice de fiabilidad promedio de los ítems, como el de las personas y el índice de fiabilidad global han resultado aceptables.

Por otro lado, el ajuste de los ítems se ha realizado en cuatro fases. En la primera, se sometió a los resultados de la escala completa a análisis factorial exploratorio, sin que fuera posible obtener una solución factorial satisfactoria; en un segundo paso, se depuró la escala según juicios de profesionales expertos, en cuanto a la observabilidad y sensibilidad de las conductas descritas por los ítems; tras eliminar aquellos ítems con valores insuficientes en observabilidad y sensibilidad, se realizó de nuevo el análisis factorial exploratorio sobre un conjunto de 24 ítems, obteniéndose una solución con dos factores claramente definidos, factores refrendados posteriormente por los resultados del análisis de componentes principales sobre los residuos de la dimensión Rasch. Por último, se ha estudiado el ajuste al modelo con la escala depurada de 24 ítems. Los resultados han puesto de manifiesto que todos ellos presentan un ajuste correcto según las expectativas del modelo. Por lo tanto, estimamos que los datos recopilados con los 24 ítems pueden ser explicados convenientemente por el modelo de escalas de clasificación.

Con relación a si los ítems se ordenan homogéneamente y de forma jerárquica respecto a las variables latentes evaluadas, los ítems que componen la escala han mostrado que se distribuyen a lo largo del continuo, sin excesivos saltos entre ellos, por lo que en principio no resultaría necesario reconstruir el instrumento añadiendo ítems destinados a llenar esos vacíos de información. Los resultados indican por tanto que los ítems se distribuyen de forma jerárquica y con un escalamiento adecuado.

Con relación al funcionamiento de las categorías de respuesta y su función de información, resultan adecuadas, con una salvedad. Como se ha visto, las CCCR han mostrado que las categorías de respuesta están ordenadas en todos los ítems, tal como exige el modelo (Wright y Masters, 1982); sin embargo, se aprecia una importante distancia entre los umbrales de las categorías "adecuado" y "muy adecuado", en comparación a las distancias observadas entre los umbrales de las otras categorías. Este resultado puede ser indicativo de la necesidad de incorporar al menos una nueva categoría de respuesta que contribuya a llenar dicho vacío de información.

En ambas subescalas (HA y AI), la zona de alineamiento de los ítems que componen se corresponde aproximadamente con la mitad de los sujetos de la muestra, y en ambos casos la media de las personas fue muy similar a la de los ítems. Dada la distribución de la dificultad de los ítems, las escalas parecen adecuadas para medir rangos medios de la variable latente. Un motivo interesante de investigación futura sería generar un banco de ítems aptos para medir la habilidad social en niveles altos y bajos de la variable.

Como limitaciones de la presente investigación, cabe señalar las siguientes. En primer lugar, el carácter incidental de la selección de los sujetos implica que no sea posible la generalización de los resultados a la población. Sería adecuado, para paliar tal inconveniente, utilizar muestras probabilísticas en futuros estudios sobre este problema. En segundo lugar, la cantidad de varianza explicada por la dimensión de Rasch es algo escasa. Estimamos que esta circunstancia se debe a la reducida dispersión en la dificultad de los ítems, toda vez que la varianza explicada depende conjuntamente de la dispersión de personas e ítems. En tercer lugar, cabría señalar algún solapamiento entre los ítems en cuanto a su dificultad (la diferencia en *logits* entre algunos de ellos es muy pequeña), si bien el contenido de dichos ítems alude a aspectos distintos del constructo evaluado, por lo que a priori podrían permanecer en la escala. En último lugar, no es claro que los todos los ítems utilizados representen adecuadamente las variables evaluadas, ni que los ítems en su conjunto posean suficiente validez de contenido; sería por ello deseable la construcción de nuevos ítems, que diesen cuenta de aspectos de las habilidades sociales no evaluados por la SPSS-R, y que fueran apropiados para su aplicación por observadores externos en población con discapacidad intelectual. Del mismo modo, sería deseable, de cara a futuras investigaciones, incorporar medidas de constructos relacionados, tales como las destrezas comunicativas (en el marco de la conducta adaptativa), y las dimensiones de calidad de vida, al objeto de comprobar cómo se relacionan con las habilidades sociales.

Referencias

- Abad, F. J., Olea, J., Ponsoda, V. y García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.
- Ayala, L. J. (2009). *The theory and practice of item response theory*. Nueva York, NY: Guilford.
- Bellack, A. S. (1983). Recurrent problems in the behavioral assessment of social skill. *Behavioral Research and Therapy*, 21, 29-41.
- Bielecki, J. y Swender, S. L. (2004). The assessment of social functioning in individuals with mental retardation: a review. *Behavior Modification*, 28, 694-708.
- Bond, T. G. y Fox, C. M. (2001). *Applying the Rasch model: fundamental measurement in the human sciences*. Mahwah, NJ: LEA.
- Caballo, V. E. (1993). *Manual de evaluación y entrenamiento de las habilidades sociales*. Madrid: Siglo XXI.
- Crocker, L. y Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.
- Embretson, S. E. y Hershberger, S. L. (1999). *The new rules of measurement*. Mahwah, NJ: LEA.
- Embretson, S. E. y McCollam, K. M. S. (2000). Psychometric approaches to understanding and measuring intelligence. En R. J. Sternberg (dir.), *Handbook of intelligence* (pp. 423-444). Cambridge: Cambridge University Press.
- Embretson, S. E. y Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: LEA.

- Fidalgo, A. M. (2005). Enfoque de la teoría de respuesta a los ítems. En J. Muñiz, A. M. Fidalgo, E. García-Cueto, R. Martínez y R. Moreno (dirs.), *Análisis de los ítems* (pp. 79-131). Madrid: La Muralla.
- Gómez, L. E., Verdugo, M. A., Arias, B. y Arias, V. (2010). A comparison of alternative models of individual quality of life for social service recipients. *Social Indicators Research, 101*, 109-126.
- Gómez, L. E., Verdugo, M. A. y Arias, B. (2010). Calidad de vida individual: avances en su conceptualización y retos emergentes en el ámbito de la discapacidad. *Behavioral Psychology/Psicología Conductual, 18*, 453-472.
- Gómez, L. E., Verdugo, M. A., Arias, B. e Irurtia, M. J. (2011). Evaluación de los derechos de las personas con discapacidad intelectual: estudio preliminar. *Behavior Psychology/Psicología Conductual, 19*, 207-222.
- Gresham, F. M. y Elliot, S. N. (1990). *Social Skills Rating System: manual*. Circle Pines, MN: American Guidance Service.
- Hambleton, R. K., Swaminathan, J. y Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- International Test Commission (2006, julio). *The ITC 5th International Conference on Psychological and Educational Test Adaptation across Language and Cultures*. Bruselas, Bélgica.
- Jacobson, J. W., Mulick, J. A. y Rojahn, J. (dirs.) (2007). *Handbook of intellectual and developmental disabilities*. Nueva York, NY: Springer.
- La Greca, A. M., Stone, W. L., Bell, D. J. y Charles, R. (1982). Assessing the problematic interpersonal skills of mentally retarded individuals in a vocational setting. *Applied Research in Mental Retardation, 3*, 37-53.
- Lecavalier, L. y Butter, E. M. (2010). Assessment of social skills and intellectual disability. En D. W. Nangle, D. J. Hansen, C. A. Erdley y P. J. Norton, (dirs.), *Practitioner's guide to empirically based measures of social skills* (pp.179-190). Nueva York, NY: Springer.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*, 103-122.
- Linacre, J. M. (2014). *WINSTEPS: Rasch measurement computer program*. Recuperado de <http://www.winsteps.com/winman/index.htm>.
- Linacre, J. M. (2008). *A user's guide to WINSTEPS-MINISTEP: Rasch-model computer programs*. Chicago. Diposible en: winsteps.com.
- Linacre, J. M. y Wright, B. D. (1999). *WINSTEPS: Multiple choice, rating scale, and partial credit Rasch analysis* [Programa informático]. Chicago: MESA.
- Lorenzo-Seva, U. y Ferrando, P. J. (2006). FACTOR: a computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers, 38*, 88-91.
- Lowe, M. R. y Cautela, R. (1978). A self-report measure of social skills. *Behavior Therapy, 9*, 535-544.
- Matson, J. L. (1995). *The Matson Evaluation of Social Skills for Individuals with Severe Retardation*. Baton Rouge, LA: Scientific Publishers.
- Matson, J. L. y Swiezy, N. (1994). *Social skills training with autistic children*. En J. L. Matson (dir.), *Autism in children and adults: etiology, assessment, and intervention* (pp. 241-260). Belmont, CA: Brooks/Cole.
- Matson, J. L., Helsel, W. J., Bellack, A. S. y Senatore, V. (1983). Development of a rating scale to assess social skills deficits in mentally retarded adults. *Applied Research in Mental Retardation, 4*, 399-407.
- Matson, J. L., Rotari, A. F. y Helsel, W. J. (1983). Development of a rating scale to measure social skills in children: the Matson Evaluation of Social Skills in Youths (MESSY). *Behaviour Research and Therapy, 21*, 335-340.

- Meyer, L. H., Reichle, J., McQuarter, R., Cole, D. A., Vandercook, T., Evans, I., Neel, R. y Kishi, G. (1985). *Assessment of Social Competence (ASC): a scale of social competence functions (revised)*. Syracuse, NY: Syracuse University.
- Muñiz, J., Elosua, P. y Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25, 151-157.
- Nangle, D. W., Hansen, D. J., Erdley, C. A. y Norton, P. J. (2010). *Practitioner's guide to empirically based measures of social skills*. Nueva York, NY: Springer.
- Preacher, K. J. y McCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13-43.
- Prieto, G. y Delgado, A. R. (1999). Medición cognitiva de las aptitudes. En J. Olea, V. Ponsoda y G. Prieto (dirs.), *Tests informatizados: fundamentos y aplicaciones* (pp. 207-226). Madrid: Pirámide.
- Prieto, G. y Delgado, A. R. (2000) Utilidad y representación en la psicometría actual. *Metodología de las Ciencias del Comportamiento*, 2, 111-127.
- Prieto, G. y Delgado, A. R. (2003) Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15, 94-100.
- Prieto, G. y Dias, A. (2004). Uso del modelo de Rasch para poner en la misma escala las puntuaciones de distintos tests. *Actualidades en Psicología*, 19, 5-23.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA Press.
- Rasch, G. (1977). On specific objectivity: an attempt at formalizing the request for generality and validity of scientific statements. En M. Blegvad (dir.), *The Danish yearbook of philosophy* (pp. 58-94). Copenhagen: Munksgaard.
- Schalock, R. L., Bonham, G. S. y Verdugo, M. A. (2008). Quality of life conceptual and measurement frameworks: from concept to application in the field of intellectual disabilities. *Evaluation y Program Planning*, 31, 181-190.
- Schalock, R. L., Keith, K. D., Verdugo, M. A. y Gómez, L. E. (2010). Quality of life model development in the field of intellectual disability. En R. Kober (dir.), *Enhancing quality of life for people with intellectual disability: from theory to practice* (pp. 17- 32). Nueva York, NY: Springer.
- Schalock, R. L. y Verdugo, M. A. (2002). *Handbook on quality of life for human service practitioners*. Washington, DC: American Association on Mental Retardation.
- Schalock, R. L. y Verdugo, M. A. (2007). El concepto de calidad de vida en los servicios y apoyos para personas con discapacidad intelectual. *Siglo Cero*, 38, 21-36.
- Schalock, R. L. y Verdugo, M. A. (2012a). *A leadership guide for today's disabilities organizations: overcoming challenges and making change happen*. Baltimore: Brookes.
- Schalock, R. L. y Verdugo, M. A. (2012b). A conceptual and measurement framework to guide policy development and systems change. *Journal of Policy and Practice in Intellectual Disabilities*, 7, 71-81.
- Timmerman, M. E. y Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16, 209-220.
- Van de Vijver, F. y Hambleton, R. K. (1996). Translating tests: some practical guidelines. *European Psychologist*, 1, 89-99.
- Wright, B. D. y Stone, M. H. (1979). *Best test design. Rasch measurement*. Chicago, IL: MESA.
- Wright, B. D. y Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA.

RECIBIDO: 12 de mayo de 2014

ACEPTADO: 22 de septiembre de 2014