

Construction and study of validity evidence of the Teaching Assessment Scale

MÔNIA APARECIDA DA SILVA

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

WAGNER DE LARA MACHADO

Pontifícia Universidade Católica de Campinas, SP, Brazil

LUCIANE MARIA PILOTTO

Centro Universitário Univates, Lajeado, RS, Brazil

BÁRBARA BACKES

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

REGINA BASSO ZANON

Sociedade Educacional Três de Maio, Três de Maio, RS, Brazil

PATRÍCIA VASCONCELOS MACHADO

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

ANA PAULA COUTO ZOLTOWSKI

Instituto Brasileiro de Gestão de Negócios, Porto Alegre, RS, Brazil

REBECA VERAS DE ANDRADE VIEIRA

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

RENATA GIULIANI ENDRES

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

MARTINA FRANCALANCI

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

VICTOR ADRIEL DE JESUS OLIVEIRA

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

JEFFERSON SILVA KRUG

Pontifícia Universidade Católica de Porto Alegre, RS, Brazil

DENISE RUSCHEL BANDEIRA

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

ABSTRACT

The objective of this study was to present the construction of the Teaching Assessment Scale (TAS) and the validity evidence for it. This scale was developed through a literature review and interviews with graduate students. The evidence for content validity of the TAS was evaluated by ten referees, and this was followed by the analysis of the evidence for construct validity and reliability. The tetrachoric correlation matrix was submitted for exploratory factorial analysis, and the Hull method was used to decide the number of dimensions to be retained. Item response theory (IRT) analysis was performed using the rating scale model with the result that seven items needed to be excluded. The Kaiser-Meyer-Olkin (KMO) index and Bartlett's Test of Sphericity indicated that the polychoric correlation matrix was factorable. The Hull method suggested the retention of a factor, with the eigenvalue of 15.49. The factor's reliability measures were $\alpha=0.96$ and $\omega=0.95$. As a result, the TAS is considered helpful in evaluating higher education teaching methods in Brazil.

KEYWORDS

Teaching Assessment Scale; self-assessment; education higher; item response theory.

CONSTRUÇÃO E ESTUDO DE EVIDÊNCIAS DE VALIDADE DA ESCALA DE AVALIAÇÃO DOCENTE

RESUMO

Este estudo objetivou apresentar a construção da Escala de Avaliação Docente (EADoc) e suas evidências de validade. A escala foi desenvolvida com base em uma revisão de literatura e de entrevistas com estudantes de graduação. A EA-Doc foi avaliada por dez juízes quanto a evidências de validade de conteúdo, com subsequente análise de evidências de validade de construto e fidedignidade. A matriz de correlação tetracórica foi submetida a análises fatoriais exploratórias, e o método Hull foi usado para a decisão do número de dimensões a serem retidas. Foram conduzidas análises de teoria de resposta ao item (TRI), pelo modelo *rating scale*, que excluiu sete itens. O índice Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett indicaram a fatorabilidade da matriz de correlações policóricas dos itens. O método Hull apontou a retenção de um fator, com autovalor de 15,49. As medidas de fidedignidade para o fator foram de $\alpha=0,96$ e $\omega=0,95$. Considera-se que a EADoc pode auxiliar na avaliação de práticas da educação superior no Brasil.

PALAVRAS-CHAVE

Escala de Avaliação Docente; autoavaliação; ensino superior; teoria de resposta ao item.

CONSTRUCCIÓN Y ESTUDIO DE PRUEBAS DE VALIDEZ DE LA ESCALA DE EVALUACIÓN DOCENTE

RESUMEN

En este estudio se presentan los procesos de construcción de la Escala de Evaluación Docente y sus pruebas de validez. La escala fue desarrollada a partir de una revisión de la literatura y entrevistas con estudiantes de pregrado. La escala fue sometida a diez jueces para investigar evidencias de validez de contenido, seguido por el análisis de evidencias de validez de constructo y fiabilidad. La matriz de correlaciones tetracóricas se sometió a análisis factorial exploratorio y el método HULL se utilizó para decidir el número de dimensiones a ser retenidas. Se realizaron los análisis de Teoría de Respuesta a los Ítems (TRI) por el modelo de escala de clasificación y se excluyeron siete ítems. El índice Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett y el índice de KMO indica la factorabilidad de la matriz de correlaciones policóricas de ítems. El método HULL indicó la retención de un factor con el valor propio de 15,49. Las medidas de fiabilidad para el factor fueron $\alpha=0,96$ y $\omega=0,95$. Se considera que la escala puede ayudar en la evaluación de las prácticas de enseñanza superior en Brasil.

PALABRAS CLAVE

Escala de Evaluación Docente; autoevaluación, educación superior; teoría de respuesta a los ítems.

INTRODUCTION

According to Korthagen (2004), the definition of a good teacher consists of a complex construct that involves both objective and subjective aspects. The author reports that, starting in the 1950s, there was a significant effort on the part of researchers to understand this construct *vis-à-vis* the teacher's skills. Therefore, studies were developed to correlate the teachers' behaviors to the students' learning, resulting in lists composed of behavioral indicators identified as the abilities of a good teacher.

In the 1970s, researchers started to base their studies on a new paradigm called Humanistic Based Teacher Education (HBTE). This approach assumed that teachers had personal characteristics capable of influencing their professional conduct, apart from the isolated role of their competencies (Zeichner, 1983). However, HBTE met with resistance from many specialists, and both approaches have been targets of debate and controversy up to today.

Korthagen (2004), with the aim of understanding what defines a good teacher in a more holistic way, proposes a model involving different components: environment, behavior, competencies, beliefs, identity, and purpose. Environment encompasses aspects such as the classroom and the students, while behavior refers to the conduct of the teacher himself as contrasted with, for example, his activities and his students. Competencies cover, among other aspects, the knowledge and skills of teachers in relation to their teaching technique. These competencies are directly influenced by the teacher's beliefs (the fourth component), more specifically those related to the teaching-learning process. Identity refers to how teachers understand their role as professionals and is directly related to purpose, which involves their reasons for being a teacher, and to the inspiration for pursuing the occupation. According to Korthagen (2004), basing the definition of a good teacher on this model permits to understand this construct in all its complexity, without ignoring fundamental aspects that influence the quality of professional performance, such as beliefs and identity. In this manner, starting with a broad conceptualization of the characteristics of a good teacher, three dimensions which seek to synthesize the levels described by Korthagen (2004) are presented:

1. what a good teacher is;
2. what a good teacher knows;
3. how a good teacher works.

These dimensions underpin the construction of the Teaching Assessment Scale (TAS), intended for higher education teachers, which will be described ahead.

The "who a good teacher is" dimension primarily involves the notions of professional identity and self-concept (Korthagen, 2004). Self-concept can be understood as the personal significance of aptitudes, interests, values and choices, which jointly comprise the unique characteristics of individual in fulfilling their professional role (Super; Savickas; Super, 1996). When it comes to the teacher, professional identity is consequently understandable as a collection of self-concepts, that is, interests, values, goals, and personal history that are considered vocationally relevant (Korthagen, 2004). It deals, therefore, with a realm involving relationship

skills, personality types, and emotional states. The following elements are situated in this dimension: enjoyment of what one does; valuing one's profession; engagement with the learning process; displaying professionalism; valuing the relationship with students; building a supportive environment; demonstrating unconditional respect and positive consideration for students; being open-minded; having a sense of professional identity; being empathic, understanding, and respectful with others; having sense of humor; and being flexible and responsible (Korthagen, 2004; Osmun; Copeland, 2011; Pachane, 2012; Sutkin *et al.*, 2008).

Related to the "what a good teacher knows" dimension, Shulman (1987) notes that the teacher's knowledge is characterized by area of expertise as much as by content, knowledge of the curriculum, and knowledge of educational contexts. According to Harden and Crosby (2000), the teacher is responsible for transmitting new and relevant knowledge in an appropriate manner. The authors, when theorizing about the roles played by the teacher, stress that, in the process of transmitting information, good teachers should be capable of sharing personal thoughts and reflections with the students, making their line of reasoning and their particular view on the field clear. In this sense, they could, for example, present information that the student would not normally find in texts on the subject, as well as draw links between the content, the curriculum, and other aspects of the subject area in a critical and reflective manner. It needs to be stressed, therefore, that good teachers ought to have knowledge of discipline's content (its substance, logic, form, and epistemology) and the ability to reflect deeply upon that which they propose to teach. Accordingly, one can understand this dimension as involving attributes related to the realms of content, experience, practical and theoretic bases, continuing education, and practical and theoretical revision (Azer, 2005; Feitoza; Cornelsen; Valente, 2007; Pachane, 2012).

Finally, the third and last dimension is called "how a good teacher works". Regarding the word *work* in this sense, Reed (1989) asserted that it is not so much the content that is important as its preparation and presentation. To the author, the primary purpose of a class should be to provide pertinent materials, presented in a stimulating manner and designed to facilitate comprehension. Concurrent with this assertion, among the techniques used by the good teacher, studies cite: using exhaustive questioning to engage students in discussion; knowing how to deliver the right amount of supervision and independence; developing a supportive relationship with students; emphasizing solutions to problems; giving feedback; being clear in presenting content; having control of the classroom; knowing how to relate theory to practice; using appropriate language; stimulating high-level critical thinking skills; presenting difficult concepts in an understandable form; employing diverse teaching strategies; and stressing teamwork and collaborative learning (Azer, 2005; Feitoza; Cornelsen; Valente, 2007; Osmun; Copeland, 2011; Pachane, 2012).

The good teacher has been seen as a professional capable of motivating and engaging students regarding academic tasks and respecting their learning styles, above and beyond presenting themes in a clear, organized way. In the teacher's own perception, it is necessary to create a good plan and be well-prepared for class, ally practice with theory, be organized and communicative, like challenges, and be a motivator and conductor of the students' learning (Samples; Copeland, 2013; Duarte, 2013).

Thus, keeping in mind the variability and complexity of the aspects of the before-mentioned dimensions, it is important to introduce psychometrically valid metrics capable of measuring that construct. To this end, Avrichir and Dewes (2006) developed an instrument for self-assessment of teacher performance for graduate school professors. It consists of a scale intended to evaluate two factors:

1. teacher interest and challenge;
2. relationship with the student and evaluation.

Regarding to teacher efficacy, Tschannen-Moran and Hoy (2001) have developed the Ohio State Teacher Efficacy Scale instrument, composed of 24 items divided into three factors, which encompasses:

1. instructional strategy efficacy;
2. classroom management efficacy;
3. student engagement efficacy.

Harden and Crosby (2000), for their part, presented a model of the functions performed by teachers of medicine and an instrument for evaluating them. The authors emphasized that there are 12 primary functions involved in the teaching of medicine and developed a questionnaire to evaluate the perception of teachers concerning the importance of these functions and of their commitment to each of them. Also in the area of health, Stone *et al.* (2002) conducted semi-structured interviews with U.S. medical school teachers, from which emerged the following factors:

1. underlying humanitarianism;
2. familiarity with educational principles and practices;
3. appreciation of both the advantages and disadvantages of teaching;
4. self-image as a teacher.

By means of a qualitative case study approach, data from a two-year university developmental program was explored and analyzed with the result that factors surrounding teacher education fell into three groups:

1. personal (cognitive and emotional factors unique to the individual);
2. relational (connections and interactions with others);
3. contextual (the program itself and external work environments) (Lieff *et al.*, 2012).

None of the instruments already cited has been adapted to or is valid for use in the Brazilian context. In Brazil, the use of instruments for teacher evaluation and self-assessment in teaching institutions, particularly in those for higher education, does occur. However, these instruments are generally devised by the institutions themselves and have not usually been studied psychometrically before being used. It is likely that many of these instruments were created in response to the 2004 creation of the National System for Higher Education Evaluation (SINAES, or in Portuguese Sistema Nacional de Avaliação da Educação Superior). SINAES proposed an institutional evaluation composed of several complementary instruments

and included the implementation of a system of self-assessment to be undertaken within the institutions themselves via Internal Evaluation Committees. Despite following a set script, the institutional self-assessment is implemented through distinct instruments to which the higher education institutions (HEI) can add items they judge pertinent (Ristoff; Giolo, 2006). One of the most common and constant components of the self-assessment processes is done by the student on teaching work (Gomes; Borges, 2008). At the national level, teaching evaluation lacks instruments that consider the different aspects of the construct for evaluating the teacher's practice. Besides this, it is important to develop an instrument that specifically attempts to evaluate graduate-level educators' teaching practices in the Brazilian context, given the operating peculiarities and specifics in the field.

Consequently, the objective of this study was to present the construction of the instrument called the TAS — Professor Version, along with evidence for its validity.

METHOD

CONSTRUCTION OF THE TEACHING ASSESSMENT SCALE — PROFESSOR VERSION

To construct the scale, a literature review starting with the definitions of the three dimensions of the instrument — what a good teacher is, what a good teacher knows, and how a good teacher works — was carried out. Additionally, interviews with graduate students were conducted to survey how they define a good teacher. Students from various majors at public and private institutions in three Brazilian states (Rio Grande do Sul, Minas Gerais and Ceará) were contacted to be interviewed. The interviews took place in person or via e-mail, and the students were asked to describe a good graduate school teacher, considering their skills and competencies. Twenty-two graduate students participated in this stage. The number of interviews was determined according to the saturation criterion, that is, the interviews stopped when the students' descriptions of good teachers began to repeat (Fontanella; Ricas; Turato, 2008).

The students' responses pertaining to the characteristics of a good teacher were analyzed and categorized based on a qualitative content analysis (Bardin, 1977), thereby generating a series of items that composed the scale. The items were presented to a focus group of 11 graduate students. The group analyzed the items, adapting them to the criteria for item preparation proposed by Pasquali (2010). Items with highly similar content were excluded. In the end, 71 items, distributed among the three theoretical dimensions, were selected. The response options for the TAS were defined on a five-level Likert scale, in which 1 represents "never"/"not at all" and 5 "always"/"completely", according to how well the professor embodied the statements posed. At the end of this process, the instrument was dubbed the TAS — Professor Version.

The completed scale was submitted for evaluation to ten judges for objective assessment of evidence for content validity, as suggested in Pasquali (1998). All the

judges were teachers with graduate school teaching experience (from 2 to 25 years); some had a background in psychometry and others in education. The judges received both the dimension definitions and the dimension-sorted items. They were asked to evaluate the items' comprehensibility and relevance and pertinence to the assigned dimension. Space was also provided for them to note any writing problems or suggest the removal of items or the addition of new items.

The results of the judges' analyses were discussed by the authors, and items were modified or excluded as necessary. At the end of this stage, 57 items remained in the TAS. A pilot study was carried out to assess the preliminary characteristics of this version of the instrument. Thirty-two private and public university professors from 30 graduate programs with an average of 10.85 years' time teaching (standard deviation (SD) = 10.59) participated. Using the analysis of the responses from the pilot study, four items highly correlated (over 0.60) with other items were removed as they were considered redundant. The version of the TAS resulting from the pilot study thus contained 53 items.

EXAMINATION OF THE EVIDENCE FOR CONSTRUCT VALIDITY AND TRUSTWORTHINESS

For the analysis of the evidence for construct validity and trustworthiness of the TAS, 194 professors active in graduate programs in different regions of Brazil, particularly in the South (63.9%) and Southeast (26.8%), responded. Chart 1 summarizes the primary descriptive characteristics of the sample.

The majority of professors were female (61.9%), with average age of 41.88 years (SD=11.9) and average time active in higher education of 12 years (SD=10.2). The most common degree attained was doctorate (48.1%). The majority of professors was active in one or more graduate programs (primarily Psychology, Engineering and Administration), predominantly in private universities (53.6%), and the greater part was employed full time (59.3%).

PROCEDURES

This project was submitted and approved by Committee for Ethics in Research of the Institute of Psychology at the Universidade Federal do Rio Grande do Sul (UFRGS) (no. 450,393, 4 November 2013). The version of the TAS resulting from our construction process, along with a sociodemographic data questionnaire, was made available online via SurveyMonkey®. Those invited to participate were contacted by e-mail or social networks. The participants had to confirm their participation by accepting the Informed Consent Term before filling out the scale.

DATA ANALYSIS

Given the ordinal nature of the responses, exploratory factorial analysis using the unweighted least squares (ULS) and mean and variance adjusted weighted least squares (WLS) (Muthén; Du Toit; Spisic, 1997) extraction methods were carried out on the polychoric correlation matrix (Holgado-Tello *et al.*, 2010). The Hull method (Lorenzo-Seva, Timmerman, Kiers, 2011) was used to determine the number of dimensions to be retained for analysis.

This method contrasts the degrees of freedom with the adjustment indices of several possible solutions for the same matrix, and the factorial solution with the best balance between both parameters is retained. Following this, item response theory (IRT) analyses were undertaken using the rating scale model of Andrich (1978), adapted for polytomous items. The model independently estimates the difficulty of the items (δ) and the position of individuals on the estimated linear latent continuum (θ) in log-odds units (logits). The adjustments of items to the measurement model (infit and outfit), the dimensionality of the measurement residuals (principle contrasts), and the location dependence of the items (residual correlations) were estimated.

RESULTS

The Kaiser-Meyer-Olkin (KMO) index was 0.89, and Bartlett's test (4993.2, degree of freedom (df)=1378, $p < 0.001$) indicated that the polychoric matrix correlation of the items was factorable. The Hull method suggested retaining a factor, with the eigenvalue of 15.49. The one-dimensional model gave an adjustment of $\chi^2=2874.12$ (1325), $p < 0.001$, using the ULS method, and $\chi^2=2506.786$

Chart 1 – Descriptive data from the teacher sample (n = 194).

| Variable | n | % |
|------------------------------|-----|------|
| Gender | | |
| Female | 120 | 61.9 |
| Male | 74 | 38.1 |
| Highest degree earned | | |
| Bachelor's | 1 | 0.5 |
| Specialist | 8 | 4.1 |
| Master's | 74 | 38.2 |
| Doctorate | 111 | 57.2 |
| Type of university taught at | | |
| Public | 90 | 46.4 |
| Private | 104 | 53.6 |
| Region | | |
| South | 124 | 63.9 |
| Southeast | 52 | 26.8 |
| Center-West | 10 | 5.2 |
| Northeast | 8 | 4.1 |
| Type of employment | | |
| Full-time | 115 | 59.3 |
| Hourly | 79 | 40.7 |

(1325), $p < 0.001$, using WLS method. The reliability measurements for the factor were $\alpha = 0.96$ and $\omega = 0.95$. The factorial loads of the items were all suitable, being above 0.32.

Subsequently, the same set of items was analyzed using the rating scale model. The items' adjustment to the measurement model was analyzed using the infit and outfit values. These indices pertain to the residuals, and should always lie between 0.5 and 1.5. The closer to the unit, the better the adjustment between the level of difficulty of the item and level of the individuals' latent trait which either endorses or not the item in question. Chart 2 presents the results of this analysis. In it, the items are ordered by difficulty parameter, and the infit and outfit values are presented. The reliability measurement of the items estimated by the rating scale model was 0.96.

The dimensionality was obtained by means of principle contrasts in which the residual variance of the items is submitted to principle component analysis. The presence of contrasts with eigenvalues greater than or equal to 2 indicates possible dimensions influencing the response patterns beyond measurement. Finally, the residual correlations of the items were estimated to investigate local dependence cases between them. High residual correlations indicate that a particular endorsed item's probability is not determined solely as a function of a latent trait, but depends on the response of another item. According to these criteria, seven items were excluded (5, 9, 14, 28, 29, 46 and 49). The final version of the scale, with 46 items, is presented in Appendix 1.

DISCUSSION AND FINAL CONSIDERATIONS

From this study, a Brazilian instrument for self-assessment of teaching practice for graduate school teachers was constructed. The instrument may be useful for teachers in monitoring the activities that teaching demands, as well as for reflection and improvement of one's own practice. With its basis in literature, the conceptions of students regarding the practices of a good teacher, and the judgments of specialists, the TAS — Professor Version has potential to be a valuable instrument in the evaluation of higher education practices in Brazil.

Darling-Hammond (2010) states that the participation of teachers in the process of evaluating their own performance can reflect on the effectiveness of their teaching. This is because the evaluation is a diagnostic process which can form the basis for necessary changes in the actions of, as well as in the training of, these professionals. In addition, in the Center for American Progress' program Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching, the author suggests teachers who continually evaluate their own performance tend to adopt new and more effective teaching methodologies. This view of evaluation aligns with the one already recommended by Marques (2010). For that author, evaluation is a process of constant monitoring and improvement in the quality of higher education. Thus, there is a need for good instruments, such as the TAS, which enable systematic evaluation of teaching practice.

Additionally, it should be noted that there is a scarcity of studies of teaching evaluation instruments for the Brazilian context that takes into account this construct's various facets. On the other hand, one must recognize that many of the country's universities are developing their own measurements for this end, many of them evaluating the performance of teachers based on the perception of their students (Gomes; Borges, 2008). The information coming from these measurements are important for institutional monitoring and the formulation of specific programs aimed at improving the quality of instructions. However, these instruments are generally not being studied psychometrically and, thus, their quality is questionable. In this sense, it is understood that the TAS, having been constructed based on both theory and empirical data and respecting all the necessary psychometric procedures for the creation and validation of such an instrument (Pasquali, 1998; Urbina, 2007), can be useful for evaluating higher education teaching practice. After searching the Brazilian literature, it was determined that this is the first self-assessment instrument for Brazilian teaching created according to the before-cited standards.

In the process of construction, principles for creating such instruments based on those proposed by Pasquali (1999; 2010) were followed. This author recommends three procedure types be followed: theoretical, empirical, and analytical. Theoretical procedures include literature reviews, surveying the empirical evidence on the conception of a good teacher and creation of the instrument items. This step unfolds in the empirical procedures, which consist of the pilot study and data collection for analysis of the validity evidence. The utilization of statistical analyses for confirming the psychometric properties of the TAS corresponded to the analytical procedures and completed the construction process. Thus, TAS — Professor Version was constructed with the theoretical and methodological rigor recommended by the literature for creating such instruments was confirmed. The scale possesses psychometric qualities tested and approved for self-assessment of graduate school teaching practice and is in accord with international standards for educational and psychological testing (AERA, APA, NCME, 1999).

The analyses suggest the instrument has a one-dimensional structure, in contrast with the theoretical model of three dimensions adopted for its construction. One can imagine that the three theoretical dimensions might be difficult to separate empirically, and that, for example, a teacher's flexible personal style will be associated with his way of managing possible modifications to the lesson plan. Hence, the dimensions dynamically influence each other and compose distinct patterns of behavior. The parameter analysis of the item difficulty permitted establishing groups of items with similar contents, reflecting different levels of teaching ability. For example, in Chart 2, the easier items (item sequence 13 to 22) have their roots in the broader and more formal aspects, almost bureaucratic, of the function of teaching. They involve basic ways of being, doing, and knowing for the role, and are the minimum one would expect from a teacher. On the second level (item sequence 31 to 50), there are the skills involved in mastering, monitoring, and control of the teaching process. Relational aspects, such as incentive, feedback, and fostering a critical stance are integrated into a deeper level of the content being worked. Finally (the sequence from 19 to 3), in the last stage, are the advanced characteristics of a teacher, which

Chart 2 – Instrument’s finalized items, difficulty parameters, and items’ adjusted measurements.

| Item | δ | Infit | Outfit |
|---|----------|-------|--------|
| 19. Is flexible in adapting content | -1.42 | 1.17 | 1.18 |
| 38. Is capable of teaching differing students | -1.63 | 0.97 | 0.97 |
| 34. Masters several methodologies | -1.72 | 1.04 | 1.06 |
| 21. Uses diverse methodologies | -1.76 | 1.17 | 1.17 |
| 37. Asks for feedback about classes | -1.78 | 1.35 | 1.30 |
| 6. Keeps students’ attention | -1.84 | 0.63 | 0.67 |
| 30. Gives negative feedback | -1.87 | 1.10 | 1.14 |
| 36. Considers sociocultural diversity | -1.89 | 1.15 | 1.14 |
| 51. Handles conflicts | -1.96 | 0.63 | 0.63 |
| 35. Is organized | -2.25 | 1.35 | 1.46 |
| 25. Clarifies epistemological bases | -2.31 | 1.07 | 1.10 |
| 15. Is able to say no | -2.34 | 1.30 | 1.24 |
| 18. Guides searches for extra materials | -2.36 | 0.99 | 0.94 |
| 12. Demands concerning content | -2.40 | 1.22 | 1.19 |
| 8. Permits informality/humor | -2.41 | 1.09 | 1.09 |
| 41. Admits his/her knowledge is limited | -2.48 | 1.18 | 1.19 |
| 44. Is familiar with the course curriculum | -2.50 | 1.56 | 1.52 |
| 17. Is punctual | -2.54 | 1.24 | 1.45 |
| 3. Supplements content according to interest | -2.54 | 1.09 | 1.05 |
| 31. Adopts critical stance toward content | -2.59 | 0.84 | 0.80 |
| 48. Verifies that students are following | -2.61 | 0.74 | 0.74 |
| 20. Encourages critical appreciation | -2.62 | 1.11 | 1.12 |
| 16. Gives positive feedback | -2.63 | 0.86 | 0.83 |
| 45. Reconciles theory and research | -2.67 | 0.99 | 1.00 |
| 42. Is familiar with practical applications | -2.72 | 0.82 | 0.84 |
| 7. Answers questions outside of class | -2.74 | 1.35 | 1.28 |
| 33. Likes being with students | -2.84 | 1.10 | 1.09 |
| 39. Uses diverse resources | -2.87 | 0.94 | 0.98 |
| 4. Promotes students’ curiosity | -2.89 | 1.11 | 1.12 |
| 47. Clarifies doubts about the profession | -2.91 | 1.00 | 0.94 |
| 2. Fosters relevant discussion | -2.92 | 0.68 | 0.70 |
| 40. Promotes participation | -3.07 | 0.68 | 0.67 |
| 11. Relates linked ideas | -3.07 | 0.80 | 0.91 |
| 26. Likes to encourage autonomy | -3.08 | 0.66 | 0.67 |

Continue...

Chart 2 – Continuation.

| Item | δ | Infit | Outfit |
|--|--------------|-------------|-------------|
| 50. Demonstrates mastery of content | -3.11 | 0.73 | 0.73 |
| 13. Has sufficient theoretical knowledge | -3.18 | 0.91 | 1.41 |
| 52. Is clear about assessments | -3.20 | 1.10 | 1.07 |
| 27. Is up-to-date in the area being taught | -3.22 | 0.97 | 0.93 |
| 10. Speaks so that everyone can hear | -3.26 | 0.91 | 1.02 |
| 32. Reconciles theory with practice | -3.33 | 0.74 | 0.74 |
| 1. Is happy being a teacher | -3.41 | 1.11 | 1.04 |
| 43. Is open to dialog | -3.53 | 0.76 | 0.75 |
| 23. Gives examples of applications of theory | -3.64 | 0.77 | 0.70 |
| 53. Helps with doubts | -3.75 | 0.73 | 0.72 |
| 24. Respects arrangements | -3.89 | 0.84 | 0.88 |
| 22. Is diligent | -4.58 | 1.03 | 1.16 |
| Average (standard deviation) | -2.70 (0.65) | 0.99 (0.22) | 1.01 (0.23) |

demand a high level of adaptability and handling of students' demands, however without missing the clear benchmarks for evaluating the process.

Given the one-dimensionality of the TAS discovered in the present sample — in contrast to theories positing higher dimensionality —, new studies with broader samples should be undertaken to confirm the factorial structure. Beyond this, studies investigating other evidence for the validity of the instrument based on external criteria, such as the judgment of students concerning the competencies of their teachers, are recommended. With this in mind, a version of the TAS for students is already being constructed. Researches comparing the practice of teachers of different subject areas who operate in both public and private sectors will be able to gather useful information on which to base changes in the training and professional operation of higher education teachers in Brazil.

We believe that the present study contributes to the advance of knowledge and to the evaluation of teaching practice in Brazil as it presents a specific and valid instrument for this end. It is hoped that its use can have an impact on four levels: individual, institutional, political, and theoretical-scientific. In individual terms, teachers will be able to evaluate their skills and competencies, and reflect on aspects to be improved in their work. At an institutional level, the HEI can use the results of the TAS, along with the results of other instruments, to implement strategic changes in the process of teaching. On the level of educational policy, it is hoped that this instrument can contribute to the SINAES, by helping to standardize the recording of information concerning teaching practice. Finally, from the theoretical-scientific point of view, the set of information produced by this and future researches could help rethink the theoretical definition of the good teacher.

REFERENCES

- AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION – AERA, APA, NCME. *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association, 1999. Disponível em: <<http://www.intestcom.org>>. Acesso em: 10 Jan. 2015.
- ANDRICH, D. A rating formulation for ordered response categories. *Psychometrika*, v. 43 n. 4, p. 561-573, dez. 1978.
- AVRICHIR, I.; DEWES, F. Construção e Validação de um Instrumento de Avaliação do Desempenho Docente. *Revista Eletrônica de Ciência Administrativa*, v. 5, n. 2, p. 1-16, jul./dez. 2006. Disponível em: <<http://revistas.facecla.com.br/index.php/recadm/index>>. Acesso em: 23 set. 2013.
- AZER, S. A. The qualities of a good teacher: how can they be acquired and sustained? *Journal of the Royal Society of Medicine*, v. 98, n. 2, p. 67-69, fev. 2005.
- BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 1977.
- COMREY, A. L.; LEE, H. B. *A first course in factor analysis*. 2. ed. Hillsdale: Erlbaum, 1992.
- DARLING-HAMMOND, L. *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*. Washington: Center for American Progress, out. 2008. Disponível em: <https://scale.stanford.edu/system/files/teacher_effectiveness.pdf>. Acesso em: 24 jan. 2015.
- DUARTE, F. P. Conceptions of good teaching by good teachers: case studies from an Australian university. *Journal of University Teaching & Learning Practice*, v. 10, n. 1, p. 3-14, jan./abr. 2013.
- FEITOZA, L. A.; CORNELSEN, J. M.; VALENTE, S. M. P. Representação do bom professor na perspectiva dos alunos de arquivologia. *Perspectivas em Ciência da Informação*, Belo Horizonte: Universidade Federal de Minas Gerais, v. 12, n. 2, p. 158-167, maio/ago. 2007.
- FONTANELLA, B. J. B.; RICAS, J.; TURATO, E. R. Amostragem por saturação em pesquisas qualitativas em saúde: contribuições teóricas. *Caderno de Saúde Pública*, v. 24, n. 1, p. 17-22, jan. 2008.
- GOMES, C. M. A.; BORGES, O. Limite da validade de um instrumento de avaliação docente. *Avaliação Psicológica*, v. 7, n. 3, p. 391-401, 2008.
- HAIR, J. F. et al. *Multivariate data analysis*. 7. ed. Upper Saddle River: Prentice Hall, 2009.
- HARDEN, R. M.; CROSBY, J. AMEE Guide N° 20: The good teacher is more than a lecturer - the twelve roles of the teacher. *Medical Teacher*, v. 22, n. 4, p. 334-347, 2000.
- HOLGADO-TELLO, F. et al. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, v. 44, n. 1, p. 153-166, jan. 2010.
- KORTHAGEN, F. A. J. In search of the essence of a good teacher: Towards a more holistic approach in teacher education. *Teaching and Teacher Education*, v. 20, n. 1, p. 77-97, jan. 2004.

- LIEFF, S. *et al.* Who am I? Key influences on the formation of academic identity within a faculty development program. *Medical Teacher*, v. 34, n. 3, p. 208-215, mar. 2012.
- LORENZO-SEVA, U.; TIMMERMAN, M. E.; KIERS, H. A. L. The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, v. 46, n. 2, p. 340-364, abr. 2011.
- MACCALLUM, R. C. *et al.* Sample Size in Factor Analysis. *Psychological Methods*, v. 4, n. 1, p. 84-89, jan./mar. 1999.
- MARQUES, M. C. O novo sistema de avaliação e acreditação do ensino superior: impacto, repercussões e benefícios. *Universo Contábil*, v. 6, n. 1, p. 142-155, jan./mar. 2010.
- MUTHÉN, B. O.; DU TOIT, S. H. C.; SPISIC, D. *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*, 1997. Disponível em: <http://www.statmodel.com/bmuthen/articles/Article_075.pdf>. Acesso em: 14 jan. 2015.
- OSMUN, W. E.; COPELAND, J. The occasional Teacher – Part 2: the good Teacher. *Canadian Journal of Rural Medicine*, v. 16, n. 1, p. 23-24, jan./mar. 2011.
- PACHANE, G. G. Quem é seu melhor professor universitário e por quê? Características do bom professor universitário sob o olhar de licenciandos. *Educação. Revista do Centro de Educação*, v. 37, n. 2, p. 307-319, maio/ago. 2012.
- PASQUALI, L. Princípios de Elaboração de Escalas Psicológicas. *Revista de Psiquiatria Clínica*, v. 25, n. 5, p. 206-213, set./out. 1998.
- _____. *Instrumentação psicológica*. Porto Alegre: Artmed, 2010.
- REED, C. What makes a good teacher? *BioScience*, California: University of California Press, v. 39, n. 8, p. 555-557, set. 1989.
- RISTOFF, D.; GIOLO, J. (Orgs.). *Educação superior brasileira: 1991-2004*. Brasília: MEC/INEP, 2006.
- SAMPLES, J. W.; COPELAND, S. E. The universality of good teaching: a study of descriptors across disciplines. *International Journal of Teaching and Learning in Higher Education*, v. 25, n. 2, p. 176-188, maio 2013.
- SHULMAN, L. S. Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, v. 57, n. 1, p. 1-22, fev. 1987.
- STONE, S. *et al.* Identifying oneself as a teacher: the perceptions of preceptors. *Medical Education*, v. 36, n. 2, p. 180-185, fev. 2002.
- SUPER, D. E.; SAVICKAS, M. L.; SUPER, C. M. The life-span, life-space approach to careers. In BROWN, D.; BROOKS, L. (Orgs.), *Career choice and development*. 3. ed., San Francisco: Jossey-Bass, p. 121-178, 1996.
- SUTKIN, G. *et al.* What makes a good clinical teacher in medicine? A review of the literature. *Academic Medicine*, v. 83, n. 5, p. 452-466, maio 2008.
- TSCHANNEN-MORAN, M.; HOY, A. W. Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, Amsterdam: Elsevier, v. 17, n. 7, p. 783-805, out. 2001.
- URBINA, S. *Fundamentos da Testagem Psicológica*. Porto Alegre: Artmed, 2007.
- ZEICHNER, K. Alternative paradigms of teacher education. *Journal of Teacher Education*, Washington, D.C.: Sage Publication, v. 34, n. 3, p. 3-9, maio 1983.

Appendix 1 – Teaching Assessment Scale (TAS) — Professor Version.

(Follow attached a free translation of the TAS — Professor Version. The only appropriate version of TAS for use is the original in Portuguese. To an adequate use of it in another language, we strongly suggest a process of cultural adaptation. For more information about it, please read Borsa, Damasio and Bandeira, 2012. For any use of TAS, please contact the first or the last author of this manuscript).

This is an instrument for reflecting on your teaching activity in higher education. With it, you will have the opportunity to identify and track your teaching competencies, and improve or develop them. In order for this process to be as effective as possible, be sincere in your responses and think carefully about your teaching work in one college-level discipline.

Discipline: _____

Respond to each item according to how well it applies to you using the 1 to 5 scale below:

| | | | | |
|------------------------|--------------|--------------|-----------|------------------------|
| 1 - Never / Not at all | 2 - A little | 3 - Somewhat | 4 - Often | 5 - Always/ Completely |
|------------------------|--------------|--------------|-----------|------------------------|

| | | | | | |
|--|---|---|---|---|---|
| 1. I am happy being a teacher | 1 | 2 | 3 | 4 | 5 |
| 2. I foster discussion relevant to the content | 1 | 2 | 3 | 4 | 5 |
| 3. I supplement content according to the interests of the students | 1 | 2 | 3 | 4 | 5 |
| 4. I promote students' curiosity for other sources of information | 1 | 2 | 3 | 4 | 5 |
| 5. I keep students' attention on the class | 1 | 2 | 3 | 4 | 5 |
| 6. I clarify, outside of class, students' doubts about the subject matter | 1 | 2 | 3 | 4 | 5 |
| 7. I permit moments of informality/humor in the classroom | 1 | 2 | 3 | 4 | 5 |
| 8. I speak so that all the students can hear me | 1 | 2 | 3 | 4 | 5 |
| 9. I relate linked ideas in my lecturing | 1 | 2 | 3 | 4 | 5 |
| 10. I am demanding of the students regarding the content being worked | 1 | 2 | 3 | 4 | 5 |
| 11. I have sufficient theoretical knowledge of the content being worked | 1 | 2 | 3 | 4 | 5 |
| 12. I am able to say "no" to students when necessary | 1 | 2 | 3 | 4 | 5 |
| 13. I give positive feedback to my students when it's pertinent | 1 | 2 | 3 | 4 | 5 |
| 14. I am punctual regarding the start and end of class | 1 | 2 | 3 | 4 | 5 |
| 15. I guide students' search for materials to complement the content worked in class | 1 | 2 | 3 | 4 | 5 |
| 16. I am flexible about adapting content according to the students' interests | 1 | 2 | 3 | 4 | 5 |
| 17. I promote critical appreciation of the content being taught | 1 | 2 | 3 | 4 | 5 |
| 18. I make use of diverse teaching methodologies in the classroom | 1 | 2 | 3 | 4 | 5 |
| 19. I am diligent | 1 | 2 | 3 | 4 | 5 |
| 20. I provide contemporary examples of the application of theories | 1 | 2 | 3 | 4 | 5 |
| 21. I respect arrangements made with students | 1 | 2 | 3 | 4 | 5 |
| 22. I make the epistemological bases for the content clear | 1 | 2 | 3 | 4 | 5 |
| 23. I like to encourage student autonomy | 1 | 2 | 3 | 4 | 5 |
| 24. I keep myself up-to-date on what I am teaching | 1 | 2 | 3 | 4 | 5 |
| 25. I give negative feedback to my students when it's necessary | 1 | 2 | 3 | 4 | 5 |

Appendix 1 – Continuation.

| | | | | | |
|--|---|---|---|---|---|
| 26. I adopt a critical stance regarding the content of the classes I present | 1 | 2 | 3 | 4 | 5 |
| 27. Whenever possible, I reconcile presented theory with professional practice | 1 | 2 | 3 | 4 | 5 |
| 28. I like being with my students | 1 | 2 | 3 | 4 | 5 |
| 29. I have mastery of several teaching methodologies | 1 | 2 | 3 | 4 | 5 |
| 30. I am organized about the tasks teaching demands of me | 1 | 2 | 3 | 4 | 5 |
| 31. I take the students' sociocultural diversity into account | 1 | 2 | 3 | 4 | 5 |
| 32. I ask for feedback from my students about the classes | 1 | 2 | 3 | 4 | 5 |
| 33. I feel capable of teaching students with different levels of learning | 1 | 2 | 3 | 4 | 5 |
| 34. I use diverse resources, when available, in class (e.g., audio-visual material, graphical material, computers, etc.) | 1 | 2 | 3 | 4 | 5 |
| 35. I promote students' participation in class | 1 | 2 | 3 | 4 | 5 |
| 36. I admit, in front of my students, that my knowledge has limits | 1 | 2 | 3 | 4 | 5 |
| 37. I am familiar with the practice of the content being worked | 1 | 2 | 3 | 4 | 5 |
| 38. I am open to dialog with my students | 1 | 2 | 3 | 4 | 5 |
| 39. I am familiar with the curricular organization of the course I am teaching | 1 | 2 | 3 | 4 | 5 |
| 40. I promote reconciliation between presented theory and evidence from scientific research | 1 | 2 | 3 | 4 | 5 |
| 41. I clarify students' doubt regarding my profession | 1 | 2 | 3 | 4 | 5 |
| 42. I verify whether the students are following my explanations or not during class | 1 | 2 | 3 | 4 | 5 |
| 43. I demonstrate mastery of the content when I speak | 1 | 2 | 3 | 4 | 5 |
| 44. I handle in-class conflicts appropriately | 1 | 2 | 3 | 4 | 5 |
| 45. I am clear about evaluation criteria | 1 | 2 | 3 | 4 | 5 |
| 46. I help students in class in clarifying their doubts | 1 | 2 | 3 | 4 | 5 |

ABOUT THE AUTHORS

MÔNIA APARECIDA DA SILVA has a doctorate in psychology from the Universidade Federal do Rio Grande do Sul (UFRGS).
E-mail: moniapsi@yahoo.com.br

WAGNER DE LARA MACHADO has a doctorate in psychology from the Universidade Federal do Rio Grande do Sul (UFRGS). He is a professor at the Pontifícia Universidade Católica de Campinas (PUC-Campinas).
E-mail: wag_psico@yahoo.com.br

LUCIANE MARIA PILOTTO has a doctorate in dentistry from the Universidade Federal do Rio Grande do Sul (UFRGS). She is a professor at the Centro Universitário Univates (UNIVATES).
E-mail: lutipilotto@yahoo.com.br

BÁRBARA BACKES has a doctorate in psychology from the Universidade Federal do Rio Grande do Sul (UFRGS).
E-mail: barbara.edas@gmail.com

REGINA BASSO ZANON has a doctorate in psychology from the Universidade Federal do Rio Grande do Sul (UFRGS). She is a professor at the Faculdade Três de Maio (SETREM).
E-mail: rebzanon@gmail.com

PATRÍCIA VASCONCELOS MACHADO has a master in psychology at the Universidade Federal do Rio Grande do Sul (UFRGS).
E-mail: patricia-machado@igp.rs.gov.br

ANA PAULA COUTO ZOLTOWSKI has a doctorate in psychology from the Universidade Federal do Rio Grande do Sul (UFRGS). She is a professor at the Instituto Brasileiro de Gestão de Negócios (IBGEN).
E-mail: ana_zoltowski@yahoo.com.br

REBECA VERAS DE ANDRADE VIEIRA is a doctoral candidate in psychology at the Universidade Federal do Rio Grande do Sul (UFRGS).
E-mail: rebecavieirapsico@gmail.com

RENATA GIULIANI ENDRES is a doctoral candidate in psychology at the Universidade Federal do Rio Grande do Sul (UFRGS).
E-mail: rgendres01@yahoo.com.br

MARTINA FRANCALANCI is a graduate in psychology from the Università di Bologna (Italy).
E-mail: martina.francalanci@studio.unibo.it

VICTOR ADRIEL DE JESUS OLIVEIRA is a doctoral candidate in computer science at the Universidade Federal do Rio Grande do Sul (UFRGS).
E-mail: victorajoliveira@gmail.com

JEFFERSON SILVA KRUG has a doctorate in psychology from the Universidade Federal do Rio Grande do Sul (UFRGS). He is a professor at the Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS).

E-mail: jefferson.krug@pucrs.br

DENISE RUSCHEL BANDEIRA has a doctorate in psychology from the Universidade Federal do Rio Grande do Sul (UFRGS). She is a professor at the same institution.

E-mail: deniserbandeira@gmail.com

Received in March 26, 2015

Approved in October 20, 2015