

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

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

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

Denilson Junio Marques SOARES¹
Talita Emidio Andrade SOARES²
Wagner dos SANTOS³

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 EAEA 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.

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-se que este trabalho abra

¹ Federal Institute of Minas Gerais (IFMG), Piumhi – MG – Brazil. Professor. PhD student in Education (Ufes). ORCID: <https://orcid.org/0000-0003-3075-3532>. E-mail: denilson.marques@ifmg.edu.br

² Federal University of Espírito Santo (Ufes), Vitória – ES – Brazil. PhD student in Education. ORCID: <https://orcid.org/0000-0003-2692-4941>. E-mail: talitaeandrade@gmail.com

³ Federal University of Espírito Santo (Ufes), Vitória – ES – Brazil. Professor in the Graduate Program in Education. Leader of the Institute for Research in Education and Physical Education (Proteoria/Ufes). PhD in Education (Ufes). ORCID: <https://orcid.org/0000-0002-9216-7291>. E-mail: wagnercefd@gmail.com

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.*

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.*

Introduction

In the educational field, the evaluation is configured in various ways: evaluation of learning; of curricula and programs; of educational systems; of professionals; and of public policies (AFONSO, 2000). According to Stadler (2017), in all of them, evaluation has the crucial role of analyzing the quality of education by monitoring the educational process or reading its results.

When this process is carried out by agents external to the school, the evaluation is said to be external. Most of the time, it is applied on a large scale, that is, to a large number of people. In summary, external evaluations applied on a large scale have been used by government agencies as tools to guide policies, projects, programs and educational actions in schools and communities, and can be classified according to: its nature, referring to its scope; its purposes / objectives; and its effects (BAUER; ALAVARSE; OLIVEIRA, 2015; ORFIELD; WALD, 2000).

The classification regarding effects considers the impacts that the evaluation brings to its target population. In this aspect, high stakes tests can provide inputs for decision-making processes and thus bring out conspicuous consequences for the actors of the school process.

In contrast, low stakes tests do not offer direct or indirect effects for these subjects (BAUER; ALAVARSE; OLIVEIRA, 2015; ORFIELD; WALD, 2000).

In Brazil, the National Policy for Evaluation of Basic Education, established through the publication of Ordinance No. 458 of May 5, 2020 (BRAZIL, 2020), is composed of three external assessments applied on a large scale: the System for Evaluation of Basic Education (Saeb), created in 1990; the National High School Exam (Enem), created in 1998; and the National Exam for Certification of Skills of Youth and Adults (Encceja), created in 2002. The art. 8 and art. 18 of the mentioned ordinance give to Saeb and Enem, respectively, the same general objective: "[...] to assess the mastery of skills and abilities expected throughout basic education, according to the Common National Curricular Base - BNCC and the corresponding national curriculum guidelines" (BRAZIL, 2020, p. 2). Encceja, according to art. 12, is responsible for assessing "[...] the competencies and skills required for the conclusion of primary and secondary education, as the case may be. (BRAZIL, 2020, p. 2).

In common, all of them have the National Institute of Educational Studies and Research Anísio Teixeira (Inep), a federal agency linked to the Brazilian Ministry of Education (MEC), as the external agent responsible for their organization and management. Furthermore, these national assessments can be considered to have a high impact on the students to whom they are intended and, consequently, on the other subjects that make up the daily life of schools. Encceja, for example, can be used with the purpose of certifying different levels of education; and Saeb and Enem, as single, alternative or complementary mechanisms for the access of these students to higher education.

It should also be noted that, based on the results obtained in these assessments, it is possible to make a diagnosis of Brazilian basic education and of factors that may interfere with student performance, because, along with them, questionnaires are applied to explain the levels of learning demonstrated by the students assessed from the various contextual information collected. With that, the political use of these instruments has been occurring for decades and the appropriation of their results to outline contemporary educational policies is frequent (MILITÃO; SANTANA; PERBONI, 2019).

However, there is still a gap regarding the studies aimed at analyzing the attitudes of teachers towards these assessments. In a survey conducted in January 2022, in the Periodical Portal of the Coordination for the Improvement of Higher Level Personnel (Capes), through expanded consultation to the electronic databases Scientific Electronic Library Online (SciELO), Scopus and Web of Science (WoS), using the descriptors ["External evaluation"

OR "External evaluations"] AND ["teachers' attitude" OR "teachers' attitude"], no research produced with this purpose was mapped.

Thus, this article aims to present the Scale of Attitudes towards External Evaluations applied on a large scale (EAAE in the Portuguese acronym) and the analyses performed in order to find evidence of its content validity. This is one of the initial stages of its construction process.

The attitudes construct, used in this process, has its origins in the early twentieth century, from the study of sociologists Thomas and Znaniecki (1918), who conceptualize it as an individual state and determinant for the action of individuals. According to Sarti, Vendramini, and Camilo (2021), to explain it, there are different models in the literature. The tripartite one is the most relevant, mainly because it encompasses multiple psychological factors. This model considers the following interrelated dimensions: cognitive, affective, and behavioral.

According to Camilo (2021), the cognitive dimension covers perceptions, thoughts, and concepts in relation to the object under investigation, seeking to capture what subjects know (or believe they know) about it; the affective dimension, in turn, refers to the emotions and feelings of these subjects; and the behavioral dimension encompasses their actions and/or intentions to act in the face of the object. In this perspective, the EAAE was developed in order to capture what the basic education teachers, who constitute the target population of the scale, think (cognitive dimension), feel (affective dimension), and how they behave (behavioral dimension) when facing the external evaluations applied in large scale in Brazil.

Thus, this article is structured in three sections, in addition to this introduction. The first section presents the method used to conduct the research. Next, the results achieved are highlighted. Subsequently, the article ends with a discussion of the results obtained, accompanied by the final considerations, followed by the references.

Method

Participants

The items of the EAAE were initially submitted to the evaluation of ten expert judges: eight doctoral professors and two doctoral students, all linked to graduate programs in education and/or psychology in the five Brazilian regions. The selection of these professionals

was intentional, considering their recognized competence in the area of educational and/or psychological evaluation.

Based on the results obtained in the expert judges' analysis, the pre-selected items to compose the scale underwent a semantic evaluation, by applying the instrument to a sample of 30 basic education teachers (pilot application). The purpose of this step was to verify whether the items were understandable in terms of language, considering the subjects belonging to the target population of the scale, as indicated by Glaser and Strauss (2017). The sample size was based on the saturation criterion (GLASER; STRAUSS, 2017).

Instrument

The SAEA (under development) initially had 42 items developed by researchers at the Institute for Research in Education and Physical Education at the Federal University of Espírito Santo, from studies that consider the constitutive and operational definition of the investigated phenomenon (PASQUALI, 2010), developed since 2017 within the institute.

To this end, the attitudes construct was considered, composed of the cognitive (15 items), affective (12 items), and behavioral (15 items) dimensions, in order to capture what teachers of basic education (target audience) think, feel, and how they behave, respectively, in relation to external assessments that are applied on a large scale in Brazil.

Each dimension is accompanied by a guiding phrase: for the cognitive dimension, a response is requested referring to what is believed (beliefs, knowledge, information and/or opinions) in relation to the external evaluations applied in large scale; for the affective dimension, a response based on feelings is requested; and for the behavioral dimension, the answer must be grounded on the daily actions in relation to the object.

The items that make up the instrument were elaborated in the form of assertions and, as response options, they are structured on a Likert-type scale of five points of agreement, ranging from 1 (strongly disagree) to 5 (strongly agree). In summary, it is expected that lower (higher) scores indicate less (more) positive/favorable attitudes towards the object investigated.

Procedures

The analyses were made possible by sending an invitation letter by e-mail, accompanied by a request to evaluate the previously prepared items. Thus, after the acceptance and agreement of the participant to the Free and Informed Consent Form (FICF), access to the analysis form was made available. The confidentiality of the participants was maintained throughout the process.

In the evaluation, each expert judge should assign an ordinal value, between 1 and 5, according to the quality of the item. Three aspects were considered: language clarity (1. not at all clear - 5. totally clear); practical relevance (1. not at all relevant - 5. totally relevant); and theoretical relevance (1. not at all relevant - 5. totally relevant) of the items. This procedure took as reference the studies of Hernandez-Nieto (2002), Cassepp-Borges, Balbinotti and Teodoro (2010) and Pacico (2015).

As indicated by Cassepp-Borges, Balbinotti and Teodoro (2010), the clarity of language consists in the analysis of the language used in the items, considering the characteristics of the target audience. The practical relevance, in turn, aims to analyze whether the item is, in fact, important in the constitution of the instrument. Theoretical relevance, on its turn, seeks to assess the degree of association between the item and the theory that supports it.

For the semantic analysis, a questionnaire was prepared in which the participants of the pilot application could classify the language adopted as clear, difficult or incomprehensible. At this stage, we used Pasquali's study as theoretical support (2010).

Statistical Analysis

The quantitative analysis was performed by calculating the arithmetic mean and standard deviation of the responses assigned by the expert judges for each item, in each aspect analyzed. In summary, the higher (lower) the value for the former, the more (less) well evaluated the item is, on average. As for the standard deviation, high (low) values indicate a lower (higher) regularity of answers, signaling a possible divergence of opinions among raters.

To determine the level of agreement between the expert judges, we used the Content Validity Coefficient (CVC), proposed by Hernández-Nieto (2002). This is an indicator used to calculate the level of consensus, by calculating the average of the value attributed by the

judges in each item of the instrument, expressed in percentage terms, subtracting response bias. In mathematical terms:

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

in which x_i represents the value x assigned to the item by the judge $i; K$ represents the number of judges; and x_M , the maximum score on the scale used.

In addition to the CVC for clarity of language, practical relevance, and theoretical relevance, the mean for all items and the overall CVC of the scale were also calculated. As indicated by Polit and Beck (2006), items with CVC equal to or greater than 80% were considered valid. All analyses were conducted with the aid of Microsoft Excel® software.

Results

Based on the expert judges' evaluations, of the 42 items previously elaborated, 12 were excluded for presenting repeated information or CVC lower than 80%. Of the remaining items, five were reformulated considering the suggestions of the evaluators in order to give objectivity to the writing or facilitate their understanding.

Thus, 30 items were previously selected to compose the SAE: 12 from the cognitive component, eight from the affective component, and ten from the behavioral component. The semantic analysis, carried out after the pilot application of the instrument, did not indicate any problems regarding its understanding for the sample of the target audience investigated (N=30).

The total score obtained by the respondents can vary from 30 to 150 points, with the average point equal to 90 points. Thus, scores above (below) this value indicate more (less) positive than negative attitudes of teachers toward external evaluations applied on a large scale.

Tables 1, 2 and 3 show these items, organized by dimension and accompanied by the guiding question, mean and standard deviation of the Likert scale values assigned by the judges, as well as the CVC for each criterion considered (clarity of language, practical relevance and theoretical relevance of the item).

Chart 1 – Mean, standard deviation and CVC of the items of the cognitive component of the EAAE

	Cognitive Component Items	C. L.		P. P		R. T.	
		M (SD)	CVC	M (SD)	CVC	M (SD)	CVC
Answer based on what you believe (your knowledge, opinions and/or beliefs) about the external assessments applied on a large scale.	Adequately assess the quality of teaching and learning	4,50 (0,85)	0,90	4,60 (0,52)	0,92	4,90 (0,32)	0,98
	They are the main instrument for monitoring the effectiveness of educational policies	4,70 (0,67)	0,94	4,80 (0,63)	0,96	4,90 (0,32)	0,98
	They satisfactorily fulfill the purpose of measuring students' learning levels	4,20 (1,03)	0,84	4,40 (1,35)	0,88	4,40 (1,35)	0,88
	They are useful for adequately selecting students for other stages/levels of education	4,70 (0,67)	0,94	4,50 (1,27)	0,90	4,60 (1,26)	0,92
	Have an impact on the selection of contents taught in school	4,40 (0,84)	0,88	4,30 (1,34)	0,86	4,10 (1,37)	0,82
	Adequately measure the level of knowledge of the students in relation to the programmatic contents foreseen in the Common National Curricular Base (BNCC)	4,40 (0,84)	0,88	4,90 (0,32)	0,98	5,00 (0,00)	1,00
	Its curricular matrices give equal value to all the knowledge foreseen in the Common National Curricular Base (BNCC)	4,60 (1,26)	0,92	4,20 (1,69)	0,84	4,20 (1,69)	0,84
	They allow for the analysis of social and cultural aspects of the students	4,20 (1,03)	0,84	4,80 (0,63)	0,96	4,80 (0,63)	0,96
	The following are necessary instruments to reduce educational inequalities	4,80 (0,63)	0,96	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	They are important instruments of accountability to society	4,80 (0,63)	0,96	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Important instruments for monitoring teaching practice are	4,90 (0,32)	0,98	4,40 (1,35)	0,88	4,40 (1,35)	0,88
	They are efficient instruments to promote financial bonuses for education professionals	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

Note: M = arithmetic mean; S.D. = standard deviation; C. L. = clarity of language; P. P.= practical relevance; R. T. = theoretical relevance.

Source: Prepared by the authors (2021)

Chart 2 – Mean, standard deviation and CVC of the items of the affective component of the EAAE

	Affective Component Items	C. L.		P. P		R. T.	
		M (SD)	CVC	M (SD)	CVC	M (SD)	CVC
Answer based on your feelings about the external evaluations applied on a large scale	I appreciate this kind of evaluation	5,00 (0,00)	1,00	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	Looking forward to its results	4,80 (0,42)	0,96	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	I trust its results	5,00 (0,00)	1,00	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	I trust the school/student comparisons that are made from them	4,60 (0,84)	0,92	4,70 (0,67)	0,94	4,80 (0,63)	0,96
	I trust the certifications of completion of a level/stage of education that are made from them	4,70 (0,67)	0,94	4,80 (0,63)	0,96	4,80 (0,63)	0,96
	I feel that my knowledge is valued by them	5,00 (0,00)	1,00	4,90 (0,32)	0,98	5,00 (0,00)	1,00
	I notice that the subject I teach is valued by them	4,40 (0,84)	0,88	4,30 (0,82)	0,86	4,50 (0,53)	0,90
	I like to guide my work from its results	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

Note: M = arithmetic mean; S.D. = standard deviation; C. L. = clarity of language; P. P.= practical relevance;

R. T. = theoretical relevance.

Source: Prepared by the authors (2021)

Chart 3 – Mean, standard deviation and CVC of the items of the behavioral component of the EAAE

	Behavioral Component Items	C. L.		P. P		R. T.	
		M (SD)	CVC	M (SD)	CVC	M (SD)	CVC
Answer based on your daily actions in relation to the external evaluations applied on a large scale	I plan my lessons based on its results	5,00 (0,00)	1,00	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	I incorporate its results into my daily actions in the classroom	4,50 (0,85)	0,90	4,60 (1,26)	0,92	4,60 (1,26)	0,92
	I organize the school timetable to enhance its realization	5,00 (0,00)	1,00	4,60 (1,26)	0,92	4,60 (1,26)	0,92
	I prioritize teaching the content that is most frequent in these assessments	4,70 (0,67)	0,94	4,90 (0,32)	0,98	4,90 (0,32)	0,98
	I talk to students about its importance	5,00 (0,00)	1,00	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	I recommend participating in these evaluations	4,50 (1,08)	0,90	4,60 (0,84)	0,92	4,60 (0,84)	0,92
	I provide feedback on the results obtained in class	4,90 (0,32)	0,98	5,00 (0,00)	1,00	5,00 (0,00)	1,00
	I solve questions from previous assessments in class	4,80 (0,42)	0,96	4,90 (0,32)	0,98	4,80 (0,63)	0,96
	I use previous questions in my school assessments (internal)	4,70 (0,67)	0,94	5,00 (0,00)	1,00	5,00 (0,00)	1,00

I participate in specific training for a better use of its results	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

Note: M = arithmetic mean; S.D. = standard deviation; C. L. = clarity of language; P. P. = practical relevance; R. T. = theoretical relevance.

Source: Prepared by the authors (2021)

In addition to the CVC for each criterion, the CVC of the scale as a uniform whole was also calculated. Thus, for the clarity of language aspect, the overall CVC was 0.92, for practical relevance 0.93, and for theoretical relevance 0.94. The overall CVC of the scale was 0.93, confirming its content validity according to the analysis methods used in this research.

As for the descriptive analyses, we noticed a better average, in the judges' evaluation, for the criterion theoretical relevance (M=4.73), compared to the others (M=4.70 in both). As for the standard deviation, this evaluation is more heterogeneous for the criterion practical relevance (S.D. = 0.64) and more homogeneous for clarity of language (S.D. = 0.57). For the criterion theoretical relevance, the calculated standard deviation was 0.60.

Discussion and final remarks

This article was developed in order to present the construction process and evidence of content validity of the Scale of Attitudes towards External Assessments applied on a large scale (EAAE). This instrument was built from the adoption of the tripartite model of the attitudes construct, in order to capture what teachers of basic education think, feel and how they behave when facing external evaluations applied on a large scale.

The construction process and the search for evidence of content validation of the EAAE were strictly guided by procedures recognized by the specialized literature, consulted for the development of the instrument. The use of the CVC, in the expert judges' analysis, and the semantic analysis showed good results, providing subsidies for the conclusion that the instrument presents evidence of content validity, pointing to the clarity of language, theoretical pertinence, and practical relevance of the items that compose it.

In this regard, it should be noted that four items showed 100% agreement in all aspects assessed: two of the affective component ("I appreciate this type of assessment" and "I trust its results") and two of the behavioral ("I plan my classes based on its results" and "I talk to students about its importance"), indicating a perfect unanimous assessment among the raters. Thus, it is theoretically expected that these items provide more information about the attitudes of basic education teachers toward external evaluations applied on a large scale.

However, to confirm this hypothesis, as well as to assess other psychometric aspects of the scale, it is necessary to conduct further validity analyses, such as those based on the internal structure of the instrument, the relationships with external measures or the response pattern to items. This will be done in later studies, considering that the EAAE is still under construction.

As a main limitation of this study, we highlight the fact that the analyses were carried out in the online format, which makes it difficult to control variables such as assessors' tiredness and distraction, which can compromise the results found. On the other hand, it was possible to count on the collaboration of professionals who work in all regions of the country, providing a considerably comprehensive assessment of the scale items.

Finally, it should be noted that, when completed, the EAAE may subsidize a new research agenda in the educational field, contributing to the growth and improvement of scientific production. From its application to the target audience, some questions can be answered, for example: a) How do basic education teachers position themselves in relation to external evaluations applied on a large scale? b) Is there a significant relationship between social/demographic/economic variables and more positive attitudes towards them? c) Do more positive attitudes influence the educational results achieved? Among others.

With this, the EAAE can offer subsidies for the establishment of efficient, strategic, and innovative solutions aimed at improving the quality of education, as evidenced by these assessments. Thus, the construction of this scale is extremely relevant, especially given the scarcity of studies that address the subject investigated.

ACKNOWLEDGEMENTS: To the Espírito Santo State Foundation for the Support of Research and Innovation (FAPES) for supporting the development of this research.

REFERENCES

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. Available at: <http://www.scielo.br/pdf/ep/v41nspe/1517-9702-ep-41-spe-1367.pdf>. Access on: 03 Jan. 2022.

BRAZIL. **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. Available at: <https://www.in.gov.br/web/dou/-/portaria-n-458-de-5-de-maio-de-2020-255378342>. Access on: 05 Jan. 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. Available at: <https://www.usf.edu.br/galeria/getImage/427/4000894494237433.pdf>. Access on: 5 Jan. 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 estadístico**. 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. Available at: <https://periodicos.fclar.unesp.br/iberoamericana/article/view/12927>. Access on: 22 Dec. 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. Available at: <https://thenation.s3.amazonaws.com/pdf/testingtesting2000.pdf>. Access on: 05 Jan. 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. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/nur.20147>. Access on: 15 Jan. 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. Available at: <https://www.scielo.br/j/ciedu/a/zJ9mkwzgzpzyP5zRZjhsYsfh/abstract/?lang=pt>. Access on: 10 Dec. 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. Available at: <http://tede2.uepg.br/jspui/handle/prefix/2394>. Access on: 15 Jan. 2022.

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

How to reference this article

SOARES, D. J. M.; SOARES, T. E. A.; SANTOS, W. Evidence of content validity of the Scale of Attitudes towards External Assessments applied on a large scale (EAAE). **Revista Ibero-Americana de Estudos em Educação**, Araraquara, v. 17, n. 3, p. 1805-1817, July/Sept. 2022. e-ISSN: 1982-5587. DOI: <https://doi.org/10.21723/riaee.v17i3.16191>

Submitted: 26/01/2022

Revisions required: 15/04/2022

Approved: 06/05/2022

Published: 01/07/2022

Processing and publishing by the Editora Ibero-Americana de Educação.

Correction, formatting, standardization and translation.