

**EVIDENCIA DE LA VALIDEZ DEL CONTENIDO DE LA ESCALA DE ACTITUD
HACIA LAS EVALUACIONES EXTERNAS A GRAN ESCALA (EAAE)**

***EVIDÊNCIAS DE VALIDADE DE CONTEÚDO DA ESCALA DE ATITUDES
PERANTE AS AVALIAÇÕES EXTERNAS APLICADAS EM LARGA ESCALA (EAAE)***

***EVIDENCE OF CONTENT VALIDITY OF THE SCALE OF ATTITUDES TOWARDS
EXTERNAL ASSESSMENTS APPLIED ON A LARGE SCALE (EAAE)***

Denilson Junio Marques SOARES¹
Talita Emidio Andrade SOARES²
Wagner dos SANTOS³

RESUMEN: Este artículo presenta el proceso de construcción y evidencia de validez de contenido de la Escala de Actitudes hacia las Evaluaciones Externas aplicadas a gran escala (EAAE). Este instrumento fue desarrollado con el fin de captar lo que piensan, sienten y se comportan los docentes de educación básica a través de este tipo de evaluación. Se trata de un estudio con enfoque cuantitativo, estructurado en dos maneras: evaluación por jueces expertos, mediante el cálculo del Coeficiente de Validez de Contenido (CVC); y aplicación previa, realizada con una muestra de la población objeto de estudio. La versión final de la escala contó con 30 ítems, cuyo CVC global fue de 0,92, evidenciando su validez de contenido y señalando la claridad del lenguaje, la relevancia práctica y teórica del instrumento. Se espera que este trabajo aporte datos importantes para futuras investigaciones, con miras a continuar el proceso de validación de la EAAE, y en consecuencia, posibilitar su uso en estudios acerca de la evaluación educativa.

PALABRAS CLAVE: Escala de actitudes. Evaluaciones externas a gran escala. Validez de contenido. Evaluación educativa. Psicometría.

RESUMO: Este estudo teve como objetivo desenvolver e verificar as evidências de validade de conteúdo da Escala de Atitudes perante as Avaliações Externas aplicadas em larga escala (EAAE). Esse instrumento foi construído de modo a capturar o que os professores da educação básica pensam, sentem e como se comportam mediante esse tipo de avaliação. Trata-se de um estudo com abordagem quantitativa, estruturado em duas vertentes: avaliação de juízes especialistas com base no cálculo do Coeficiente de Validade de Conteúdo (CVC); e aplicação piloto, realizada com uma amostra da população-alvo. A versão final da escala contou com 30 itens, cujo CVC geral mostrou-se em 0,92, evidenciando sua validade de conteúdo e sinalizando para a clareza de linguagem, pertinência prática e relevância teórica do instrumento. Espera-

¹ Instituto Federal de Minas Gerais (IFMG), Piumhi – MG – Brasil. Maestro. Estudiante de Doctorado en Educación (UFES). ORCID: <https://orcid.org/0000-0003-3075-3532>. E-mail: denilson.marques@ifmg.edu.br

² Universidad Federal de Espírito Santo (UFES), Vitória - ES - Brasil. Estudiante de Doctorado en Educación. ORCID: <https://orcid.org/0000-0003-2692-4941>. E-mail: talitaandrade@gmail.com

³ Universidad Federal de Espírito Santo (UFES), Vitória – ES – Brasil. Profesor en el Programa de Posgrado en Educación. Líder del Instituto de Investigación en Educación y Educación Física (Proteoria/UFES). Doctorado en Educación (UFES). ORCID: <https://orcid.org/0000-0002-9216-7291>. E-mail: wagnercefd@gmail.com

se que este trabalho abra perspectivas para novas pesquisas, com vistas a dar continuidade ao processo de validação da EAAE e, conseqüentemente, possibilitar seu uso em estudos sobre avaliação educacional.

PALAVRAS-CHAVE: *Escala de atitudes. Avaliações externas aplicadas em larga escala. Validade de conteúdo. Avaliação educacional. Psicometria.*

ABSTRACT: *This study presents the process of development of the Scale of Attitudes towards External Assessments applied on a large scale (EAAE) and evidence of its content validity. This instrument was developed in order to capture what basic education teachers think, feel and how they behave through this type of assessment. This is a study with a quantitative approach, structured in two stages: evaluation by expert, using the Content Validity Index (CVI); and pilot application was conducted with a sample of the target population. The final version of the EAAE was reduced to 30 items. The overall CVI of the scale was 0.92, evidencing its content validity and signaling the clarity of language, practice relevance and theoretical relevance of the instrument. It is hoped that this study will open perspectives for further research, with a view to continuing the EAAE validation process and, consequently, enabling its use in studies on educational assessment.*

KEYWORDS: *Attitude scales. External assessments applied on a large scale. Content validity. Educational assessment. Psychometry.*

Introducción

En el ámbito educativo, la evaluación se configura en varias modalidades: evaluación del aprendizaje; planes de estudio y programas; sistemas educativos; de profesionales; y políticas públicas (AFONSO, 2000). Según Stadler (2017), en general, la evaluación tiene el papel crucial de analizar la calidad de la educación mediante el seguimiento del proceso educativo o la lectura de sus resultados.

Cuando este proceso es realizado por agentes ajenos a la escuela, la evaluación se dice externa. La mayoría de las veces, su aplicación ocurre a gran escala, es decir, para un gran número de personas. En resumen, las evaluaciones externas aplicadas a gran escala han sido utilizadas por organismos gubernamentales como herramientas que guían políticas, proyectos, programas y acciones educativas entre escuela y comunidad, y pueden clasificarse según: su naturaleza, en referencia a su alcance; propósitos/ propósitos; y sus efectos (BAUER; ALAVARSE; OLIVEIRA, 2015; ORFIELD; WALD, 2000).

La clasificación de los efectos considera los impactos que la evaluación trae a su población objetivo. A este respecto, las evaluaciones de alto impacto (*high stakes tests*) pueden ofrecer subsidios para los procesos de toma de decisiones y, por lo tanto, traer consecuencias

notorias para los actores del proceso escolar. Por otro lado, las evaluaciones de bajo impacto (*low stakes tests*) no ofrecen efectos directos o indirectos a estos sujetos (BAUER; ALAVARSE; OLIVEIRA, 2015; ORFIELD; WALD, 2000).

En Brasil, la Política Nacional de Evaluación de la Educación Básica, instituida a través de la publicación de la Ordenanza N° 458 del 5 de mayo de 2020 (BRASIL, 2020), está compuesta por tres evaluaciones externas aplicadas a gran escala: el Sistema de Evaluación de la Educación Básica (Saeb), creado en 1990; el Examen Nacional de Bachillerato (Enem), creado en 1998; y el Examen Nacional de Certificación de Habilidades de Jóvenes y Adultos (Enceja), creado en 2002. El artículo 8 y el artículo 18 de dicha ordenanza confieren a Saeb y Enem, respectivamente, el mismo objetivo general: "[...] evaluar el dominio de las competencias y habilidades esperadas a lo largo de la educación básica, de acuerdo con la Base Nacional de Currículo Común - BNCC y las correspondientes directrices curriculares nacionales" (BRASIL, 2020, p. 2). Enceja, de acuerdo con el artículo 12, procede medir "[...] las competencias y habilidades requeridas para la finalización de la escuela primaria y secundaria, según sea el caso" (BRASIL, 2020, p. 2).

En común, todos ellos tienen al Instituto Nacional de Estudios e Investigaciones Educativas Anísio Teixeira (Inep), una autoridad federal vinculada al Ministerio de Educación de Brasil (MEC), como agente externo responsable de su organización y gestión. Además, estas evaluaciones, de carácter nacional, pueden considerarse de alto impacto para los estudiantes a los que están destinadas y, en consecuencia, para las demás asignaturas que constituyen la vida cotidiana de las escuelas. Enceja, por ejemplo, se puede utilizar con el propósito de certificar diferentes niveles de educación; y Saeb y Enem, como mecanismo único, alternativo o complementario para el acceso de estos estudiantes a la educación superior.

También es necesario destacar que, a partir de los resultados obtenidos en estas evaluaciones, es posible realizar un diagnóstico de la educación básica brasileña y de los factores que pueden interferir en el desempeño del estudiante, porque, junto con ellos, se aplican cuestionarios que permiten explicar los niveles de aprendizaje demostrados por los estudiantes evaluados a partir de la diversa información contextual recopilada. Con esto, el uso político de estos instrumentos se ha producido durante décadas y es frecuente la apropiación de sus resultados para delinear las políticas educativas contemporáneas (MILITÃO; SANTANA, SANTANA, PERBONI, 2019).

Sin embargo, todavía existe una brecha con respecto a los estudios dirigidos a analizar las actitudes de los docentes hacia estas evaluaciones. En una encuesta realizada en enero de

2022, en el Portal de revistas de la Coordinación para el Perfeccionamiento del Personal de Educación Superior (Capes), a través de consulta ampliada con bases de datos electrónicas *Scientific Electronic Library Online (SciELO)*, *Scopus* e *Web of Science (WoS)*, utilizando los descriptores ["Evaluación externa" OR "Evaluaciones externas"] AND ["actitud de los maestros" OR "actitud de los maestros"], no se mapeó ninguna investigación producida para este propósito.

Así, este artículo pretende presentar la Escala de Actitudes hacia las Evaluaciones Externas aplicadas a gran escala (EAAE) y los análisis realizados con el fin de encontrar evidencia de validez de su contenido. Esta es una de las primeras etapas de su proceso de construcción.

Las actitudes de constructo, utilizadas en este proceso, tienen su origen a principios del siglo XX, a partir del estudio de los sociólogos Thomas y Znaniecki (1918), quienes lo conceptualizan como un estado individual y determinante para la acción de los individuos. Según Sarti, Vendramini y Camilo (2021), para explicarlo, existen diferentes modelos en la literatura. El tripartito es el más relevante, sobre todo porque incluye múltiples factores psicológicos. Este modelo considera las siguientes dimensiones interrelacionadas: cognitiva, afectiva y conductual.

Según Camilo (2021), la dimensión cognitiva abarca percepciones, pensamientos y conceptos con relación al objeto investigado, buscando captar lo que los sujetos saben (o creen saber) sobre él; la dimensión afectiva, a su vez, se refiere a las emociones y sentimientos de estos sujetos; y conductual abarca sus acciones y/o intenciones de actuar ante el objeto. En este aspecto, la EAAE fue desarrollada con el fin de captar lo que los maestros de educación básica, que se constituyen como la población objetivo de la escala, piensan (dimensión cognitiva), sienten (dimensión afectiva) y cómo se comportan (dimensión conductual) a través de evaluaciones externas aplicadas a gran escala en Brasil.

Así, este artículo se estructura en tres apartados, además de esta introducción. En el primero, se presenta el método utilizado para llevar a cabo la investigación. Luego, se evidencian los resultados obtenidos. Posteriormente, cerrando el artículo, se lleva a cabo una discusión sobre los resultados obtenidos, acompañada de las consideraciones finales, seguidas de las referencias.

Método

Participantes

Los ítems de la EAAE fueron sometidos inicialmente a la evaluación de diez jueces expertos: ocho profesores de doctorado y dos estudiantes de doctorado, todos vinculados a Programas de Posgrado en Educación y/o Psicología de las cinco regiones brasileñas. La selección de estos profesionales fue intencional, considerando la competencia reconocida en el área de evaluación educativa y/o psicológica.

A partir de los resultados obtenidos en el análisis de los jueces expertos, los ítems preseleccionados para conformar la escala fueron sometidos a una evaluación semántica, mediante la aplicación del instrumento a una muestra de 30 docentes de educación básica (aplicación piloto). El propósito de esta etapa fue verificar si los ítems eran comprensibles en cuanto al lenguaje, considerando los sujetos pertenecientes a la población objetivo de la escala, como lo indican Glaser y Strauss (2017). El tamaño de la muestra se basó en el criterio de saturación (GLASER; STRAUSS, 2017).

Instrumento

La EAAE (en proceso de elaboración) contó inicialmente con 42 ítems elaborados por investigadores del Instituto de Investigaciones en Educación y Educación Física de la Universidad Federal de Espírito Santo, a partir de estudios que consideran la definición constitutiva y operativa del fenómeno investigado (PASQUALI, 2010), desarrollados desde 2017 dentro del instituto.

Para este propósito, se consideraron las actitudes de constructo, compuestas por las dimensiones cognitiva (15 ítems), afectiva (12 ítems) y conductual (15 ítems), con el fin de captar lo que los docentes de educación básica (público objetivo) piensan, sienten y cómo se comportan, respectivamente, en relación con las evaluaciones externas que se aplican a gran escala en Brasil.

Cada dimensión va acompañada de una frase guía: para la dimensión cognitiva se solicita una respuesta respecto a lo que se cree (creencias, conocimientos, información y/u opiniones) en relación con las evaluaciones externas aplicadas a gran escala; para la dimensión afectiva, se solicita una respuesta basada en los sentimientos; y, para la dimensión conductual, la respuesta debe basarse en acciones cotidianas en relación con el objeto.

Los ítems que componen el instrumento fueron elaborados en forma de sueros y, como opción de respuesta, se estructuran en una escala *Likert* de cinco puntos de acuerdo, que van desde 1 (muy en desacuerdo) hasta 5 (Totalmente de acuerdo). En resumen, se espera que las puntuaciones más bajas (más altas) indiquen menos (más) actitudes positivas/favorables hacia el objeto investigado.

Procedimientos

Los análisis fueron posibles mediante el envío, por correo electrónico, de una carta de invitación, acompañada de la solicitud de evaluación de los ítems previamente preparados. Así, tras la aceptación y consentimiento del participante al Formulario de Consentimiento Libre e Informado (TCLE), se puso a disposición el acceso al formulario de análisis. La confidencialidad de los participantes se mantuvo durante todo el proceso.

En la evaluación, cada juez experto debe asignar un valor ordinal, entre 1 y 5, de acuerdo con la calidad del ítem. Se consideraron tres aspectos: la claridad del lenguaje (1. nada claro - 5. totalmente claro); relevancia práctica (1. nada relevante - 5. totalmente pertinente); y la relevancia teórica (1. nada relevante - 5. totalmente relevante) de los ítems. Este procedimiento tomó como referencia los estudios de Hernández-Nieto (2002), Cassepp-Borges, Balbinotti y Teodoro (2010) y Pacico (2015).

Como indican Cassepp-Borges, Balbinotti y Teodoro (2010), la claridad lingüística consiste en el análisis del lenguaje utilizado en los ítems, considerando las características del público objetivo. La relevancia práctica, a su vez, tiene como objetivo analizar si el elemento es, de hecho, importante en la constitución del instrumento. La relevancia teórica busca evaluar el grado de asociación entre el ítem y la teoría que lo sustenta.

Para el análisis semántico, se elaboró un cuestionario en el que los participantes de la aplicación piloto pudieron clasificar el lenguaje adoptado como claro, difícil o incomprensible. En esta etapa, el estudio de Pasquali (2010) fue asumido como soporte teórico.

Análisis estadístico

El análisis cuantitativo se realizó calculando la media aritmética y la desviación estándar de las respuestas atribuidas por los jueces expertos para cada ítem, en cada aspecto analizado. En resumen, cuanto mayor (cuanto menor) es el valor para el primero, más (menos) bien evaluado es, en promedio, el elemento. En relación con la desviación estándar, los valores altos

(bajos) indican una menor (mayor) regularidad de las respuestas, lo que indica una posible divergencia de opiniones entre los evaluadores.

Para determinar el nivel de acuerdo entre los jueces expertos, se utilizó el Coeficiente de Validez de Contenido (CVC), propuesto por Hernández-Nieto (2002). Es un indicador utilizado para calcular el nivel de consenso, calculando el promedio del valor asignado por los jueces en cada ítem del instrumento, expresado en términos porcentuales, restando el sesgo de respuesta. En términos matemáticos:

$$CVC = \frac{\sum_{i=1}^K \frac{x_i}{K}}{x_M} - \left(\frac{1}{K}\right)^K,$$

en que x_i representa el valor x asignado al tema por el juez $i; K$ representa el número de magistrados; y x_M , la puntuación máxima de la escala utilizada.

Además del CVC para la claridad del lenguaje, la relevancia práctica y la relevancia teórica, también se calculó la media para todos los ítems y el CVC general de la escala. Como indican Polit y Beck (2006), los ítems con CVC igual o superior al 80% se consideraron válidos. Todos los análisis se realizaron con la ayuda del *software* Microsoft Excel®.

Resultados

Con base en las evaluaciones de los jueces expertos, de los 42 ítems previamente elaborados, 12 fueron excluidos por presentar información repetida o CVC por debajo del 80%. De los otros, cinco fueron reformulados considerando las sugerencias de los evaluadores con el fin de dar objetividad a la escritura o facilitar su comprensión.

Así, se seleccionaron previamente 30 ítems para comparar con la EAAE: 12 del componente cognitivo, ocho del componente afectivo y diez del componente conductual óptimo c. El análisis semántico, realizado a partir de la aplicación piloto del instrumento, no indicó problemas en cuanto a su comprensión para la muestra del público objetivo investigado (N=30).

La puntuación total obtenida por los encuestados puede oscilar entre 30 y 150 puntos, con un punto medio igual a 90 puntos. Así, las puntuaciones por encima (por debajo) de este valor indican actitudes más (menos) positivas que negativas de los docentes en relación con las evaluaciones externas aplicadas a gran escala.

Las tablas 1, 2 y 3 traen estos ítems, organizados por dimensión y acompañados de la pregunta guía, media y desviación estándar de los valores de la escala *Likert asignados* por los jueces, así como el CVC para cada criterio considerado (claridad lingüística, relevancia práctica y relevancia teórica del ítem).

Tabla 1 - Media, desviación estándar y CVC de los ítems del componente cognitivo de la EAAE

	Componente cognitivo Artículos	C.L.		P.P.		R.T.	
		M (DP)	CVC	M (DP)	CVC	M (DP)	CVC
Responda en función de lo que cree (sus conocimientos, opiniones y / o creencias) en relación con las evaluaciones externas aplicadas a gran escala.	Evaluar adecuadamente la calidad de la enseñanza y el aprendizaje	4,50 (0,85)	0,90	4,60 (0,52)	0,92	4,90 (0,32)	0,98
	Son el principal instrumento de seguimiento de la eficacia de las políticas educativas	4,70 (0,67)	0,94	4,80 (0,63)	0,96	4,90 (0,32)	0,98
	Cumplir satisfactoriamente con el propósito de medir los niveles de aprendizaje de los estudiantes	4,20 (1,03)	0,84	4,40 (1,35)	0,88	4,40 (1,35)	0,88
	Son útiles para seleccionar adecuadamente a los estudiantes para otras etapas / niveles de educación	4,70 (0,67)	0,94	4,50 (1,27)	0,90	4,60 (1,26)	0,92
	Impacto en la selección de contenidos impartidos en la escuela	4,40 (0,84)	0,88	4,30 (1,34)	0,86	4,10 (1,37)	0,82
	Medir adecuadamente el nivel de conocimiento de los estudiantes en relación con los contenidos del programa previstos en la Base Curricular Común Nacional (BNCC)	4,40 (0,84)	0,88	4,90 (0,32)	0,98	5,00 (0,00)	1,00
	Sus matrices curriculares dan igual valor a todos los conocimientos previstos en la Base Curricular Común Nacional (BNCC)	4,60 (1,26)	0,92	4,20 (1,69)	0,84	4,20 (1,69)	0,84
	Permitir el análisis de los aspectos sociales y culturales de los estudiantes	4,20 (1,03)	0,84	4,80 (0,63)	0,96	4,80 (0,63)	0,96
	Estas son herramientas necesarias para reducir las desigualdades educativas	4,80 (0,63)	0,96	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Estos son importantes instrumentos de rendición de cuentas a la empresa	4,80 (0,63)	0,96	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Estos son instrumentos importantes para el seguimiento de la práctica docente	4,90 (0,32)	0,98	4,40 (1,35)	0,88	4,40 (1,35)	0,88
	Son herramientas eficientes para promover bonos financieros a los profesionales de la educación	4,50 (0,71)	0,90	4,40 (0,84)	0,88	4,50 (0,85)	0,90
	Total:	4,56 (0,79)	0,91	4,61 (0,86)	0,92	4,65 (0,76)	0,93

Nota: M = media aritmética; D. P. = desviación estándar; C. L. = claridad del lenguaje; P. P. = pertinencia práctica;

R. T. = relevancia teórica.

Fuente: Elaboración propia (2021)

Tabla 2 - Media, desviación estándar y CVC de los ítems del componente afectivo de la EAEE

	Componente afectivo	C.L.		P.P.		R.T.	
		M (DP)	CVC	M (DP)	CVC	M (DP)	CVC
Responder a en función de sus sentimientos con respecto a las evaluaciones externas a gran escala	Tengo una apreciación para este tipo de evaluación	5,00 (0,00)	1,00	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Espero conocer sus resultados	4,80 (0,42)	0,96	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Confío en tus resultados	5,00 (0,00)	1,00	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Confío en las diferencias entre escuelas / estudiantes que se hacen a partir de ellos	4,60 (0,84)	0,92	4,70 (0,67)	0,94	4,80 (0,63)	0,96
	Confío en las certificaciones de finalización de un nivel/etapa de enseñanza que se llevan a cabo desde ellas	4,70 (0,67)	0,94	4,80 (0,63)	0,96	4,80 (0,63)	0,96
	Siento que mi conocimiento es valorado por ellos	5,00 (0,00)	1,00	4,90 (0,32)	0,98	5,00 (0,00)	1,00
	Me doy cuenta de que la disciplina que enseño es valorada por ellos	4,40 (0,84)	0,88	4,30 (0,82)	0,86	4,50 (0,53)	0,90
	Me gusta guiar mi trabajo a partir de tus resultados	4,90 (0,32)	0,98	4,40 (1,35)	0,88	4,40 (1,35)	0,88
	Total:	4,77 (0,44)	0,95	4,73 (0,54)	0,95	4,81 (0,39)	0,96

Nota: M = media aritmética; D. P. = desviación estándar; C. L. = claridad del lenguaje; P. P. = pertinencia práctica;

R. T. = relevancia teórica.

Fuente: Elaboración propia (2021)

Tabla 3 - Media, desviación estándar y CVC de los ítems del componente conductual de la EAEE

	Componente conductual	C.L.		P.P.		R.T.	
		M (DP)	CVC	M (DP)	CVC	M (DP)	CVC
Responda en función de sus acciones diarias en relación con las evaluaciones externas aplicadas a gran escala	Planifico mis clases a partir de tus resultados	5,00 (0,00)	1,00	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Incorporo tus resultados en mis acciones cotidianas en el aula	4,50 (0,85)	0,90	4,60 (1,26)	0,92	4,60 (1,26)	0,92
	Organizo el horario escolar para mejorar su rendimiento	5,00 (0,00)	1,00	4,60 (1,26)	0,92	4,60 (1,26)	0,92
	Priorizo la enseñanza de los contenidos que son más frecuentes en estas evaluaciones	4,70 (0,67)	0,94	4,90 (0,32)	0,98	4,90 (0,32)	0,98
	Hablo con los estudiantes sobre su importancia	5,00 (0,00)	1,00	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Recomiendo participar en estas evaluaciones	4,50 (1,08)	0,90	4,60 (0,84)	0,92	4,60 (0,84)	0,92
	Presento un feedback sobre los resultados obtenidos en el aula	4,90 (0,32)	0,98	5,00 (0,00)	1,00	5,00 (0,00)	1,00

Resuelvo preguntas de evaluaciones previas en el aula	4,80 (0,42)	0,96	4,90 (0,32)	0,98	4,80 (0,63)	0,96
Utilizo preguntas anteriores en mis evaluaciones escolares (internas)	4,70 (0,67)	0,94	5,00 (0,00)	1,00	5,00 (0,00)	1,00
Participo en formación específica para un mejor uso de tus resultados	4,00 (1,49)	0,80	4,10 (1,37)	0,82	4,20 (1,40)	0,84
Total:	4,68 (0,61)	0,94	4,74 (0,60)	0,95	4,77 (0,57)	0,950

Nota: M = media aritmética; D. P. = desviación estándar; C. L. = claridad del lenguaje; P. P. = pertinencia práctica; R. T. = relevancia teórica.

Fuente: Elaboración propia (2021)

Además del CVC para cada criterio, también se calculó el CVC de la escala como un conjunto uniforme. Así, para el aspecto de claridad del lenguaje, el CVC general fue de 0,92, para la relevancia práctica fue de 0,93 y para la relevancia teórica de 0,94. El CVC global de la escala fue de 0,93, confiriendo validez a su contenido según los métodos de análisis utilizados en esta investigación.

En cuanto a los análisis descriptivos, se percibe una mejor media, en la evaluación de los jueces, para el criterio de relevancia teórica (M=4,73), frente a los demás (M=4,70 en ambos). En cuanto a la desviación estándar, esta evaluación es más heterogénea para el criterio de pertinencia práctica (D. P. = 0,64) y más homogénea para la claridad lingüística (D. P. = 0,57). Para el criterio de relevancia teórica, la desviación estándar calculada fue de 0,60.

Discusión y consideraciones finales

Este artículo fue desarrollado con el propósito de presentar el proceso de construcción y evidencia de validez de contenido de la Escala de Actitudes hacia evaluaciones externas aplicadas a gran escala (EAAE). Este instrumento se construyó a partir de la adopción del modelo tripartito del constructo de actitudes, con el fin de captar lo que los docentes de educación básica piensan, sienten y cómo se comportan a través de evaluaciones externas aplicadas a gran escala.

El proceso de construcción y la búsqueda de evidencias de validación de contenido de las EAAE se basaron estrictamente en procedimientos reconocidos por la literatura especializada, consultada para el desarrollo del instrumento. El uso de CVC, en el análisis de los jueces expertos, y el análisis semántico demostraron buenos resultados, apoyando la conclusión de que el instrumento presenta evidencia de validez de contenido, señalando la claridad del lenguaje, la pertinencia teórica y la relevancia práctica de los ítems que lo componen.

En este sentido, es necesario destacar que cuatro ítems presentaron 100% de acuerdo en todos los aspectos evaluados: dos del componente afectivo ("Tengo aprecio por este tipo de evaluación" y "Confío en sus resultados") y dos de los conductuales ("Planifico mis clases a partir de sus resultados" y "Hablo con los alumnos sobre su importancia"), indicando una evaluación perfecta unánime entre los evaluadores. Se espera, por tanto, teóricamente, que estos ítems aporten más información sobre las actitudes de los docentes de educación básica en relación con las evaluaciones externas aplicadas a gran escala.

Sin embargo, para confirmar esta hipótesis, así como para evaluar otros aspectos psicométricos de la escala, es necesario que se realicen nuevos análisis de validez, como el basado en la estructura interna del instrumento, en las relaciones con medidas externas o en el patrón de respuesta a los ítems. Esto se hará en estudios posteriores, teniendo en cuenta que la EAAE todavía está en construcción.

Como principal limitación de este estudio, destacamos el hecho de que los análisis se realizaron en formato online, lo que dificulta el control de variables, como el cansancio y la distracción de los evaluadores, que pueden comprometer los resultados encontrados. Por otro lado, se pudo contar con la colaboración de profesionales que trabajan en todas las regiones del país, proporcionando una evaluación considerablemente integral de los ítems de la escala.

Por último, se destaca que, una vez finalizado, las EAAE podrán subvencionar una nueva agenda de investigación en el ámbito educativo, contribuyendo al crecimiento y mejora de la producción científica. Desde su aplicación al público objetivo, se pueden responder algunas preguntas, por ejemplo: a) ¿Cómo se posicionan los docentes de educación básica en relación con las evaluaciones externas aplicadas a gran escala? b) ¿Existe una relación significativa entre las variables sociales/demográficas/económicas y las actitudes más positivas hacia ellas? c) ¿Influyen actitudes más positivas en los resultados educativos alcanzados? Entre otros.

Con ello, EAAE podrá ofrecer subvenciones para el establecimiento de soluciones eficientes, estratégicas e innovadoras orientadas a mejorar la calidad de la educación, evidenciadas a través de estas evaluaciones. Así, se enfatiza que la construcción de esta escala es sumamente relevante, sobre todo en vista de la escasez de estudios que aborden el tema investigado.

GRACIAS: A la Fundación de Apoyo a la Investigación y la Innovación de Espírito Santo (FAPES) por apoyar el desarrollo de esta investigación.

REFERENCIAS

AFONSO, A. J. **Avaliação educacional:** Regulação e emancipação para uma sociologia das políticas avaliativas contemporâneas. 2. ed. São Paulo: Cortez, 2000.

BAUER, A.; ALAVARSE, O. M.; OLIVEIRA, R. P. Avaliação em larga escala: Uma sistematização do debate. **Educação e Pesquisa**, São Paulo, v. 41, n. esp., p. 1367-1382, dez. 2015. Disponible en: <http://www.scielo.br/pdf/ep/v41nspe/1517-9702-ep-41-spe-1367.pdf>. Acceso: 03 enero 2022.

BRASIL. **Portaria n. 458, de 5 de maio de 2020.** Institui normas complementares necessárias ao cumprimento da Política Nacional de Avaliação da Educação Básica. Brasília, DF: Ministro de Estado da Educação, 2020. Disponible en: <https://www.in.gov.br/web/dou/-/portaria-n-458-de-5-de-maio-de-2020-255378342>. Acceso: 05 enero 2022.

CAMILO, C. C. **Construção e estudos psicométricos da Escala de Atitude frente à atuação do psicólogo.** 2021. Tese (Doutorado em Psicologia) – Programa de Pós-Graduação Stricto Sensu em Psicologia, Universidade São Francisco, Campinas, 2021. Disponible en: <https://www.usf.edu.br/galeria/getImage/427/4000894494237433.pdf>. Acceso: 5 enero 2022.

CASSEPP-BORGES, V.; BALBINOTTI, M. A. A.; TEODORO, M. L. M. Tradução e validação de conteúdo: Uma proposta para a adaptação de instrumentos. *In:* PASQUALI, L. **Instrumentação psicológica:** Fundamentos e práticas. Porto Alegre: Artmed, 2010.

GLASER, B. G.; STRAUSS, A. L. **Discovery of grounded theory:** Strategies for qualitative research. New York: Aldine de Gruyter, 2017.

HERNÁNDEZ-NIETO, R. A. **Contribuciones al análisis estatístico.** Mérida, Venezuela: Universidad de Los Andes/Iesinfo, 2002.

MILITÃO, A. N.; SANTANA, A. C. M.; PERBONI, F. A expansão dos sistemas avaliativos e a teoria crítica: Um diálogo (im)pertinente. **Revista Ibero-Americana de Estudos em Educação**, Araraquara, v. 14, n. esp. 4, p. 2046-2056, 2019. Disponible en: <https://periodicos.fclar.unesp.br/iberoamericana/article/view/12927>. Acceso: 22 dic. 2021.

ORFIELD, G.; WALD, J. Testing, testing: The high-stakes testing mania hurts poor and minority students the most. **The Nation**, New York, v. 270, n. 22, p. 38-40, jun. 2000. Disponible en: <https://thenation.s3.amazonaws.com/pdf/testingtesting2000.pdf>. Acceso: 05 enero 2022.

PACICO, J. C. Como é feito um teste? Produção de itens. In: HUTZ, C. S.; BANDEIRA, D. R.; TRENTINI, C. M. **Psicometria**. Porto Alegre: Artmed, 2015.

PASQUALI, L. **Instrumentação psicológica**. Brasília, DF: Editora Vetor, 2010.

POLIT, D. F.; BECK, C. T. The content validity index: Are you sure you know what's being reported? Critique and recommendations. **Research in Nursing & Health**, v. 29, n. 5, p. 489-497, 2006. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/nur.20147>. Acceso: 15 enero 2022.

SARTI, A. G.; VENDRAMINI, C. M. M.; CAMILO, C. C. Evidências de validade da escala de atitudes de estudantes do ensino superior frente à estatística-EAEst. **Ciência & Educação**, Bauru, v. 27, e21042, 2021. Disponible en: <https://www.scielo.br/j/ciedu/a/zJ9mkwzpzpyP5zRZjhsYsfh/abstract/?lang=pt>. Acceso: 10 dic. 2021.

STADLER, J. C. **Prova Brasil de Matemática do 5º ano do ensino fundamental: Resultados nas plataformas Devolutivas Pedagógicas e QEdU**. Ponta Grossa, 2017. 165 f. Dissertação (Mestrado em Educação) – Universidade Estadual de Ponta Grossa, Ponta Grossa, 2017. Disponible en: <http://tede2.uepg.br/jspui/handle/prefix/2394>. Acceso: 15 enero 2022.

THOMAS, W. I.; ZNANIECKI, F. **The Polish peasant in Europe and America: Monograph of an immigrant group**. Chicago: University of Chicago Press, 1918.

Cómo hacer referencia a este artículo

SOARES, D. J. M.; SOARES, T. E. A.; SANTOS, W. Evidencia de la validez del contenido de la Escala de Actitud hacia las Evaluaciones Externas a gran escala (EAEE). **Revista Ibero-Americana de Estudos em Educação**, Araraquara, v. 17, n. 3, p. 1814-1826, jul./sept. 2022. e-ISSN: 1982-5587. DOI: <https://doi.org/10.21723/riaee.v17i3.16191>

Enviado en: 26/01/2022

Revisiones requeridas en: 15/04/2022

Aprobado en: 06/05/2022

Publicado en: 01/07/2022

Procesamiento y edición: Editora Ibero-Americana de Educação.
Corrección, formateo, normalización y traducción.