

Vol. 23, 2021/e06

Evidencia de validez de la Escala de Autoeficacia Percibida Específica de Situaciones Académicas (EAPESA)

Validity Evidence of the Academic Situation-Specific Perceived Self-Efficacy Scale (EAPESA)

Angélica Garzón Umerenkova (1) <https://orcid.org/0000-0002-6813-5521>
Javier Gil Flores (2) <https://orcid.org/0000-0003-0755-4367>
Manuel Rafael de Besa Gutiérrez (3) <https://orcid.org/0000-0003-1552-2664>

(1) Fundación Universitaria Konrad Lorenz
(2) Universidad de Sevilla
(3) Universidad de Cádiz

(Recibido: 13 de abril de 2019; Aceptado para publicación: 4 de octubre de 2019)

Cómo citar: Garzón, A., Gil, J. y de Besa, M. R. (2021). Evidencia de validez de la Escala de Autoeficacia Percibida Específica de Situaciones Académicas (EAPESA). *Revista Electrónica de Investigación Educativa*, 23, e06, 1-12.
<https://doi.org/10.24320/redie.2021.23.e06.2979>

Resumen

Se presentan las características psicométricas de la Escala de Autoeficacia Percibida Específica de Situaciones Académicas (EAPESA) en alumnado universitario español mediante análisis Rasch. Se explora la validez de constructo, la confiabilidad de la medida y de las personas, el funcionamiento de la escala Likert y el comportamiento diferencial de los ítems (DIF). Se obtuvo evidencia de la validez discriminante frente a una medida de procrastinación académica y validez de criterio frente al rendimiento. La muestra estuvo constituida por 443 universitarios de tres titulaciones, entre primer y cuarto año. Se encontraron valores adecuados de confiabilidad y de la escala Likert, así como un buen ajuste al modelo Rasch. Sin embargo, se encontró evidencia de que se puede mejorar la validez de constructo y de un posible DIF por titulación en un ítem.

Palabras clave: autoeficacia, logro académico, pruebas psicológicas

Abstract

This study employs a Rasch analysis to present the psychometric properties of the Academic Situation-Specific Perceived Self-Efficacy Scale (EAPESA, in Spanish) in students at a Spanish university. Construct validity, measurement and person reliability, the functioning of the Likert scale, and differential item functioning (DIF) were all explored, producing evidence of discriminant validity in regard to a measurement of academic procrastination and criterion validity in relation to performance. The sample comprised 443 university students from the first to fourth years of three degree programs. Adequate reliability and Likert scale values were found, as was a good fit to the Rasch model. However, evidence was also found to suggest that construct validity could be improved and that DIF by degree program may be present in one item.

Keywords: self-efficacy, academic achievement, psychological tests

I. Introducción

El origen del concepto de autoeficacia debe buscarse en la teoría sociocognitiva de Bandura (1986). Según este autor, el logro del bienestar requiere contar con un sentido positivo sobre la eficacia personal, y define la autoeficacia como la expectativa de ejecutar cierta conducta adecuadamente, así como las expectativas de éxito al valorar la propia habilidad para alcanzar ciertos resultados (Bandura, 1997). La autoeficacia percibida, por lo tanto, es la creencia que el individuo tiene sobre su capacidad para actuar en determinados ámbitos de la actividad humana y hacerlo logrando sus objetivos. De acuerdo con este enfoque teórico, una elevada autoeficacia percibida conlleva en los individuos un mayor optimismo y una mayor motivación, que les impulsa a invertir más esfuerzo, de manera más continuada, para alcanzar sus metas.

Dado que la autoeficacia se refiere a la autovaloración que hacen los individuos sobre sus propias capacidades específicas, es posible diferenciarla de otros constructos similares referidos a aspectos globales de la persona sobre sí misma, como pueden ser la autoestima o el autoconcepto (Artino, 2012). Frente al constructo autoeficacia, la autoestima hace referencia a la autovaloración global de los individuos sobre sí mismos (Bandura, 1997), sin limitarse a la valoración de la capacidad para actuar de manera exitosa. En el caso del autoconcepto, se alude a las percepciones que el individuo tiene sobre sí mismo, que se fundamentan en sus experiencias sociales previas y sus percepciones sobre la propia conducta, físico u otros aspectos de su persona (Luna y Molero, 2013).

Por otra parte, a diferencia de medidas basadas en constructos de autovaloración globales, la autoeficacia implica la autovaloración específica de la expectativa de ser competente, o no, para realizar con éxito una tarea concreta en un determinado contexto. Centrándonos en el contexto que configuran las situaciones académicas, se ha investigado ampliamente la autoeficacia profesional de los docentes (Zee y Koomen, 2016), entendiéndola como la confianza del profesor para afrontar adecuadamente las tareas de enseñanza y lograr el aprendizaje en sus estudiantes. La atención de los investigadores también se ha centrado en la autoeficacia percibida por el alumno (Artino, 2012). En relación con los estudiantes, la autoeficacia académica remite a las convicciones de los individuos de que pueden realizar con éxito las tareas académicas que les son propuestas y alcanzar las metas educativas (Feldman y Kubota, 2015).

La autoeficacia académica percibida por el alumnado es el tema en el cual se centra este trabajo. El interés por la autoeficacia académica reside en su probada asociación con el desempeño de los estudiantes; de acuerdo al estudio de Schneider y Preckel (2017) sobre 105 variables predictoras analizadas, la autoeficacia fue la segunda variable más fuerte del desempeño académico. Maddux (2016) y Valentine et al. (2004) han aportado evidencia de que una autoeficacia positiva facilita la implicación de los estudiantes, el establecimiento de metas, la motivación intrínseca, la inversión de esfuerzo en las tareas académicas, la adaptación al contexto universitario, la persistencia, el rendimiento o incluso la elección de carrera.

Posiblemente la relación entre autoeficacia y rendimiento académico sea uno de los aspectos que más interés ha suscitado en la literatura científica. Multon et al. (1991) llevaron a cabo un análisis sobre 39 estudios acerca de esta relación, publicados entre 1977 (año en que Bandura introdujo su teoría sobre la autoeficacia) y 1988. Los resultados obtenidos mediante técnicas meta-analíticas revelaron una relación positiva y significativa de la autoeficacia percibida con el rendimiento y la persistencia académica.

Trabajos posteriores han arrojado resultados consistentes en el mismo sentido. Uno de estos trabajos es la revisión de Honicke y Broadbent (2016) sobre 59 estudios publicados entre 2003 y 2015, en la que se corroboró una correlación moderada entre la autoeficacia académica y el rendimiento de los estudiantes universitarios. En un reciente metanálisis de Talsma et al. (2018) se seleccionaron 11 estudios publicados entre 1984 y 2016, en los que se midió la autoeficacia académica no como un rasgo general, sino con base en preguntas sobre la confianza en la propia habilidad para ejecutar ciertas tareas académicas; el rendimiento se operacionalizó a partir de las calificaciones obtenidas por los estudiantes. El objetivo de este metanálisis fue estudiar el sentido de la relación entre ambas variables. Las conclusiones alcanzadas indican efectos positivos y significativos de la autoeficacia sobre el rendimiento y del rendimiento académico sobre la autoeficacia, si bien este segundo efecto resultó ser significativamente mayor.

Además de la relación entre autoeficacia académica con el rendimiento, también existe evidencia de su relación con otras variables académicas relevantes. De acuerdo con la teoría social de Bandura (1997), si la expectativa es baja podría verse afectada la motivación para iniciar la tarea y alcanzar un propósito, lo que puede llevar a una conducta de evitación que redundaría en procrastinación. Varios estudios aportan evidencia de que entre la autoeficacia y la procrastinación existe una relación inversa (Steel, 2007). La baja autoeficacia se ha empleado para explicar la procrastinación, ya que cuando los estudiantes tienen una baja expectativa de éxito en sus tareas académicas tienden a aplazar con mayor frecuencia el inicio de las mismas (Steel y Klingsieck, 2015). En un estudio sobre las relaciones entre la procrastinación, la percepción de logro y la autoeficacia, se encontró que existe una especie de círculo vicioso entre la procrastinación y la autoeficacia; en el cual la autoeficacia media el efecto de la percepción de logro sobre la procrastinación. Por ello los estudiantes que tienen baja autoeficacia percibida son más susceptibles de entrar en el círculo vicioso de la procrastinación (Wäschle et al., 2014).

En cuanto a la relación entre autoeficacia, procrastinación y autorregulación académica, se ha reportado que la baja autoeficacia para la autorregulación es predictiva de la tendencia a procrastinar, en comparación a otras medidas "auto", como la autoestima. En ese sentido, la autorregulación y la autoeficacia serían determinantes de un patrón de procrastinación y otras variables que, en conjunto, afectarían el rendimiento académico (Vinothkumar et al., 2016).

1.1 Medida de autoeficacia EAPESA

La literatura científica recoge diferentes medidas de autoeficacia, entre ellas la Escala de Autoeficacia Percibida Específica de Situaciones Académicas (EAPESA), objeto de estudio del presente trabajo. De acuerdo con el autor del instrumento (Palenzuela, 1983), en su concepción original la EAPESA se diferencia de otros constructos relacionados con la autoeficacia, para los cuales existían pruebas al uso. Uno de estos constructos es la indefensión, ya que una persona puede estar segura de sus capacidades y sin embargo sentirse indefensa (Bandura, 1986, 1997). Otros conceptos de los cuales buscó alejarse en el proceso de construcción la EAPESA son el de control y el de motivación de competencia, centrándose exclusivamente en la definición original de autoeficacia percibida de Bandura (1986) aplicada al contexto académico. Por lo tanto, los ítems de la escala de autoeficacia percibida se refieren exclusivamente a si el individuo "espera (expectativa) o cree que puede llegar a ser eficaz y competente en su vida académica" (Palenzuela, 1983, p. 188).

La EAPESA originalmente fue construida y validada con un grupo de estudiantes de secundaria y universitarios de la Universidad de Salamanca (España). La escala estaba constituida por diez ítems con un coeficiente de fiabilidad de Tucker de .92; se encontró evidencia de la unidimensionalidad del constructo, correlaciones positivas con una medida de motivación al logro, autoestima, autodeterminación y motivación por competencia.

Posteriormente, en un estudio realizado por García-Fernández et al. (2010), con estudiantes de secundaria españoles se calculó el Alfa de Cronbach (.89) y fiabilidad test-retest (.87), y se obtuvieron relaciones significativas con medidas de autoeficacia académica, metas académicas, autoconcepto y rendimiento. En otro estudio realizado por García-Fernández et al. (2016) con adolescentes chilenos se encontró un Alfa de Cronbach de .88, un único factor y relaciones con la motivación, gestión del tiempo y estrategias de estudio y evaluación.

1.2 Análisis Rasch

Como se ha descrito en la anterior sección, hasta ahora el análisis de las características de la EAPESA fue realizado por su autor (Palenzuela, 1983) y en estudios posteriores (Domínguez et al., 2012; García-Fernández et al., 2010; García-Fernández et al., 2016) desde enfoques propios de la Teoría Clásica de los Tests (TCT).

A diferencia de tales aproximaciones, en este trabajo abordamos el análisis de la EAPESA mediante el modelo logístico de un parámetro de Rasch. Con el análisis Rasch es posible encontrar evidencia adicional sobre características psicométricas, que son difíciles de establecer con otras técnicas estadísticas. En este sentido, permitiría establecer evidencia sobre si los ítems cubren el rango completo del atributo medido, si la escala Likert es apropiada o si la medida diferencia personas de alta y baja habilidad en el atributo.

Hay diversas ventajas que en la literatura se atribuyen al uso del análisis Rasch, tales como que mediante este enfoque analítico es posible refinar o mejorar las pruebas, lo que conlleva mejorar la evaluación aplicada (Brandt et al., 2015; Heesch et al., 2006) y permitir la obtención de pruebas más cortas y efectivas (Settanni et al., 2015).

A diferencia de otras formas de análisis estadístico, el análisis Rasch busca determinar el grado en el cual los datos obtenidos se ajustan a las expectativas del modelo y no al contrario (Smith, 2012). El modelo Rasch se fundamenta básicamente en dos principios: la unidimensionalidad y la independencia local. La unidimensionalidad permite la estimación de la existencia de un único factor principal del instrumento y la independencia local indica que las respuestas de las personas a cualquier pregunta son independientes de sus respuestas a otra pregunta. Mientras el principio de unidimensionalidad se comprueba con los criterios específicos explicados más adelante, el principio de independencia local se comprueba en la configuración misma de las salidas del modelo, no existen unos análisis específicos para su comprobación.

Mediante la escala en Loguitos (*logits scale*) el modelo representa la habilidad del individuo, quien responde a los ítems de la prueba en diferentes magnitudes de dificultad (Bond y Fox, 2015). En este caso la habilidad, es decir, el atributo latente a la medida, se interpreta como la *percepción de autoeficacia académica*, de acuerdo al componente específico para el que fue construida la medida.

El presente estudio busca proveer un análisis de las propiedades psicométricas de la EAPESA para estudiantes universitarios españoles, a fin de constatar: la dimensionalidad, el ajuste de los ítems al modelo, el funcionamiento de la escala Likert de respuesta, la validez de constructo, la confiabilidad y el funcionamiento diferencial de los ítems (DIF) por género y titulación.

II. Método

Se realizó un muestreo por conveniencia, los participantes fueron 443 estudiantes que en el curso 2017-2018 se encontraban matriculados en la Universidad de Cádiz (España), en las titulaciones de Grado en Psicología (37.5%), Grado en Educación Primaria (44.9%) y Grado en Educación Infantil (17.6%). Del total de participantes, 350 (79%) eran mujeres y 93 (21%) hombres. La media de edad se situó en los 21.57 años (DT 3.99) con un rango de 18 a 51 años. Un 39% de los participantes era alumnado de nuevo ingreso que cursaba el primer año de estudios, mientras que el 61% restante se distribuyó entre los años segundo y cuarto. Además, 123 (27.8%) reportaron estar realizando un trabajo remunerado al tiempo que cursaban los estudios universitarios, con una media de 13.6 horas de dedicación por semana, el 13.5% reportó planificar su tiempo nunca o casi nunca, el 35.4% reportó planificarlo a veces y el 50.8% reportó planificar su tiempo casi siempre o siempre.

El instrumento utilizado fue la escala EAPESA, desarrollada por Palenzuela (1983) para medir autoeficacia percibida en contextos académicos y, como se mencionó en la introducción, se basó en la definición de Bandura (1986) en el marco de la teoría del aprendizaje social.

En un estudio realizado por Domínguez et al. (2012) con estudiantes universitarios peruanos, se encontró una única dimensión que explica el 55% de la varianza total del instrumento y un Alfa de Cronbach de .89. A partir de los resultados de dicho estudio, los autores recomendaron la eliminación del ítem 9 (Soy de esas personas que no necesitan estudiar para aprobar una asignatura o pasar un curso completo de la universidad) por razones empíricas y de contenido, ya que el ítem no alcanzó el índice de homogeneidad mínimo. Palenzuela (1983) ya había detectado algunas dificultades con este ítem, cuya eliminación aumentaba la fiabilidad de la escala; en cuanto al contenido, este ítem presenta un posible sentido

ambiguo, ya que no define aspectos de la autoeficacia percibida sino que hace referencia a la aparente falta de necesidad de estudiar para aprobar las asignaturas. Por estas razones, para los fines del presente estudio se optó por emplear la versión de 9 ítems, sin emplear el ítem numerado originalmente como "9".

Adicionalmente se aplicó una encuesta sociodemográfica que incluía, entre otras, preguntas sobre la edad, género, titulación, o el desempeño de un trabajo remunerado durante los estudios universitarios cursados, además del autoinforme del participante acerca del rendimiento académico logrado en el primer cuatrimestre del curso 2017-2018.

La recogida de datos se realizó en el segundo cuatrimestre del curso académico 2017-2018, concretamente durante el mes de marzo de 2018. De este modo, en el momento de aplicar los instrumentos de recogida de datos, el alumnado ya contaba con resultados académicos en forma de calificaciones correspondientes a las materias cursadas durante el primer cuatrimestre. Antes de la aplicación se contactó con profesorado que se impartía asignaturas en las titulaciones de Grado en Psicología, Educación Infantil y Educación Primaria, con el objetivo de conseguir su colaboración. Se buscó que en la muestra estuvieran presentes estudiantes de primer año y alumnado que contaba ya con uno o más años de experiencia en sus estudios universitarios. La aplicación de los cuestionarios la realizaron los investigadores dentro del aula donde habitualmente se desarrollan las clases. La duración aproximada fue de 15 minutos, incluyendo una breve presentación del estudio, la explicación sobre los instrumentos y el modo de consignar las respuestas por parte del alumnado. Los participantes aceptaron voluntariamente tomar parte en el estudio y se les garantizó el anonimato y la confidencialidad de sus respuestas. No hubo compensaciones por participar y se garantizó que no habría consecuencias negativas para quienes declinaran participar.

El análisis Rasch se realizó mediante el paquete estadístico Winsteps versión 3.72.3. En primer lugar se llevó a cabo un análisis de bondad de ajuste de los datos empíricos al modelo, teniendo en cuenta su dimensionalidad y el ajuste al modelo de cada ítem. Luego se establecieron el parámetro de dificultad b , la confiabilidad tanto para personas como para los ítems, el funcionamiento de las categorías de respuesta y, por último, un análisis del comportamiento diferencial de los ítems (DIF) por género y titulación.

III. Resultados

3.1 Bondad de ajuste al modelo Rasch

Dimensionalidad. En la Tabla 1 se muestran los resultados de los análisis realizados con el objetivo de valorar el supuesto de unidimensionalidad para la escala de autoeficacia percibida. Para su comprobación se pueden tener en cuenta tres criterios para establecer o descartar la posibilidad de la existencia de una única dimensión latente mediante el análisis de componentes principales:

- Primer criterio, la prueba tiene una dimensión si la varianza explicada por la medida está en los siguientes rangos: $\geq 40\%$ (suficiente); $\geq 30\%$ (moderada) y $\geq 20\%$ (mínima aceptable) (Linacre, 2012). De acuerdo con los resultados (ver Tabla 1) la escala presenta una varianza explicada por la medida superior al 40% (59.6%), siendo suficiente.
- Segundo criterio, comprobar si la cantidad de varianza explicada por el primer contraste de los residuales no es superior a la varianza explicada por la dificultad de los ítems (varianza explicada por los ítems en la Tabla 1); en este caso la varianza explicada por el primer contraste (8%) no es superior a la varianza explicada por los ítems (16.9%), con lo cual se cumpliría con el criterio.
- Tercer criterio, observar que el primer contraste de los residuales de la varianza bruta sin explicar sea inferior a 2 valores propios (Smith, 2012), ya que si tiene un valor superior indicaría que hay al menos dos ítems que estarían apuntando dimensionalmente en un sentido diferente. Como puede observarse en la Tabla 1, el primer contraste de los residuales es inferior a dos valores propios (1.78) con lo cual se cumpliría con el tercer criterio.

De acuerdo con estos resultados, se puede afirmar que la escala está constituida por una única dimensión.

Tabla 1. Varianza de residuos estandarizados

Autoeficacia Percibida	Valores propios	Observada	Esperada
Total de varianza bruta	22.25	100%	100%
Varianza bruta explicada por las medida	13.25	59.6%	59.3%
Varianza bruta explicada por las personas	9.49	42.7%	42.5%
Varianza bruta explicada por los ítems	3.75	16.9%	16.8%
Varianza bruta sin explicar (total)	9.00	40.4%	40.7%
Varianza bruta sin explicar en 1er contraste	1.78	8.0%	19.8%

Nota. Se esperan valores similares en los porcentajes de la varianza bruta observada y esperada.

Ajuste al modelo de los ítems. Teniendo en cuenta los resultados de ajuste (Tabla 2) todos los ítems presentan un buen ajuste al modelo, ya que sus valores están dentro de los parámetros establecidos para el MNSQ (Media Cuadrática) entre .5 y 1.5.

Tabla 2. Estimaciones del infit y outfit para cada ítem

Ítem	Medida	Modelada SE	MNSQ INFIT	MNSQ OUTFIT	PT-Medida Corr.	PT-Medida Exp.
5	1.55	.10	1.47	1.50	.71	.77
7	-2.04	.10	1.05	1.09	.69	.71
8	.92	.10	1.07	1.08	.77	.77
1	-.25	.10	.90	.95	.75	.76
3	.32	.10	.94	.93	.77	.76
2	-.03	.10	.93	.91	.76	.76
4	.16	.10	.88	.86	.79	.76
6	-.68	.10	.85	.80	.78	.75
9	.05	.10	.82	.81	.81	.76
Media	.00	.10	.99	.99		
Dt.	.95	.00	.19	.20		

Notas. Los ítems están ordenados de acuerdo al ajuste al modelo, del menos ajustado al más ajustado. Modelada SE (Error Estándar Modelado, siendo la precisión de la replicación de la medida y no debe exceder el valor de 3.5); MNSQ (Media Cuadrática), PT-Medida Corr. (correlaciones entre ítem y habilidad de las personas) y PT-Medida Exp. (correlaciones esperadas entre ítem y habilidad de las personas).

El modelo calcula una escala en loguitos (*logits scale*) y se toman como indicadores de ajuste valores MNSQ del infit y outfit entre .5 y 1.5 loguitos (Bond y Fox, 2015) con un valor esperado de 1. Valores superiores a 1.5 indicarían que el ítem es errático y valores por debajo de .5 indicarían que el ítem es muy predecible; los valores superiores a 2 son una amenaza potencial a la calidad de la medida (Linacre, 2012).

En la Tabla 2 también puede comprobarse que no hay correlaciones negativas entre el ítem y la habilidad (columna PT-Medida Corr.). De acuerdo al parámetro del modelo, se espera que la habilidad de las personas correlacione con la dificultad de los ítems. Por lo tanto, este dato indica la alineación entre el ítem y la habilidad; cuanto más alto el valor mejor, y los valores negativos implicarían una revisión de los resultados, debido a que habría una violación de la predicción del modelo basado en la alineación entre habilidad y dificultad (Garzón y Gil, 2017).

De acuerdo con los resultados, los valores de las correlaciones tienden a ser altas, siendo el menor valor .69 para el ítem 7 y el de mayor valor .81 para el ítem 9. El valor positivo y elevado de todas las correlaciones

permite afirmar que los ítems se encuentran alineados en la misma dirección de la variable latente.

Por otra parte, en la columna PT-Medida Exp. puede comprobarse que los valores de las correlaciones observadas son muy cercanos a las expectativas del modelo (Tabla 2), indicando nuevamente que el ajuste al modelo es adecuado y está en la dirección esperada.

3.2 Confiabilidad de la medida y de las personas

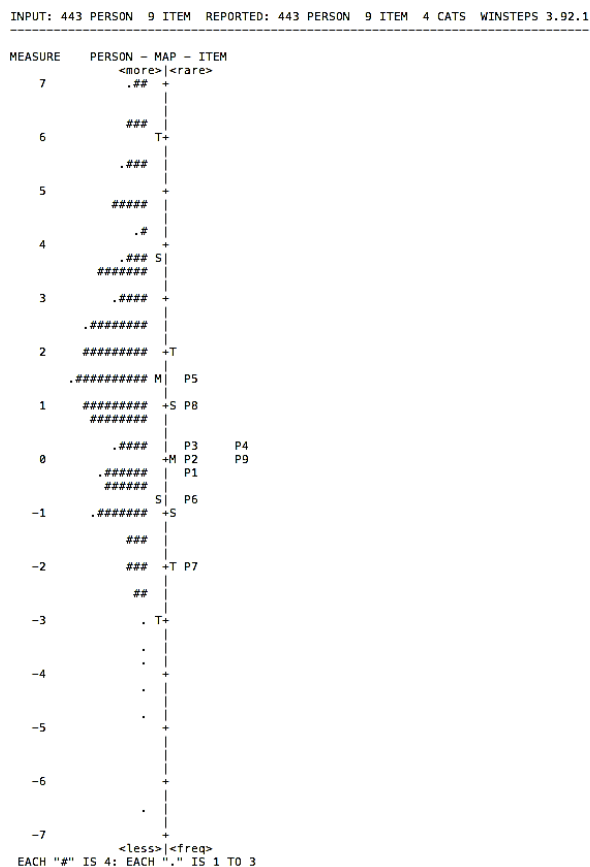
La confiabilidad de los ítems fue de .99 y la confiabilidad para las personas de .89, ambos coeficientes se interpretan como un Alfa de Cronbach. En ese sentido, los valores de confiabilidad para las personas y los ítems obtenidos para la prueba son adecuados.

En cuanto a la separación para para las personas (valor de 2.81), se considera un índice adecuado cuando se obtienen valores a partir de 2. En este caso, el valor 2.81 indica que el instrumento es lo suficientemente sensible para identificar personas con alta y baja habilidad en la variable medida, al menos para separar dos grupos (Smith, 2012). La separación para los ítems (valor de 9.20) es adecuado, lo que indica que la muestra es lo suficientemente grande para confirmar la jerarquía de dificultad de los ítems, es decir, la validez de constructo del instrumento (Smith, 2012).

3.3 Estimación e interpretación del parámetro *b*

En cuanto a evidencia de validez adicional, el orden de los ítems aporta evidencia adicional a los investigadores sobre qué tan bien están definiendo el atributo (Figura 1).

Figura 1. Mapa Wright de personas e ítems para los 9 ítems



El modelo de Rasch establece la evidencia de validez de constructo de acuerdo a la jerarquía de los ítems y el rango en que la distribución de las personas se alinea con los ítems; lo cual se puede observar en el Mapa Wright que muestra a la derecha la distribución de los ítems y a la izquierda la distribución de las personas (Figura 1). En este caso el ítem más difícil es el ítem 5 (No me importa que los profesores sean exigentes y duros, pues confío en mi propia capacidad académica). De esta manera se puede observar si el orden predicho de la dificultad de los ítems coincide con el orden resultante tras el análisis de los datos con el modelo y tales comparaciones aportan evidencia sobre qué tan bien se ajusta el comportamiento de la medida con lo que predice la teoría (Boone, 2016).

El mapa sitúa a las personas con mayor habilidad (atributo) en la parte superior y en la parte inferior a quienes presentan menor habilidad. Sin embargo, de acuerdo a los resultados del mapa (Figura 1) aunque se constata una adecuada distribución de los ítems, estos son insuficientes para cubrir el rango de habilidad de los individuos, principalmente en los niveles más elevados de autoeficacia.

Como se aprecia en la parte superior del mapa, hay personas con un nivel en el rasgo latente medido que se sitúa bastante por encima de los parámetros de dificultad para los ítems.

En consecuencia, EAPESA se beneficiaría de la inclusión de ítems de mayor nivel de dificultad para el atributo medido y mejoraría así las evidencias de validez relacionadas con el constructo en tanto cubra un mayor rango del atributo.

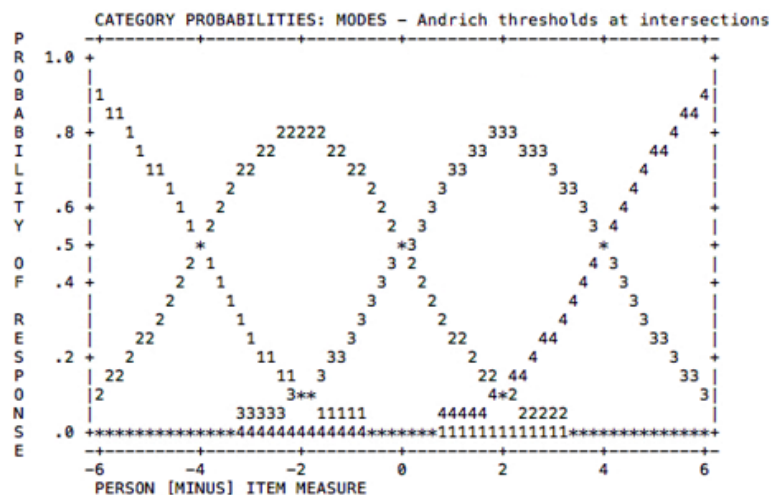
3.4 Funcionamiento de las categorías de respuesta

Las categorías de respuesta son: nunca (1), algunas veces (2), bastantes veces (3) y siempre (4). Mediante el Rating Scale Model (RSM) para escalas de ítems politómicos que permite comprobar el orden de las categorías y que cada una de ellas genere una clara diferenciación.

El modelo Rasch politómico para múltiples categorías se construye a partir de las aplicaciones sucesivas del modelo Rasch dicotómico, aplicando condicionalmente a los pares sucesivos de categorías adyacentes.

Las Curvas de Probabilidad de las Categorías de Andrich (Figura 2) muestran que las cuatro categorías presentaron un orden correcto (la sucesión va del número 1 al 4) y cada una de ellas se diferencia a lo largo de las mediciones del atributo, obteniendo crestas diferenciadas. Puede concluirse que la categorización de la escala Likert de respuesta es satisfactoria.

Figura 2. Curvas de Probabilidad de las Categorías



3.5 Análisis del funcionamiento diferencial de los ítems (DIF)

En el presente estudio se ha comprobado la invarianza de la medida de autoeficacia académica entre hombres y mujeres, y entre las titulaciones. No se encontraron valores que indicaran un posible DIF por género, ya que ninguno de los ítems superó los valores de "DIF contraste" por encima de .5 loguitos. El criterio empleado (.5 loguitos) hace referencia a la dificultad de un ítem entre los grupos. Por otra parte, se tuvo en cuenta que los valores t fueran superiores a 2 y con diferencias significativas ($p \leq .05$) (Bond y Fox, 2015).

Por titulación, para el ítem 7 (Tabla 3) se presenta evidencia de un posible DIF. El ítem 7 es más fácil para la titulación 2 (Grado en Educación Primaria) que para la 3 (Grado en Educación Infantil).

Tabla 3. Resumen del análisis del comportamiento diferencial de los ítems por titulación

Ítem	Titulación	Contraste DIF	t	Valor p
7	2-3	.60	2.06	0.0416

Nota. Los valores del contraste DIF indican que el ítem tiene más dificultad para los términos con valor positivo.

IV. Discusión y conclusiones

El presente estudio tuvo como objeto el análisis de las características psicométricas de la EAPESA mediante un análisis Rasch. Este instrumento no había sido previamente validado y calibrado psicométricamente mediante enfoques de análisis enmarcados en la Teoría de Respuesta al Ítem (TRI).

Se obtuvieron los principales indicadores psicométricos de confiabilidad y evidencia de validez, teniendo en cuenta que no se habían realizado estudios recientes de la EAPESA en población universitaria española.

Se encontraron valores ajustados al modelo Rasch, lo que indica evidencia de validez de constructo y un buen funcionamiento de cada uno de los ítems. No obstante, la escala podría beneficiarse de la incorporación de ítems que midan mayores grados o niveles del atributo (autoeficacia percibida) a fin de aumentar su sensibilidad en personas con alta autoeficacia percibida y, con ello, su validez de constructo.

Las categorías de respuesta (del 1 al 4) funcionan adecuadamente y tienden a discriminar a lo largo del atributo medido.

En cuanto a la confiabilidad, tanto para la medida como para las personas es adecuada. El análisis Rasch aporta datos adicionales de confiabilidad para las personas, que usando otros modelos estadísticos no es posible estimar claramente. En este caso, el valor de separación para las personas fue de 2.81, lo que indica que la EAPESA puede diferenciar claramente entre dos grupos de personas de acuerdo a los niveles del atributo "autoeficacia percibida".

Este tipo de confiabilidad para las personas es importante en el contexto de la evaluación educativa a nivel práctico, ya que nos ofrece información sobre el grado de sensibilidad o capacidad de discriminación del instrumento para diferenciar a las personas de acuerdo al grado en el que poseen el atributo medido.

Por otro lado, al analizar el posible comportamiento diferencial de los ítems, no se encontró DIF por género y sólo el ítem 7 (Si me lo propongo, creo que tengo la suficiente capacidad para obtener un buen récord académico) presenta alguna evidencia sobre un posible funcionamiento diferencial por titulación. Pero dado que la existencia de algunos valores DIF no implica necesariamente un sesgo en el ítem, se deberá seguir investigando para establecer si existe realmente un sesgo mediante la identificación de los factores relacionados con el posible DIF, por ejemplo, los procesos cognitivos que podrían estar implicados y una evaluación sistemática de hipótesis de DIF a partir de consideraciones teóricas o empíricas (Hagquist y Andrich, 2017).

Siendo una escala corta y de fácil aplicación, puede ser de utilidad para fines del diagnóstico educativo y en el desarrollo de planes para el mejoramiento académico. La precisión y discriminación de la medida, así como su relación con otras variables como el rendimiento académico, favorecen el empleo de la escala para diversos propósitos educativos.

Aunque el trabajo desarrollado cuenta entre sus fortalezas con el enfoque de análisis adoptado y con la amplitud de la muestra utilizada, es posible señalar también algunas debilidades. Entre sus limitaciones se encuentra la restricción de la muestra a una única institución educativa y la falta de heterogeneidad que supone haber trabajado sólo con titulaciones del ámbito de la psicología y la educación. De la misma manera, habría sido conveniente la inclusión de otras medidas relevantes a fin de establecer evidencia adicional de validez de criterio o predictiva.

Una línea interesante de cara a futuros estudios es el desarrollo de la escala a partir de la inclusión de nuevos ítems relacionados con la autoeficacia percibida, con el propósito de ampliar el rango de medida del atributo, particularmente con ítems de mayor dificultad.

Referencias

- Artino, A. R. (2012). Academic self-efficacy: from educational theory to instructional practice. *Perspectives on Medical Education*, 1(2), 76-85. <https://doi.org/10.1007/s40037-012-0012-5>
- Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. W. H. Freeman and Company.
- Bond, T. G. y Fox, C. M. (2015). *Applying the Rasch Model: fundamental measurement in the human sciences*. Routledge. <https://doi.org/10.4324/9781315814698>
- Boone, W. J. (2016). Rasch Analysis for instrument development: why, when, and how? *CBE—Life Sciences Education*, 15(4), 1-7. <https://doi.org/10.1187/cbe.16-04-0148>
- Brandt, S., Moulton, M. y Duckor, B. (2015). Advances in Rasch modeling: new applications and directions: guest editorial. *Psychological Test and Assessment Modeling*, 57(3), 338-341.
- Domínguez, S., Villegas, G., Yauri, C., Mattos, E. y Ramírez, F. (2012). Propiedades psicométricas de una escala de autoeficacia para situaciones académicas en estudiantes universitarios peruanos. *Revista de Psicología- Universidad Católica San Pablo*, 2(1), 27-39. <https://revistas.ucsp.edu.pe/index.php/psicologia/article/view/8>
- Feldman, D. B. y Kubota, M. (2015). Hope, self-efficacy, optimism, and academic achievement: distinguishing constructs and levels of specificity in predicting college grade-point average. *Learning and Individual Differences*, 37, 210-216. <https://doi.org/10.1016/j.lindif.2014.11.022>
- García-Fernández, J. M., Inglés, C. J., Torregrosa, M. S., Ruiz-Esteban, C., Díaz-Herrero, A., Pérez-Fernández, E. y Martínez-Monteagudo, M. C. (2010). Propiedades psicométricas de la Escala de Autoeficacia Percibida Específica de Situaciones Académicas en una muestra de estudiantes españoles de Educación Secundaria Obligatoria. *European Journal of Education and Psychology*, 3(1), 61-74. <https://doi.org/10.30552/ejep.v3i1.51>
- García-Fernández, J. M., Inglés, C. J., Vicent, M., González, C., Pérez Sánchez, A. M. y Lagos San Martín, N. (2016). Validación de la Escala de Autoeficacia Percibida Específica de Situaciones Académicas en Chile y su relación con las estrategias de aprendizaje. *Revista Iberoamericana de Diagnóstico y Evaluación*, 41(1), 118-131.

- Garzón, A. y Gil, J. (2017). Propiedades psicométricas del TMBS en universitarios. *Revista Electrónica de Investigación Educativa*, 19(4), 50-60. <https://doi.org/10.24320/redie.2017.19.4.1340>
- Hagquist, C. y Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes*, 15(181), 1-8. <https://doi.org/10.1186/s12955-017-0755-0>
- Heesch, K. C., Masse, L. C. y Dunn, A. L. (2006). Using Rasch modeling to re-evaluate three scales related to physical activity: enjoyment, perceived benefits and perceived barriers. *Health Education Research*, 21(1), 58-72. <https://doi.org/10.1093/her/cyl054>
- Honicke, T. y Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: a systematic review. *Educational Research Review*, 17, 63-84. <https://doi.org/10.1016/j.edurev.2015.11.002>
- Linacre, J. M. (2012). Winsteps® Rasch Tutorial 4. <http://www.winsteps.com/a/winsteps-tutorial-4.pdf>
- Luna, N. C. y Molero, D. (2013). Revisión teórica sobre el autoconcepto y su importancia en la adolescencia. *Revista Electrónica de Investigación y Docencia*, (10), 43-64. <https://revistaselectronicas.ujaen.es/index.php/reid/article/view/991/818>
- Maddux, J. E. (2016). Self-efficacy. En S. Trusz y P. Babel, *Interpersonal and Intrapersonal Expectancies* (pp. 41-46). Routledge. <https://doi.org/10.4324/9781315652535>
- Multon, K. D., Brown, S. D. y Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: a meta-analytic investigation. *Journal of Counseling Psychology*, 38(1), 30-38. <https://doi.org/10.1037/0022-0167.38.1.30>
- Palenzuela, D. (1983). Construcción y validación de una escala de autoeficacia percibida específica de situaciones académicas. *Análisis y Modificación de Conducta*, 9(21), 185-219. <http://dx.doi.org/10.33776/amc.v9i21.1649>
- Settanni, M., Longobardi, C., Sclavo, E., Fraire, M. y Prino, L. E. (2015). Development and psychometric analysis of the student-teacher relationship scale-short form. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00898>
- Schneider, M. y Preckel, F. (2017). Variables associated with achievement in higher education: a systematic review of meta-analyses. *Psychological Bulletin*, 143(6), 565-600. <https://doi.org/10.1037/bul0000098>
- Smith, E. (2012). *Practical Rasch Measurement-Core Topics* (tutorial 3). Statistics.com.
- Steel, P. (2007). The nature of procrastination: a meta-analytic and theoretical review of quintessential selfregulatory failure. *Psychological Bulletin*, 133(1), 65-94. <https://doi.org/10.1037/0033-2909.133.1.65>
- Steel, P. y Klingsieck, K. (2015). Procrastination. En J. D. Wright (Ed.), *The international encyclopedia of the social y behavioral sciences* (pp. 73-78). Elsevier.
- Talsma, K., Schüz, B., Schwarzer, R. y Norris, K. (2018). I believe, therefore I achieve (and vice versa): a meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences*, 61, 136-150. <https://doi.org/10.1016/j.lindif.2017.11.015>
- Valentine, J. C., DuBois, D. L. y Cooper, H. (2004). The relation between self-beliefs and academic achievement: a meta-analytic review. *Educational Psychologist*, 39(2), 111-133.

https://doi.org/10.1207/s15326985ep3902_3

Vinothkumar, M., Vindya, K. y Rai, V. (2016). Moderating roles of hardiness and self-efficacy in the relationship between flow and academic procrastination on academic performance: a structural equation model approach. *The International Journal of Indian Psychology*, 3(2), 77-89.

<https://doi.org/10.25215/0302.044>

Wäschle, K., Allgaier, A., Lachner, A., Fink, S. y Nückles, M. (2014). Procrastination and self-efficacy: tracing vicious and virtuous circles in self-regulated learning. *Learning and instruction*, 29, 103-114.

<https://doi.org/10.1016/j.learninstruc.2013.09.005>

Zee, M. y Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being. *Review of Educational Research*, 86(4), 981-1015.

<https://doi.org/10.3102/0034654315626801>