



Construct Validity and Reliability of the Scale of Attitudes towards External Assessments Applied on a Large Scale

Denilson Junio Marques Soares¹ 

Wallacy Campos Prado² 

Talita Emidio Andrade Soares³ 

Wagner dos Santos⁴ 

Abstract: The Scale of Attitudes towards External Assessments applied on a Large-Scale was developed to capture what basic education teachers think, feel, and how they behave in response to this type of assessment. Considering the potential of the instrument to support management decisions in the field of educational assessment policies, this article aimed to evaluate construct validity, via factor analysis, as well as scale reliability, via composite reliability of the scale, based on its application to a sample of teachers affiliated with the Education Department of Espírito Santo/Brazil. This is a quantitative, nonexperimental, instrumental study that involved 405 teachers from the public school network, whose results indicate adequate psychometric indices and a satisfactory factor structure consistent with the proposed three-dimensional attitude construct. The statistical coefficients found at the level of the analyses performed legitimize its use for the development of public policies and effective practices in the educational field.

Keywords: psychometrics, attitudes scale, teachers, educational research, factor analysis

Validade de Construto e Confiabilidade da Escala de Atitudes perante as Avaliações Externas Aplicadas em Larga Escala

Resumo: A Escala de Atitudes perante as Avaliações Externas aplicadas em Larga Escala foi desenvolvida para capturar o que os professores da educação básica pensam, sentem e se comportam mediante esse tipo de avaliação. Considerando as potencialidades do instrumento para subsidiar decisões gerenciais no âmbito das políticas de avaliação educacional, esta pesquisa objetivou avaliar a validade do construto, mediante a Análise Fatorial, e a confiabilidade da escala, por meio da fidedignidade composta, a partir da aplicação para uma amostra de professores vinculados à Secretaria de Educação do Espírito Santo/Brasil. Trata-se de um estudo quantitativo, não experimental, de cunho instrumental, que envolveu 405 docentes da rede, cujos resultados indicam para adequados índices psicométricos e uma estrutura fatorial satisfatória e condizente com a proposta tridimensional do construto atitudes. Os coeficientes estatísticos encontrados ao nível das análises realizadas legitimam sua utilização para o desenvolvimento de políticas e práticas efetivas no âmbito educacional.

Palavras-chave: psicometria, escala de atitudes, professores, pesquisa educacional, análise fatorial

Validez de Constructo y Confiabilidad de la Escala de Actitudes ante las Evaluaciones Externas a Gran Escala

Resumen: La Escala de Actitudes ante las Evaluaciones Externas aplicadas a Gran Escala se desarrolló para identificar lo que el profesorado de educación básica piensa, siente y hace frente a este tipo de evaluación. Teniendo en cuenta el potencial del instrumento para respaldar las decisiones gerenciales en las políticas de evaluación educativa, este artículo tuvo como objetivo evaluar la validez del constructo, mediante el análisis factorial, y la confiabilidad, mediante la fiabilidad compuesta de la escala, basándose en su aplicación a una muestra de profesores de la Red de Educación de Espírito Santo/Brasil. Se trata de un estudio cuantitativo, no experimental, de naturaleza instrumental, que involucró a 405 docentes; y los resultados indican índices psicométricos adecuados y una estructura factorial satisfactoria y coherente con la propuesta tridimensional del constructo de actitudes. Los coeficientes estadísticos encontrados a partir de los análisis realizados legitiman su uso para el desarrollo de políticas públicas y prácticas efectivas en el campo educativo.

Palabras clave: psicometría, escala de actitudes, profesores, investigación educacional, análisis factorial

¹Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, Ouro Preto - MG, Brazil

²Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Rio de Janeiro - RJ, Brazil

³Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais, Passos - MG, Brazil

⁴Universidade Federal do Espírito Santo, Vitória - ES, Brazil

Support: Article derived from the research project entitled "Standardized evaluation in the states of Espírito Santo/Brazil and Baja California/Mexico: dilemmas and tensions of Paebes and Exeims-BC," funded by the Foundation for Research and Innovation Support of Espírito Santo, Public Notice No. 03/2021 – Universal, No. SIAFEM: 2021-1WK88.

Correspondence address: Denilson Junio Marques Soares. Instituto Federal de Educação, Ciências e Tecnologia de Minas Gerais - Campus Ouro Preto - Rua Pandiá Calógeras, 898, Bauxita, Ouro Preto, CEP 35400-000. E-mail: denilson.marques@ifmg.edu.br

In the educational context, assessments have evolved into a multifaceted domain, transcending specific theories, processes, and methods. This approach is manifested in the development and guidance of evidence-based public policies, which seek to establish quality standards aligned with the constant social, cultural, scientific, and technological transformations that drive innovation and knowledge production in an increasingly globalized world.

When assessments are implemented by external agents to the school, it is called external assessment, which is generally applied on a Large Scale, i.e., for a large number of people, providing important information for the monitoring of educational systems. Thus, external Large

Scale assessments have gained relevance in the national and international scenarios, being highlighted for their role as instruments in the development of public policies that impact teaching practice, aiming to improve the quality of teaching (Soares et al., 2022a).

Studies developed by educational researchers have highlighted this movement, which is aimed at understanding what impacts the policies of accountability, as manifested by these assessments, can bring to teaching practice (Baidoo-Anu & Ennu Baidoo, 2022). However, there is still a lack of studies in the literature aimed at assessing the attitudes of teaching professionals towards such assessment.

The *Scale of Attitudes towards External Assessments applied on a Large Scale* (EAAE), proposed by Soares et al. (2022a), was developed for this purpose. It is a 30-item instrument, composed of statements, which seeks to capture what basic education teachers think, feel, and how they behave towards this type of assessment. The EAAE employs a five-point Likert scale, 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 evaluated.

During EAAE development, content validity was analyzed by experts, based on the calculation of the Content Validity Coefficient (CVC), and a pilot application, conducted with a sample of the target population (Soares et al., 2022a). However, considering the potential of EAAE to support decision-making within the scope of Large Scale assessment policies, its validation process requires further studies to confirm its validity and reliability, as the use of scales with good psychometric parameters is essential to ensure the accuracy and usefulness of the results in different contexts, including education.

In this aspect, this research aimed to evaluate construct validity, via Factor Analysis, as well as scale reliability, via composite reliability, based on its application to a sample of teachers affiliated with the Education Department of Espírito Santo/Brazil.

The EAAE is part of the field of study on attitudes that originated in the early 20th century, based on the contributions of sociologists Allport (1935) and Thomas and Znaniecki (1919). In search of a definition that would fit the various theories and perspectives of the field, based on a systematic review of the literature on the concept of attitude, Eagly and Chaiken (1993) define it as a hypothetical construct related to a “a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (Eagly & Chaiken, 1993).

The specialized literature present various models to explain attitude, of which the tripartite model, introduced by Rosenberg and Hovland (1960), is the most relevant, as it encompasses multiple psychological factors (Mazana et al., 2019; Svenningsson et al., 2022). This model, assumed by EAAE, considers the following interrelated dimensions: cognitive, affective, and behavioral.

The cognitive component is related to the beliefs, perceptions, concepts, and knowledge that the individual

has about the attitudinal object, being usually elicited in its verbal or written form. The affective component, in turn, refers to feelings, emotions, and sensations, assuming a connotation of the individual’s evaluative judgment in relation to the object in question, positively or negatively. As indicated by Svenningsson et al. (2022), a special case of the affective dimension is interest, analytically understood as an emotional schema that also includes cognitive dimensions.

The theory of planned behavior (TPB), initially proposed by Ajzen and Fishbein (1980), proposes that these two components (cognitive and affective) determine, in part, the behavioral intention, which is the immediate motivational factor for the behavior itself. In this aspect, behavioral intention can be seen as a direct result of the affective-cognitive consistency of the subjects (Svenningsson et al., 2022). However, according to Ankiewicz (2019), this influence can be positive or negative, depending on other factors that also affect behavior, such as situational and cultural factors. The analysis of the internal structure of the instrument allows us to examine these relationships.

Methods

This is a quantitative, non-experimental, instrumental study (Carretero-Dios & Pérez, 2007), conducted with a cross-sectional design, which consists of the search for evidence of structural validity of the EAAE.

Participants

The EAAE was applied to a non-probabilistic convenience sample of 405 teachers linked to the Education Department of Espírito Santo/Brazil. After treating missing cases and extreme values (univariate and multivariate outliers), 367 responses were considered valid ($n = 367$). As the scale consists of 30 items, there are approximately 12.23 subjects per item, which is higher than the recommendations of Hair et al. (2021) that suggest, as a general rule, a minimum sample of 200 respondents and an ideal ratio of at least 10 subjects per item.

In this sample, there was a predominance of women (62.13%), whites (56.4%), graduates of federal universities (51.5%) with a teaching degree (72.48%) or a teaching degree and a bachelor’s degree (19.35%), who work in high school (77.11%), in a single school (61.85%), with an average workload of 36 hours ($SD = 10.18$), and on a permanent basis (51%). The age of participants ranged from 22 to 69 years ($M = 40.5$, $SD = 9.56$) and the average time of teaching was 13 years ($SD = 9.32$). Regarding schooling, 16.08% held a bachelor’s degree as their highest qualification, 54.5% were specialists, 21.53% were masters, 6.81% were doctors, and 1% were post-doctorates.

The participating teachers worked in 29 of the 78 municipalities of Espírito Santo, of which most lived and worked in the Metropolitan Region of Greater

Vitória (64.31%). As for the subject they teach, most are in the area of Languages (28.34%), followed by Natural Sciences (21.25%), Human Sciences (19.62%), and Mathematics (18.80%), respectively. Approximately 12% of participants reported working in other disciplines or did not want to state the discipline in which they work.

Instruments

The *Scale of Attitudes towards External Assessments applied on a Large Scale* (EAAE) was used. It consists of 30 items, elaborated in the form of assertions and structured on a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). It seeks to capture what basic education teachers (target audience) think, feel, and how they behave towards this type of assessments that are applied on a Large Scale in Brazil.

To this end, it considers the attitudes construct, composed of the cognitive (12 items), affective (8 items), and behavioral (10 items) dimensions. Each dimension is accompanied by a guiding phrase: for the cognitive dimension, participants were requested to answer based on what they believe (beliefs, knowledge, information, and/or opinions) towards external assessments applied on a Large Scale; for the affective dimension, a response based on feelings was requested; and, for the behavioral dimension, a response based on daily actions was requested.

The values from attitude measurement, both for the general scale and for each of its dimensions, were obtained from the arithmetic addition of the answers given by the participant on the respective items. Thus, the EAAE score varies from 30 to 150 points, with a neutral score of 90 points. In summary, higher values (above the neutral point) reveal more positive attitudes and, on the other hand, lower values (below the neutral point) indicate more negative attitudes towards external assessments applied on a Large Scale.

Procedures

Data collection. The scale was applied in the online (38.15%) and face-to-face (61.85%) formats. For the online method, an online survey form was used, hosted on the Google Forms platform, with dissemination via e-mails, sent by the Education Department of Espírito Santo (SEDU), and via the WhatsApp messaging application, in specific groups of teachers in the network. For the face-to-face method, subjects were approached in their workplaces.

Notably, in both methods, participants signed an informed consent form before responding, which explained the objectives of the study and ensured the confidentiality of the information provided. In the description of the applied instrument, it was emphasized that participation was voluntary, and it was possible to abandon it at any time, without penalty. The absence of right or wrong answers was also emphasized, and the anonymity of the participants was assured.

Data analysis. Initially, descriptive statistics were estimated for the scale score and its dimensions. Then, to assess the factorial structure of the EAAE and thus search for evidence of construct validity, an Exploratory Factor Analysis (EFA) was performed. To verify the possibility of factoring the data, two indices were analyzed: the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, which needs to be at least 0.60 to support this type of analysis, and the Bartlett's test of Sphericity, whose chi-square value must be statistically significant (Tabachnick & Fidell, 2007).

The analysis was implemented using a polychoric correlation matrix and robust diagonally weighted least squares (RDWLS) extraction method. To define the number of factors to be extracted, the Parallel Analysis technique was used with random permutation of the observed data and the assumed rotation was the Robust Promin (Timmerman & Lorenzo-Seva, 2011). The Hull method was also used to aid in deciding the number of dimensions to be retained (Timmerman & Lorenzo-Seva, 2011).

To confirm the hypothetical factor structure found via the EFA, a confirmatory factor analysis (CFA) was performed to verify whether the hypothetical factor structure was adequate to the observed variables, thus consolidating the theoretical model previously identified by the EFA (Hair et al., 2021).

The adequacy of the model was assessed using the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI) fit indexes. According to the literature, adequate RMSEA values should range from 0.05 to 0.08, which can go up to 0.10, and CFI and TLI values should be above 0.90, or preferably 0.95 (Timmerman & Lorenzo-Seva, 2011).

The factor loadings and thresholds of the items were also assessed. These indicators were analyzed to investigate in depth the accuracy of the items, using factor loadings, as well as the difficulty limits of the items, using thresholds, assessed using the Reckase parameterization (Reckase, 1985).

Finally, a Gaussian graphical model was estimated, regularized by L1 regularization technique (LASSO) with the selection of the EBIC model, which was presented in a network structure, in which the nodes represent the questionnaire items and the lines (edges) represent the relationship between the questionnaire items, aiming to identify the strength of the correlation between them (Epskamp & Fried, 2018).

Ethical Considerations

The research was approved by the Research Ethics Committee on Human Subjects of the Universidade Federal do Espírito Santo, CAAE No. 57014722.2.0000.5542 and was authorized by the Education Department of Espírito Santo.

Results

Table 1 shows the descriptive statistics estimated for the applied EAAE and its cognitive, affective, and behavioral

dimensions. Note that the variation in scores indicates the relevance of the scale to discriminate positive and negative attitudes towards external assessments applied on a Large Scale.

Regarding the factorial structure of the scale, the Bartlett's sphericity test ($131.02, gl = 29, p < 0.001$) and

KMO (0.953) suggested the interpretability of the correlation matrix of the items. The parallel analysis and the Hull method suggested three factors as being the most representative for the data, as indicated by Figures 1 and 2, respectively. In the first, the eigenvalues and random data obtained from the resampling process via bootstrap methods were presented.

Table 1
Descriptive Statistics of the EAAE Application

Scale and its dimensions	Quantity of items	Mean Score	Standard Deviation	Coefficient of Variation	Range
Cognitive	12	34.55	10.00	28.94%	12–60
Affective	8	22.32	7.60	34.05%	8–40
Behavioral	10	34.7	7.80	22.48%	10–50
Full Scale	30	91.57	21.97	23.99%	30–150

Figure 1
Parallel Analysis Scree Plots

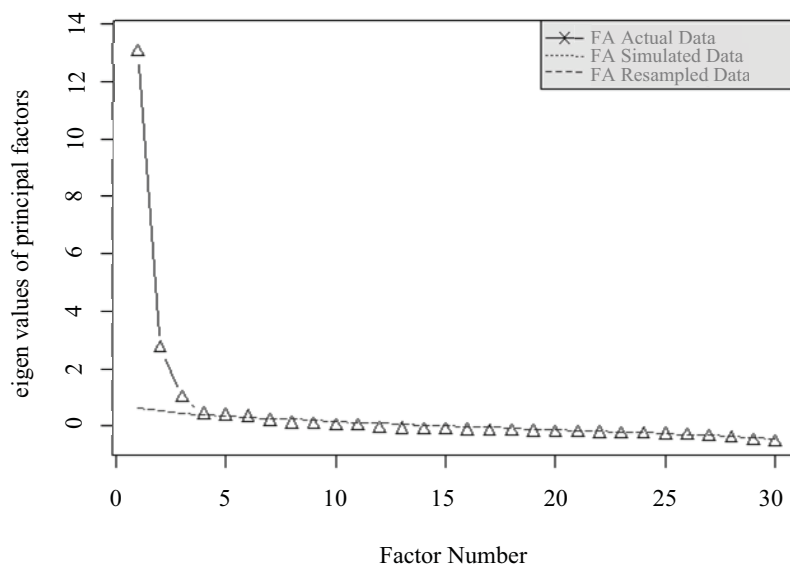
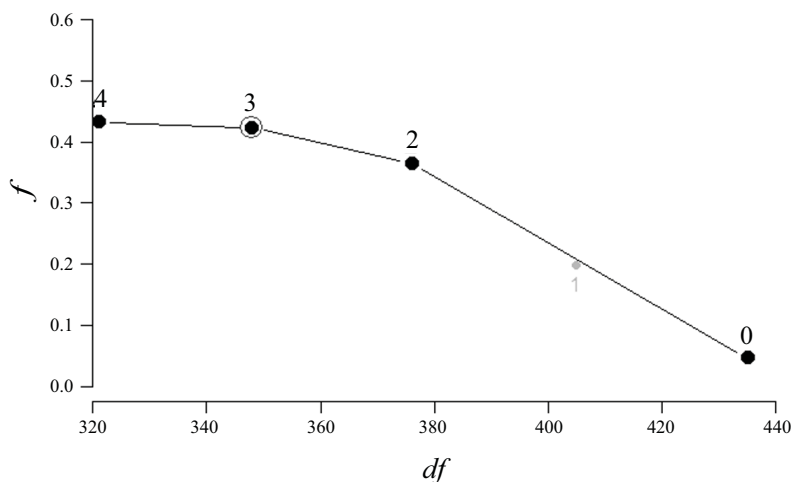


Figure 2
Hull Method



From the determination of the three-dimensional model for the structuring of the scale, confirmed by the CFA, we sought to estimate the factor loadings and thresholds of the items, which

can be observed in Table 2. The variance explained by the scale and its dimensions was also reported, as well as the composite reliability indices (Timmerman & Lorenzo-Seva, 2011).

Table 2
Psychometric analyses of EAAE

Item	Factor I	Factor II	Factor III	Threshold _{1,2}	Threshold _{2,3}	Threshold _{3,4}	Threshold _{4,5}
1 – Adequately assess the quality of teaching and learning	0.734			-1.232	-0.017	0.471	1.843
2 – They are the main instrument for monitoring the effectiveness of educational policies	0.615			-1.470	-0.463	-0.003	1.430
3 – Satisfactorily fulfill the purpose of measuring students' learning levels	0.792			-1.189	0.072	0.565	1.773
4 – They are useful to properly select students for other stages/levels of education	0.722			-1.277	-0.106	0.448	1.773
5 – Impact the selection of content taught in school	0.484	0.312		-1.843	-0.788	-0.280	1.149
6 – Adequately measure the level of knowledge of students based on the National Common Curricular Base (BNCC)	0.742			-1.277	-0.099	0.541	1.682
7 – Its curricular matrices give equal value to all the knowledge provided for in the National Common Curricular Base (BNCC)	0.601			-1.217	-0.210	0.426	1.555
8 – They allow the analysis of the social and cultural aspects of the students	0.707			-1.014	-0.196	0.345	1.603
9 – They are necessary instruments to reduce educational inequalities	0.829			-1.003	-0.044	0.418	1.511
10 – They are important instruments of accountability to society	0.764			-1.176	-0.330	0.338	1.533
11 – They are important instruments for monitoring teaching practice	0.820			-0.969	-0.182	0.345	1.710
12 – They are efficient instruments to promote financial bonuses to education professionals	0.751			-0.926	0.024	0.581	1.682
Items related to the cognitive dimension (12 items)							
13 – I appreciate this kind of assessment	0.371		0.520	-1.025	-0.168	0.557	1.773
14 – I yearn for finding their results			0.471	-1.325	-0.581	-0.017	1.470
15 – I trust their results			0.725	-1.025	-0.134	0.517	1.682
16 – I trust the comparisons between schools and students based on them			0.844	-0.875	0.010	0.597	1.773
17 – I trust the certifications of completion of a level/stage of education that are performed from them			0.772	-1.003	-0.168	0.541	1.806
18 – I feel that my knowledge is valued by these assessments			0.729	-0.855	-0.038	0.647	1.881
19 – I realize that the subject I teach is valued by them			0.700	-0.826	-0.134	0.441	1.555
20 – I like to align my work based on their results			0.802	-1.014	-0.381	0.323	1.603
Items related to the affective dimension (8 items)							
21 – I plan my classes based on their results		0.744		-1.393	-0.605	-0.003	1.376
22 – I incorporate their results into my daily actions in the classroom		0.748		-1.533	-0.798	-0.161	1.430
23 – I organize the school schedule to value its accomplishment		0.718		-1.533	-0.807	-0.127	1.232
24 – I prioritize the teaching of the contents that are most frequent in these assessments		0.739		-1.555	-0.836	-0.106	1.341
25 – I talk to students about its importance		0.743		-2.208	-1.393	-0.698	0.855
26 – I recommend participating in these evaluations		0.694		-2.018	-1.511	-0.875	0.798
27 – I provide feedback on the results obtained in the classroom		0.808		-1.490	-0.980	-0.287	1.162
28 – I solve questions from previous external assessments in class		0.873		-1.555	-0.948	-0.389	1.123
29 – I use questions from previous external assessments in internal school assessments		0.740		-1.430	-0.707	-0.245	1.189
30 – I participate in specific training for a better use of their results		0.592		-1.176	-0.533	0.182	1.358
Items related to the behavioral dimension (10 items)							
Explained Variance = 64%	Factor I 25%	Factor II 21%	Factor III 18%				
Composite reliability	0.9375430	0.9389661	0.7222134				

As Table 2 shows, the factor structure confirms the differentiation of the three dimensions of the attitude construct. Factor I, which presented the highest percentage of explained variance ($R^2 = 25\%$), integrates the items related to the cognitive dimension; Factor II includes all those related to the behavioral component ($R^2 = 21\%$); and Factor III, all those related to the affective component ($R^2 = 18\%$).

As for the factor loadings, associated with the precision of the items, there are adequate and relatively high values in their respective factors, ranging from 0.484 to 0.873 on the scale. Only two items presented a cross-load pattern (i.e., items with factor loadings above 0.30 in more than one factor), namely Item 5 and 13. However, Pratt's importance measures (Wu & Zumbo, 2017) demonstrated that both items were more strongly explained by their original factors. The fit indices of the instrument were adequate ($\chi^2 = 777.708$, $gl = 348$; $p < 0.001$; RMSEA = 0.058 (0.053 – 0.064); CFI = 0.997; TLI = 0.996). The composite reliability of the factors was also acceptable (above 0.70) for all factors.

Regarding the thresholds, estimated via the item response theory, no unexpected pattern of response was found, with a gradual increase in the difficulty of response along the interval scale, that is, as the response category on the scale increased, so did the level of latent trait required for endorsement. Thus, the difficulty is greater when the answer option of the item is closer to the alternative “I totally agree.” Thus, items 1 “Adequately assess the quality of teaching and learning” and 18 “I feel that my knowledge is valued by them” presented greater difficulty and items 25 “I talk to students about its importance” and 26 “I recommend participation in these assessments” were easier to answer.

The correlations obtained between the cognitive factor and the affective and behavioral factors were 0.778 and 0.574, respectively. Between them (affective and behavioral), the correlation was 0.530. The networks of partial correlations between the EAAE items are represented in Figure 3, in which the size and density of the edges between the nodes (which represent the EAAE items) indicate the strength of the existing correlation.

Figure 3

Partial correlation networks between EAAE items

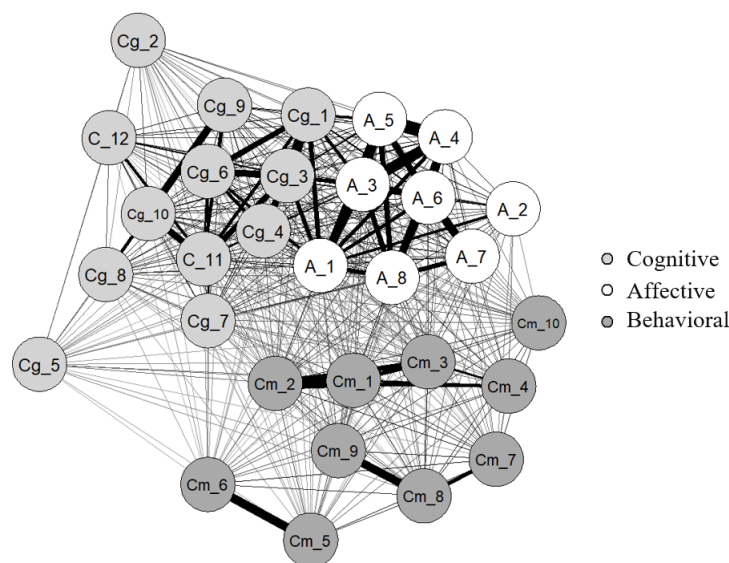


Figure 3 reveals a network with many connections between the nodes, and especially strong connections emerge within each factor and between some items of the cognitive and affective factors, reinforcing the existence of a strong correlation between them. From Figure 3, it is also possible to infer, although subjectively, the tripartite structure of the EAAE.

Discussion

Initially, it should be noted that the mean values obtained by the cognitive and affective dimensions of the scale, indicated in Table 1, reflect, for the sample analyzed,

negative attitudes towards external assessments applied on a Large Scale. On the other hand, the behavioral dimension indicates positive attitudes, signaling that despite negative beliefs and feelings, in general, the teachers have positive responses in the behavioral component, reflecting aspects of cognitive dissonance (Yahya & Sukmayadi, 2020). For the scale as a whole, the mean score also indicates positive attitudes.

This result is revealing and dialogues with the significant increase in assessment systems based on accountability policies on exams and the use of management models based on corporate rationality, called the Global Education Reform Movement (GERM) (Falabella, 2021). These policies involve credentialing, promotion, and inspection processes,

as well as rewarding or punishing schools and teachers, resulting in greater school control. This management model reflects the market ideology in education.

Parcerisa et al. (2022) postulate that these policies regularize teaching practice, outlining its behaviors to the State's intentions around external assessments. Thus, even if teachers disagree with Large Scale assessment policies, they adopt consonant practices via political mechanisms of coercion created by the State, which may explain the cognitive dissonance evidenced.

It should be noted that the state of Espírito Santo, Brazil, has assumed external assessment as the central axis of educational policies since the beginning of this century. To this end, several initiatives have been implemented to monitor student performance and the quality of education offered by schools, which are used as a basis for decision-making and the implementation of public policies towards education.

Among them, we highlight: the emergence of the Basic Education Assessment Program of Espírito Santo (PAEBES), in 2000, with the declared objective of assessing the performance of the state public network of elementary and secondary education and its reformulations; the implementation of the new common curricular base, based on the notions of competencies and skills (Espírito Santo State Department of Education, 2009); and the establishment of the School Development Index of Espírito Santo (IDE), and the performance bonus policy (Complementary Law No. 504 of 23 November 2009).

In summary, the performance bonus policy provides monetary rewards to teachers and other education professionals in the state who have achieved pre-established educational goals, based on the results of the PAEBES. The amount to be received can reach up to 150% of the teacher's base salary, which represents a significant amount for the category, with the potential to influence their daily school activities.

This practice has generated some criticism and controversy regarding its effectiveness. Soares et al. (2022b), for example, showed that bonuses can be seen as a way of pressuring teachers to achieve results at any cost, without considering the real working conditions and difficulties faced in the day-to-day life of the classroom. Moreover, the authors argue that the vertical way in which the policy was implemented generated harmful competition among teachers, pressuring them to focus only on the content that is evaluated by PAEBES, to the detriment of other important areas of knowledge, which would amount to a gaming and score inflation tactic (Baidoo-Anu & Ennu Baidoo, 2022).

On the other hand, some teachers see this policy as a way of recognizing and valuing their work, allowing an increase in their remuneration. These teachers believe that bonuses can encourage the improvement of student performance and, consequently, improve the quality of the education offered (Soares et al., 2022b). This contrast of opinions can be identified by means of the high indices obtained for the coefficient of variation obtained for the total scale and its

dimensions. This statistic reveals the absence of a uniform conception of the object among the sample participants.

Regarding the analysis of the internal structure, the results indicate that the EAAE is a tool with adequate psychometric indicators and satisfactory factorial structure that is consistent with the three-dimensional proposal of the attitude construct, given that the factor analysis revealed three factors that allowed to explain 64% of the total variance. In addition, the items presented adequate and high factor loadings in their respective factors, whose composite reliability was also acceptable. The variation in the scores also points to the relevance of the EAAE to discriminate positive and negative attitudes towards the assessments.

It should be noted that the use of scales with good psychometric parameters is essential to ensure the accuracy and usefulness of the results in different contexts, including education. In this respect, the statistical coefficients found in the psychometric analyses legitimize its use. Therefore, by applying it to a specific target audience, the EAAE can generate discussions and reflections on the impact of assessments on teaching practice, their relationship with social/demographic/economic variables, with the results achieved by different school units, among others, which can contribute to the development of more effective public policies and educational practices.

It is also necessary to highlight the association between the items of the cognitive and affective dimensions of the EAAE evidenced by Figure 3, which, in fact, was expected. As indicated by the literature, the beliefs and thoughts a person has about an object influence their emotions and feelings associated with it (Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960). In the case of this scale, it is understood that a professional who believes that external assessments applied on a Large Scale "adequately assess the quality of teaching and learning" (Item 1) and/or "satisfactorily fulfill the purpose of measuring students' learning levels" (Item 3) may feel positive emotions, such as appreciation for this type of assessment (Item 13), which explains the magnitude of the connections between these items.

However, it is necessary to consider the external validity of the scale, in terms of generalizing its results to other populations or contexts. The sample analyzed in this study was extracted from a context recognized by the specialized literature as an *Evaluative State*, which decision-making and allocation of technical and financial resources in the educational field is based on the metadata produced by external assessments applied on a Large Scale (Costa et al., 2019). It is important to consider this limitation in the interpretation of the results obtained, as well as in the application of the EAAE in other populations.

Moreover, other factors or variables may affect the attitudes of the subjects investigated that are not being measured by the scale. This may limit this construct validity and requires further studies. In fact, the evidence of validity of any instrument needs to be continuously verified and, thus, subsequent psychometric studies should be performed to investigate them in different contexts.

References

- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice-Hall.
- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 798-844). Clark University Press.
- Ankiewicz, P. (2019). Alignment of the traditional approach to perceptions and attitudes with Mitcham's philosophical framework of technology. *International Journal of Technology and Design Education*, 29, 329-340. <https://doi.org/10.1007/s10798-018-9443-6>
- Baidoo-Anu, D., & Ennu Baidoo, I. (2022). Performance-based accountability: Exploring Ghanaian teachers perception of the influence of large-scale testing on teaching and learning. *Education Inquiry*. Advance online publication. <https://doi.org/10.1080/20004508.2022.2110673>
- Carretero-Dios, H., & Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, 7(3), 863-882. <https://www.redalyc.org/pdf/337/33770319.pdf>
- Costa, A. G., Vidal, E. M., & Vieira, S. L. (2019). Avaliação em larga escala no Brasil: Entre a coordenação federativa e o ethos do Estado-avaliador [Large-scale evaluation in Brazil: Between federative coordination and the ethos of the evaluative State]. *Revista Educação em Questão*, 57(51), 1-29. <https://doi.org/10.21680/1981-1802.2019v57n51ID15806>
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt, Brace, & Jovanovich.
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617-634. <https://doi.org/10.1037/met0000167>
- Falabella, A. (2021). The seduction of hyper-surveillance: Standards, testing, and accountability. *Educational Administration Quarterly*, 57(1), 113-142. <https://doi.org/10.1177/0013161X20912299>
- Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). An introduction to structural equation modeling. In J. F. Hair, Jr., G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, & S. Ray, *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook* (pp. 1-29). Springer.
- Lei Complementar No. 504. (2009, 23 de novembro). Institui a Bonificação por Desempenho, no âmbito da Secretaria de Estado da Educação-SEDU, e dá outras providências. *Diário Oficial dos Poderes do Estado* (p. 1). <https://www3.al.es.gov.br/Arquivo/Documents/legislacao/html/LEC5042009.html?identificador=320036003500310032003A004C00>
- Mazana, M. Y., Montero, C. S., & Casmir, R. O. (2019). Investigating Students' Attitude towards Learning Mathematics. *International Electronic Journal of Mathematics Education*, 14(1), 207-231. <https://doi.org/10.29333/iejme/3997>
- Parcerisa, L., Verger, A., Pagès, M., & Browes, N. (2022). Teacher autonomy in the age of performance-based accountability: A review based on teaching profession regulatory models (2017-2020). *Education Policy Analysis Archives*, 30(100), 1-29. <https://doi.org/10.14507/epaa.30.6204>
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9(4), 401-412. <https://doi.org/10.1177/014662168500900>
- Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1-14). Yale University Press.
- Secretaria de Estado da Educação do Espírito Santo. (2009, 17 de novembro). *Portaria SEDU nº 143-R*. Estabelece Diretrizes para a Organização Curricular nas Escolas Estaduais para o Ano Letivo 2010 [SEDU Ordinance No. 143-R. Establishes Guidelines for Curricular Organization in State Schools for the 2010 Academic Year]. <https://www.legisweb.com.br/legislacao/?id=126933>
- Soares, D. J. M., Soares, T. E. A., & Santos, W. (2022a). Evidências de validade de conteúdo da escala de atitudes frente às 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)]. *Revista Ibero-Americana de Estudos em Educação*, 17(3), 1806-1818. <https://doi.org/10.21723/riiae.v17i3.16191>
- Soares, D. J. M., Soares, T. E. A., & Santos, W. (2022b). Análise do indicador de desenvolvimento das escolas estaduais do Espírito Santo [Analysis of the development indicator in Espírito Santo state schools]. *Educação & Sociedade*, 43(1), e256722. <https://doi.org/10.1590/ES.256722>
- Svenningsson, J., Höst, G., Hultén, M., & Hallström, J. (2022). Students' attitudes toward technology: Exploring the relationship among affective, cognitive and behavioral components of the attitude construct. *International Journal of Technology and Design Education*, 32, 1531-1551. <https://doi.org/10.1007/s10798-021-09657-7>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon.
- Thomas, W. I., & Znaniecki, F. (1919). *The Polish peasant in Europe and America: Monograph of an immigrant group* (Vol. 3). University of Chicago Press.

- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209-220. <https://doi.org/10.1037/a0023353>
- Wu, A. D., & Zumbo, B. D. (2017). Using Pratt's importance measures in confirmatory factor analyses. *Journal of Modern Applied Statistical Methods, 16*(2), 81-98. <https://doi.org/10.22237/jmasm/1509494700>
- Yahya, A. H., & Sukmayadi, V. (2020). A review of cognitive dissonance theory and its relevance to current social issues. *MIMBAR: Jurnal Sosial Dan Pembangunan, 36*(2), 480-488. <https://doi.org/10.29313/mimbar.v36i2.6652>

Denilson Junio Marques Soares is a Professor of the Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, Ouro Preto-MG, Brazil.

Wallacy Campos Prado is a Professor of the Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Rio de Janeiro-RJ, Brazil.

Talita Emidio Andrade Soares is a Professor of the Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais, Passos-MG, Brazil.

Wagner dos Santos is a Professor of the Universidade Federal do Espírito Santo, Vitória-ES, Brazil.

Authors' Contribution:

All authors made substantial contributions to the conception and design of this study, to data analysis and interpretation, and to the manuscript revision and approval of the final version. All the authors assume public responsibility for the content of the manuscript.

Associate editor:

Sônia Maria Guedes Gondim

Received: Aug. 06, 2023

1st Revision: Jan. 09, 2024

Approved: Feb. 10, 2024

How to cite this article:

Soares, D. J. M., Prado, W. C., Soares, T. E. A., Santos, W. (2024). Construct validity and reliability of the scale of attitudes towards external assessments applied on a large-scale. *Paidéia (Ribeirão Preto), 34*, e3410. doi: <https://doi.org/10.1590/1982-4327e3410>