

# Validation of an evaluation instrument for the flipped classroom active methodology\*<sup>1</sup>

Dionara Guarda<sup>2</sup>

ORCID: 0000-0003-3933-9717

Graciela Cabreira Gehlen<sup>3</sup>

ORCID: 0000-0002-4345-4489

Gimene Cardozo Braga<sup>4</sup>

ORCID: 0000-0002-5102-6505

Albimara Hey<sup>4</sup>

ORCID: 0000-0003-1836-8252

## Abstract

Active methodologies are understood as ways to develop the learning process and construct critical training with future professionals in the most diverse areas, favoring the autonomy of the student in their learning, encouraging individual and collective decision-making. The objective of this study, as well as its methodological path, seeks to validate the evaluation instrument used in the Flipped Classroom Active Methodology. Developed by teachers of the nursing course and based on bloom taxonomy of the Cognitive Domain, the instrument contains twenty items to be evaluated. For validation, the psychometric properties of validity and reliability related to Pasquali's psychometry were used. Verified through the Content Validity Index (CVI), it measures the proportion of evaluators who are in agreement with the instrument's items, using an agreement of at least 80% among the evaluators as a decision criterion on the pertinence and acceptance of the item. The instrument was evaluated by 8 participants considered experienced in the area, using a 4-point Likert scale. The reliability criterion, analyzed using the IBM SPSS® software, presented a Cronbach's alpha value of 0.94 and the validity criterion reached 70% agreement between the evaluators. It is concluded that the instrument is reliable, however, regarding the validity criterion, it needs adjustments to represent what it intends.

## Keywords

Active Methodologies - Flipped Classroom - Study of validation - Education in nursing.

\* English version by Elaine Rouine. The authors take full responsibility for the translation of the text, including titles of books/articles and the quotations originally published in Portuguese.

**1-** The entire data set that supports the results of this study was published in the article itself.

**2-** Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil. Contato: dionaraguarda@gmail.com

**3-** Instituto Federal do Paraná – campus Palmas, Pato Branco, PR, Brazil. Contact: graciela.gehlen@ifpr.edu.br

**4-** Instituto Federal do Paraná – campus Palmas, Palmas, PR, Brazil. Contacts: gimene.braga@ifpr.edu.br; albimara.hey@ifpr.edu.br



<https://doi.org/10.1590/S1678-4634202349248000eng>

This content is licensed under a Creative Commons attribution-type BY 4.0.

## Introduction

Active methodologies are understood as ways to develop the process of learning and develop critical training with future professionals in various areas. The use of these methodologies can favor the autonomy of the student, arouse curiosity and encourage individual and collective decision-making, bringing the student's context closer to social practices (BORGES, T. ; ALENCAR, 2014). In addition, the use of active methodologies is capable of transforming learning into a multiplicative tool for change (ROMAN *et al.*, 2017).

Active teaching methodologies promote the student's participation and commitment to their learning, encouraging them in the teaching-learning processes towards a critical-reflective position. Such methodologies require teaching with skills training, which promotes the student's critical approach to reality, reflection on problems and integration with the health service and the human being (BELLAYER, 2019).

The use of active methodologies in healthcare educational courses is an important pedagogical current to be addressed, as it supports meaningful learning and encourages students to seek the knowledge necessary for clinical practice for future use (MELLO; ALVES; LEMOS, 2014).

Over time, health education in higher education has undergone positive changes, aiming at the implementation of teaching strategies that promote the student to provider and holder of their own teaching-learning process (MELLO; ALVES; LEMOS, 2014).

Resolution CNE/CES 3/2001, which establishes the National Curriculum Guidelines for the Undergraduate Nursing Course, guides the formation of a generalist, humanist, critical-reflective professional; qualified to exercise the profession based on scientific and intellectual rigor, and based on ethical principles; among other capabilities, the nurse must be able to solve health problems, both individually and collectively; to make appropriate decisions; to the exercise of leadership, administration and management (BRASIL, 2001). Thus, it is understood that the active methodologies corroborate the formation of this professional profile.

In active methodologies teaching, the student is seen as the protagonist of his learning, leaving the teacher to mediate, which should arouse epistemological curiosity. Regarding the nursing course, this methodology has positive impacts, as it enables the future nurse to face the real challenges of the profession. In this regard, the NCGs of undergraduate nursing courses have contributed to the formation of students who are able and capable of solving problems around them and, above all, encouraging them to be citizens concerned with others (WEBER, 2018).

The nursing course at the Federal Institute of Paraná (IFPR) – Palmas campus, uses active teaching-learning methodologies. The Flipped Classroom, as one of the methodologies chosen by the course, allows students to build knowledge in a dynamic and autonomous way, since it provides for a previous study, to be carried out at home, and a space for discussion in the classroom. for the synthesis of knowledge and experience on the proposed subject. For this methodology, scientific articles are made available for individual prior reading and, subsequently, the discussion and contextualization of the same are carried out, in a group, with the mediation and guidance of the teacher,

who participates in the discussion and asks questions, exploring the knowledge and contributions of the students.

The Flipped Classroom is a proposal that demands more from the actors involved in the learning process, demands from the student autonomy and responsibility for their learning and, in addition, demands action and reflection in the classroom. As for the teacher, it requires planning, accuracy in exposing the content, reflection and self-criticism about the content, and joint analysis with the students. Students are no longer mere spectators but contribute as co-builders of the content that is being produced. Despite the advantages of this method, just inverting the classroom does not mean that a reflective and critical methodology is being applied, or that an autonomous student is being produced. For this methodology to work, changes in the pedagogical structure, planning and in the way teachers and students work are necessary (BRANCO; ALVES, 2015).

The Flipped Classroom methodology was first developed in the United States, by professors Jonathan Bergmann and Aaron Sams, as an alternative teaching didactic that presented better results. The proposal highlights the need for a change in the teaching role, which stops transmitting concepts to assume guidance/tutoring functions. In general, the class starts to revolve around the student, not the teacher anymore (LOSTADA, 2017).

Paiva et al. (2016) found benefits in active teaching-learning methodologies, in breaking with the traditional model, in changes present in teamwork, in theory-practice integration and in the development of student autonomy and a critical view of reality, favoring the evaluation formative.

This methodology presupposes that the act of evaluating must be an integral part of the entire teaching-learning process, it must be continuous and include all information produced by the interaction between teachers and students, as well as those produced among students. Peer evaluation is relevant for the learning assessment and for any necessary adjustments, so that the student can achieve the individually and collectively defined objectives (BORGES, M. *et al.*, 2014).

The Teaching-Learning Active Methodology requires consistent evaluation with the way of teaching, considering the way of producing knowledge and the contribution of each student to their learning and to the group. Considering the importance of this study for the nursing course, the validity and reliability of the evaluation instrument of the Flipped Classroom Active Methodology are questioned. Thus, the present study aims to validate the assessment instrument used in the Active Methodology of the Flipped Classroom in the nursing course at IFPR, Palmas campus.

## **Methodological route**

This is a validation study of the evaluation instrument that has already been used in the Active Methodology of the Flipped Classroom in the nursing course at the Federal Institute of Paraná - Palmas campus. The instrument was evaluated by experts in active methodologies to determine whether it is valid and reliable. The psychometric validity and reliability validation took place from June to December 2020.

It is emphasized that the evaluation instrument submitted to validation was developed by a group of teachers of the nursing course of the Federal Institute of Paraná - Palmas campus and, despite the experience of the professors, lacks analysis.

The instrument contains five (5) evaluative criteria, each with four (4) items related to concepts A, B, C and D, totaling twenty (20) items requiring evaluation.

## **Population, inclusion and exclusion criteria**

The inclusion criteria for participants in the study include: having a PhD or master's degree in the health area; professional experience (clinical, research or teaching) and article published in an indexed journal in the area of interest of the study. To assess the representativeness and scope of the instrument's items, a formal invitation was made via e-mail to medical and nursing courses at universities that use active teaching-learning methodologies, but a response was obtained from only 2 contacts interested in participating in the study. As there was difficulty in recruiting participants with this approach, the annals of a national event on active methodologies were searched for the e-mail contact of health professionals who met the inclusion criteria of the study. 21 professionals were invited via e-mail, of which only 7 responded, accepting to participate. However, when sending the questionnaire, one contact withdrew, leaving 6 to be included in the study. The objective was to reach at least 5 research participants, but 8 participants were guaranteed to evaluate the instrument. As for these 8 participants, there were no problems with delay in returning the completed questionnaire.

## **Ethical aspects**

The development of the study complied with the mandatory ethical standards for research with human beings of Resolution 466/12 of the National Health Council - CNS (BRASIL, 2013), with approval of the Research Ethics Committee of the Federal Institute of Paraná. All individuals involved in the research were informed about the objectives and procedures, and received, read and signed the Informed Consent Form (TCLE).

It is understood that participation in the research could cause risks such as tiredness in answering the questionnaire with twenty (20) items and discomfort in terms of opinion exposure. However, to minimize these risks, the participation occurred via e-mail, providing the participant with the necessary time to answer the questionnaire, as well as the guarantee of anonymity guaranteed with the TCLE and the possibility of not participating in the study at any stage of the study.

The questionnaires with the participants' answers will remain stored for five (5) years from the end of the study.

## **Instrument submitted to validation**

The instrument for evaluating the Flipped Classroom methodology was developed based on the taxonomy of Bloom *et al.* (1956).

According to Ferraz and Belhot (2010), Bloom was interested in providing a practical and useful tool that was consistent with the characteristics of higher mental processes (level of knowledge and complex abstraction) in the way they were considered and known.

Bloom's Taxonomy of the Cognitive Domain is structured in levels of increasing complexity, that is, to acquire a new skill belonging to the next level, the student must have mastered and acquired the skill of the previous level. The categories of the cognitive domain of Bloom's Taxonomy include: 1. Knowledge, 2. Understanding, 3. Application, 4. Analysis, 5. Synthesis and 6. Evaluation (FERRAZ; BELHOT, 2010).

Many educators rely on the theoretical assumptions of this cognitive domain to define objectives, strategies and evaluation systems in their educational plans. Two advantages of using the taxonomy in the educational context are the fact that it offers the basis for the development of assessment instruments and the use of differentiated strategies to facilitate, assess and encourage student performance at different levels of knowledge acquisition. In addition to encouraging educators to help their students, in a structured and conscious way, to acquire specific skills based on the perception of the need to master simpler skills (facts) in order to later master more complex ones (concepts) (FERRAZ; BELHOT, 2010).

The taxonomy brought the possibility of language standardization in the academic environment and, with that, also new discussions around subjects related to the definition of instructional objectives. In this context, learning instruments could be worked on in a more integrated and structured way, including considering technological advances that could provide new and different tools to facilitate the teaching and learning process (FERRAZ; BELHOT, 2010).

Conceived within the context and perspective of being a mediator in the progress of learning, the referred instrument for evaluation was developed, understanding, through the conception of Sirley L. Freitas, Michele G. N. Da Costa and Flavine A. De Miranda (2014), that the assessment can be perceived as a process, which is not limited to the application of daily tests, but rather a joint action in which the teacher monitors the student in relation to his performance, development and appropriation of knowledge.

## **Instrument validation**

In order to determine the quality of the instrument, psychometric properties called validity and reliability were used

Reliability is the ability of an instrument to faithfully measure a phenomenon. Validity is the ability of an instrument to accurately measure the phenomenon being studied (PILATTI; PEDROSO; GUTIERREZ, 2010).

Psychometry consists of measuring the behavior of the organism through mental processes (law of comparative judgment). It seeks to explain the meaning of the answers given by the subjects to a series of items (PASQUALI, 2009).

The items were submitted to the evaluation of the participants and later the responses were analyzed for validity and reliability.

The evaluation of the instrument was carried out through the Likert scale (1932), of four (4) points, to investigate the clarity and relevance of each item, enabling the individual analysis of content by the evaluators.

The instrument was evaluated using the Likert scale (1932), with four (4) points, to investigate the clarity and pertinence of each item, allowing the individual analysis of content by the evaluators.

The Likert scale usually presents three or more points, where the research respondent says if he agrees, if he is in doubt or if he disagrees with what is stated in the item in relation to the ability to measure what the instrument proposes (PASQUALI, 1996).

Content validation was used to demonstrate the instrument validity (PASQUALI, 2009).

Clarity/pertinence content validation was performed quantitatively using the Content Validity Index (CVI) using a 4-point Likert-type scale, 1 - Disagree, 2 - Indifferent, 3 - Partially Agree and 4 - Totally Agree. The percentage of evaluators who totally and partially agree with the item of the instrument was calculated from the responses, obtained by adding the answers “3” and “4” of each evaluator in each item of the instrument, and dividing this sum to the number total responses, resulting in the proportion of evaluators who judge the item valid (MEDEIROS *et al.*, 2015). An agreement of at least 80% between the evaluators will serve as a decision criterion on the relevance and/or acceptance of the item to which it theoretically refers (PASQUALI, 2010).

Clarity and pertinence seek to evaluate whether the concept adequately expresses what is expected to be measured and achieves the proposed objectives (ALEXANDRE; COLUCCI, 2011).

Reliability, which assesses whether an instrument is always capable of measuring what it is intended to measure in the same way, was studied using Cronbach’s alpha coefficient, which corresponds to the measure of the internal consistency of a scale. Calculating its coefficient only requires administering a single test to provide an estimate of the reliability of the entire survey (VELOSO; SHIMODA; SHIMOYA, 2015).

The Cronbach’s alpha coefficient presented by Lee J. Cronbach, in 1951, estimates the reliability of a questionnaire applied in a survey, through the average correlation between questions, that is, measuring the correlation between answers of a questionnaire through the analysis of the profile of the answers given by the respondents. Given that all items in a questionnaire use the same measurement scale, the  $\alpha$  coefficient is calculated from the variance of individual items and the variance of the sum of items for each rater using the following equation:

$$\alpha = \frac{k}{k - 1} \left[ 1 - \frac{\sum S^2 i}{S^2 t} \right]$$

where: K corresponds to the number of items in the questionnaire;  $S^2 i$  corresponds to the variance of each item;  $S^2 t$  corresponds to the total variance of the questionnaire, determined as the sum of all variances (HORA; MONTEIRO; ARICA, 2010).

For André L. P. Freitas and Sidilene G. Rodrigues (2005), the reliability classification of Cronbach's alpha coefficient with values  $< 0.60$  is considered low, from  $0.60$  to  $\leq 0.75$  considered moderate, from  $0.75$  to  $\leq 0.90$  considered high and  $> 0.90$  considered very high.

## Results and discussion

The evaluation instrument was submitted to the evaluation of 8 participants, who analyzed the instrument for content validity. According to Moura *et al.* (2008), 6 evaluators are enough to perform the task. Participants are health professionals, all of whom work in teaching, with 3 doctors in nursing, 2 masters in nursing, 1 doctor in medicine, 1 doctoral student in medicine and 1 master in psychology. Table 1 presents the instrument submitted to the evaluation of the participants.

**Table one-** Instrument for evaluating the active methodology flipped classroom of the nursing course of the IFPR *campus* Palmas

Evaluative criteria	Evaluative items related to concepts	Evaluation
1. Attendance and Participation	1. Complete (A) - Actively participates in all meetings and moments of the discussion of texts, collaborating in the collective construction of knowledge.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	2. Partially complete (B) - Actively participates in some of the moments of the discussion of the texts, collaborating in the collective construction of knowledge.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	3. Sufficient (C) - Participates in the moments of the discussion of the texts, partially collaborating in the collective construction of knowledge.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	4. Insufficient (D) - Rarely participates in the moments of the discussion of the text, not contributing to the collective construction of knowledge.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
2. Communication	5. Complete (A) - Discusses consistently by making a relationship with the studied text, demonstrating verbal ability (clarity, coherence, spontaneity) to express their thoughts.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	6. Partially Complete (B) - Discusses consistently by making a relation to the studied text, demonstrating little verbal ability (clarity, coherence, spontaneity) to express their thoughts.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	7. Sufficient (C) - Discusses consistently in relation to the studied text, demonstrating no verbal ability (clarity, coherence, spontaneity) to express their thoughts	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Concorof the Partially ( ) 4 - I totally agree
	8. Insufficient (D) - Discusses the text without coherence using empirical examples and without verbal ability	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree

<b>3. Relationship of Synthesis and Nursing Knowledge and Practices</b>	9. Complete (A) - Relates synthesis with nursing knowledge and practice, citing studies and experiences that exemplify this relationship.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	10. Partially Complete (B) - Partially relates the synthesis with nursing knowledge and/or practice, citing studies and experiences that exemplify this relationship.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	11. Sufficient (C) - Presents difficulties in relating summary to nursing knowledge and/or practice.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	12. Insufficient (D) - Does not relate synthesis to nursing knowledge and/or practice	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
<b>4. Group work</b>	13. Complete (A) - Demonstrates the ability to work in groups, critical of colleagues and the text, contributing to the learning teaching process.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	14. Partially Complete (B) - Demonstrates the ability to work in groups, critical of colleagues and the text, partially contributing to the learning teaching process.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially known ( ) 4 - I totally agree
	15. Sufficient (C) - Partially demonstrates the ability to work in groups, critical of colleagues and the text, contributing insufficiently to the learning teaching process.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	16. Insufficient (D) - Demonstrates unsatisfactory group work capacity, critical of colleagues and not, contributing to the learning teaching process.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
<b>5. Synthesis</b>	17. Complete (A) - Performs the synthesis of the text with logical reasoning, proposing the way of thinking about the practice with scientific domain.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	18. Partially Complete (B) - Synthesizes the text with logical reasoning, partially proposing alternative ways of thinking about practice with scientific domain.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	19. Sufficient (C) - Partially synthesizes the text with logical reasoning, proposing or not alternative ways of thinking about practice with scientific domain.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree
	20. Insufficient (D) - Proposes empirical alternatives and does not synthesize the text.	( ) 1 - I do not agree ( ) 2 - Indifferent ( ) 3 - Partially agree ( ) 4 - I totally agree

Source: PPC from the IFPR-Palmas nursing course (IFPR, 2019).



Table 1 illustrates the participants' responses according to their agreement with each item of the Flipped Classroom Active Methodology assessment instrument, represented by a 4-point Likert scale, 1 – Disagree, 2 – Indifferent, 3 – Partially Agree and 4 – I totally agree.

**Table 1-** Evaluation of the instrument by participants, through the 4-point Likert scale

		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Participants	A	3	4	4	1	3	3	1	1	4	3
	B	4	1	4	4	4	4	1	4	4	4
	C	4	4	4	4	4	3	1	4	3	4
	D	1	1	1	1	1	1	1	1	1	1
	E	4	4	3	4	4	3	3	4	3	3
	F	3	3	3	3	3	3	3	3	4	3
	G	3	4	4	4	4	3	1	1	3	4
	H	3	3	4	3	4	4	1	4	4	4
		Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
Participants	A	1	1	3	3	1	1	4	3	1	1
	B	1	4	4	4	1	4	4	1	4	4
	C	4	4	3	3	1	1	3	3	1	4
	D	1	1	1	1	1	1	1	1	1	1
	E	3	4	4	3	3	4	4	3	3	4
	F	4	4	3	3	3	4	4	3	4	4
	G	3	1	4	3	3	1	4	4	3	1
	H	1	1	3	3	3	1	4	4	4	4

Source: Own development.

After the participants returned the questionnaire, categorized from A to H, the instrument was analyzed for its validity and reliability. Reliability and validity are two closely related measurement properties that play complementary roles. Reliability, measured through internal consistency, concerns the homogeneity of responses from different evaluators, while validity represents the degree of certainty about the measured concept (DE BEM *et al.*, 2011).

The reliability criterion was verified using the IBM SPSS® software, which presented a 0.94 value for the Cronbach's alpha coefficient, demonstrating that the present instrument is reliable, because, according to André L. P. Freitas and Sidilene G. Rodrigues (2005), values above 0.90 are considered highly reliable.

Content validity is the result of the evaluation of specialists in the study area, who analyze the representativeness of the items in relation to the concepts that are intended to

be measured (BITTENCOURT *et al.*, 2011). For this criterion, the CVI was used, which, when applied to the entire instrument, obtained an agreement of 70% among the evaluators, however, when analyzing each item separately, varying agreements were obtained. For 9 of the 20 items the rate of agreement between raters was greater than 80%, for 3 items this rate was 75% and for the remaining 8 items the rate of agreement was less than 70%. Table 2 shows the agreement rate for each of the items in the Flipped Classroom Active Methodology assessment instrument.

**Table 2-** Agreement rate of the Content Validity Index (CVI) on the items of the instrument

	Agreement Rate		Agreement Rate
Item 1	87%	Item 11	50%
Item 2	75%	Item 12	50%
Item 3	87%	Item 13	87%
Item 4	75%	Item 14	87%
Item 5	87%	Item 15	50%
Item 6	87%	Item 16	37%
Item 7	25%	Item 17	87%
Item 8	62%	Item 18	75%
Item 9	87%	Item 19	62%
Item 10	87%	Item 20	62%

Source: Our own development.

Regarding items with agreement higher than 80% (items 1, 3, 5, 6, 9, 10, 13, 14 and 17), they achieved an agreement rate of 87% and, although suggestions were made to improve the evaluation instrument, the items proved to be adequate for what is proposed. The best evaluation of these items was generally associated with the clarity in the exposure and compatibility between the evaluative concept and the actions expected to be achieved by each student within each trial.

However, most items showed an agreement rate of less than 80%. Some evaluators justified the non-agreement and suggested changes, mainly referring to the description of the item, in order to reduce the subjectivities in the evaluation, both for the evaluator professor and for the student.

Since subjectivity is inevitably present in the evaluation processes, whether due to the choice of items, the way they are presented or the language used, it becomes necessary to clearly define the evaluation criteria to minimize it, when it is impossible to eliminate it (PARANÁ, 2013).

Although there is a specific orientation for each attribution of concepts, the minimum quality references are substantiated by subjective attributes - adequate, adequately, sufficient, insufficient -, requiring the evaluator to be more responsible for ensuring consistency and fairness in the judgment (ANDRADE, 2014).

Next, the criteria and items of the assessment instrument will be described and examined, based on the judgment and comments made by the participants.

## **Analysis of the evaluative criteria of the evaluation instrument**

The study participants made notes regarding the criteria, notably regarding the criterion “Attendance and participation”, in which it was recommended to dissociate “attendance” from “participation”, given that they are different evaluation criteria, and the student can be fully present, but with insufficient participation. Another suggestion was to change the order of the criteria, given that to assess the “Relationship of the synthesis with nursing knowledge and practices” it is imperative to come before the “Synthesis”, arranging them as follows: 1. Attendance and participation; 2. Communication; 3. Group work; 4. Synthesis; 5. Relation of the synthesis with nursing knowledge and practices.

When examining the evaluation of the participants by criteria, it was observed that some were better evaluated than others. The criterion “Attendance and participation” obtained the best evaluation of the instrument, while the criteria “Communication” and “Group work” registered the worst evaluations.

Additionally, regarding the criterion “Attendance and participation”, it was commented by an evaluator that student engagement depends on basic education training and access to tools, such as the internet and books, and the availability of extra-class time, factors that interfere with engagement as a whole and in the collective construction of knowledge.

## **Analysis of the evaluative items of the evaluation instrument**

Next, the items whose agreement rates by the participants were lower than 80% will be described and examined.

Items 2 and 4, referring to concepts B and D of the criterion “Attendance and participation”, reached 75% agreement. Item 2 describes: “he actively participates in some of the moments of discussion of the texts, collaborating in the collective construction of knowledge”, and item 4 says: “he rarely participates in the moments of discussion of the text, not contributing to the collective construction of knowledge”. There were no comments for these items.

Item 7 had the lowest concordance rate in the study, with only 25%. This item refers to the Sufficient (C) concept of the “Communication” criterion and states: “Discusses coherently, relating to the text studied, demonstrating no verbal ability (clarity, coherence, spontaneity) to express their thoughts”. The low evaluation of the item is due to the mistaken description, when it states “no skill”, which better expresses concept D, as it portrays failure in communication. In the same sense, item 8, with an agreement rate of 62%, needs adaptations to clearly represent what is being evaluated. This item addresses concept D

of the “Communication” criterion and states: “Discusses the text without coherence using empirical examples and without verbal skills”. While the other items of the “Communication” criterion took into account clarity, coherence and spontaneity, this item did not detail these factors, limiting itself only to the lack of coherence and verbal ability.

It was recommended to adapt the description of all items of this criterion, pointing out which elements are being evaluated and how many of them the student reached in each concept. For example, concept A encompasses the scope of the three elements – clarity, coherence and spontaneity -, concept B refers to the scope of two of the three, concept C only one and, finally, concept D none of the elements. Still, the need to define the understanding of what coherence is was highlighted.

Items 11 and 12, which address, in this order, “Difficulties relating the synthesis to nursing knowledge and/or practice” and “Does not relate the synthesis to nursing knowledge and/or practice”, associated with concepts C and D of the evaluative criterion “Relationship of the synthesis with nursing knowledge and practices”, reached 50% agreement. Again, subjectivity was the main obstacle for greater agreement. For example, for the concept Sufficient (C) the term “difficulties” was used and the concept Partially full (B) contains the term “partially”, which are expressions that are confused because they are within the same criterion. An alternative proposed in this sense was to quantify the studies cited by academics to assess which concept they discuss: cites 5 or more studies to exemplify the correlation (knowledge and nursing practices) and complements with experiences – concept A; cites 3 to 4 – Concept B; 1 to 2 – grade C; does not cite any study, only experiences - concept D. Another suggestion was to change the word “synthesis” to “synthesis of the text studied”, to make it clear to everyone who uses the instrument what it is about.

Items 15 and 16 obtained, respectively, 50% and 37% agreement. The items are consistent with “Partially demonstrates ability to work in groups, criticizing colleagues and the text, contributing insufficiently to the teaching-learning process” of concept C and “Demonstrates unsatisfactory ability to work in groups, criticizing colleagues and not contributing to the teaching-learning process” of concept D, both situated in the evaluative criterion “Group work”. The low agreement with the items is due to the subjective description, as well as the lack of definition in the presentation. In item 15, the statement “contributes insufficiently” does not fit the concept Sufficient (C) and in item 16, which explains the concept Insufficient (D), the following description was proposed by a participant: “Does not demonstrate ability to work in group, as it does not make constructive criticisms of colleagues and the text and does not contribute positively to the teaching-learning process”, thus elucidating the evaluation.

With regard to the word criticism, it was suggested to change it to constructive criticism or valid criticism, not considering harsh or out-of-context comments for the evaluation. Also, the term in question can be replaced by feedback. This suggestion is valid for all items in this criterion.

Regarding the “Group work” criterion, one participant highlighted its importance in the relationship between colleagues and, beyond the classroom, in the labour market.

Communication and interpersonal relationships are part of the ability to interact, live with and contact people through empathic relationships, being important in terms of productivity and the quality of the work provided by the nursing team, which demands a training process consistent with skills and skills to manage existing conflicts (BOLONHEIZ; BALDISSERA; OLIVEIRA, 2011).

In item 18, which displays “It synthesizes the text with reason, partially proposing alternative ways of thinking about the practice with scientific domínio” for the concept B of the critério “Síntese”, obtained a rate of 75% agreement. The main observation refers to the lack of objectivity in the exdisplay of the “partially”, and there is a need to detail or exemplify the meaning of the term.

In item 18, which displays “Performs the synthesis of the text with logical reasoning, partially proposing alternative ways of thinking about the practice with scientific domain” for concept B of the criterion “Synthesis”, it obtained a 75% agreement rate. The main observation refers to the lack of objectivity in the exposition of “partially”, with the need to detail or exemplify the meaning of the term.

Items 19 and 20, also of the “Synthesis” criterion, obtained 62% agreement and the main notes for these items refer to the lack of clarity of what is intended to be evaluated and to the subjectivity in the definition. The criticisms turned especially to item 19, consistent with the concept Sufficient (C), which describes “It partially performs the synthesis of the text with logical reasoning, proposing or not alternative ways of thinking about the practice with scientific domain” and causes confusion when brings “proposing or not”, because it is understood that what is expected to be achieved must be tangible for the student and for the evaluator. The excerpt “proposing or not” allows students who propose alternative ways of thinking about the practice of nursing with scientific domain to claim a B or even A concept. It is also worth noting that “proposing or not” is confused with partial, which turns the item description similar to the previous one.

The validity of the measure depends on the adequacy of the instrument in relation to what one wants to measure. Therefore, the validity of a measure is never absolute, but always relative, that is, a measuring instrument is not simply valid, but valid for a given purpose (MARTINS, 2006).

It is considered, therefore, that the evaluation instrument of the active methodology of flipped classroom of the nursing course at IFPR, Palmas campus, cannot be considered valid without alterations, since it obtained a 70% agreement rate among the evaluators, lower than the threshold value to validate without changes which is 80%.

According to Martins (2006), a measurement instrument can be reliable (presents reliability) and not necessarily be valid, that is, it can present consistency in the results it produces, but not measure what it intends to. However, to represent reality, a measuring instrument must be reliable and valid.

It is recommended to carry out the changes suggested by the evaluators and submit the instrument to a new validation process, guaranteeing the validity of the instrument to be used with confidence in the nursing course at IFPR - Palmas campus, and also so that it can be a reference instrument for other teaching institutions that work with active teaching-learning methodologies.

## Conclusion

This study submitted the evaluation instrument of the Active Methodology of Flipped Classroom of the nursing course at IFPR – Palmas campus for validation, which, according to the psychometric properties studied, presented high reliability, with a value for Cronbach's alpha coefficient of 0.94. However, regarding the validity criterion, the instrument was unable to be validated, as it reached a concordance rate of only 70% among the evaluators, thus requiring adjustments to represent what it intended. In order for the instrument to be validated, it needs to be reliable and valid.

Regarding the instrument items that were adequate for the proposed objective, achieving an agreement rate of 87% among the evaluators, such performance was attributed to its objective definition and the conformity between the evaluative concepts and the description of the items.

In the analysis of items with a concordance rate below 80%, it was noted that the main weaknesses of the instrument pointed out by the evaluators are the lack of objectivity and clarity in their descriptions. These items received suggestions for adjustments to reduce or eliminate the subjectivities that interfere in the evaluation.

The participation of the evaluators in the validation of the instrument was fundamental for the research, because, in addition to the evaluation, recommendations were made to improve the instrument, allowing a clearer and more objective description of the items, in order to provide a correct and fair evaluation of the Active Methodology of Flipped classroom.

As nursing students of this active teaching-learning method, we consider that the validation of the assessment instrument is fundamental for the transparency of the teaching-learning process, and for a good relationship between students and professors. Once the criteria and items of the assessment instrument have been clearly and objectively established, teachers and students are allowed to feel assured by the adopted assessment instrument, as it acquires the ability to faithfully elucidate the student's level of development, enabling reflections and changes in attitudes, when necessary, of all those involved in the process, in order to achieve the desired learning together.

## References

ALEXANDRE, Neusa Maria Costa; COLUCI, Marina Zambon Orpinelli. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. **Ciência & Saúde Coletiva**, Campinas, v. 16, n. 7, p. 3061-3068, 2011. <https://doi.org/10.1590/S1413-81232011000800006>

ANDRADE, Maria Antonia Brandão de. **As contribuições da avaliação dos cursos de graduação para a melhoria da qualidade da educação superior**. 2014. Dissertação (Mestrado em Educação) – Universidade Federal da Bahia, Salvador, 2014. Disponível em: <http://repositorio.ufba.br/ri/handle/ri/9019>. Acesso em: 05 nov. 2020.

BELLAVÉR, Emyr Hiago. **Ferramentas para avaliação em metodologias ativas**. Caçador, Uniarp, 2019. Disponível em: <https://uniarp.edu.br/wp-content/uploads/2021/07/E-Book-Free-Access-Ferramentas-de-avaliacao-de-metodologias-ativas-Prof.-Ms.-Emyr-Hiago-Bellaver.pdf>. Acesso em: 07 jun. 2020.

BITTENCOURT, Hélio Radke *et al.* Desenvolvimento e validação de um instrumento para avaliação de disciplinas na educação superior. **Estudos em Avaliação Educacional**, São Paulo, v. 22, n. 48, p. 91-114, jan./abr. 2011. Disponível em: <http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/1630/1630.pdf>. Acesso em: 03 jun. 2020.

BLOOM, Benjamin Samuel *et al.* **Taxonomy of educational objectives**. v. 1. New York: David McKay, 1956.

BOLONHEIZ, Lorena Mattos Galhardi Carvalho; BALDISSERA, Vanessa Denardi Antoniassi; OLIVEIRA, Raquel Gusmão de. Comunicação e relacionamento interpessoal na enfermagem: elaboração e execução de uma proposta educativa junto aos graduandos em enfermagem por meio da dialogicidade. *In*: ENCONTRO INTERNACIONAL DE PRODUÇÃO CIENTÍFICA, 6. 2011, Maringá. VII EPCC. Maringá: [s. n.], 2011. Disponível em: [http://rdu.unicesumar.edu.br/bitstream/123456789/4795/1/lorena\\_mattos\\_galhardi\\_carvalho\\_bolonheiz.pdf](http://rdu.unicesumar.edu.br/bitstream/123456789/4795/1/lorena_mattos_galhardi_carvalho_bolonheiz.pdf). Acesso em: 21 out. 2020.

BORGES, Marcos C. *et al.* Avaliação formativa e feedback como ferramenta de aprendizado na formação de profissionais da saúde. **Revista de Medicina**, Ribeirão Preto, v. 47, n. 3, p. 324-331, 03 nov. 2014. <https://doi.org/10.11606/issn.2176-7262.v47i3p324-331>

BORGES, Tiago Silva; ALENCAR, Gidélia. Metodologias Ativas na promoção da formação crítica do estudante: o uso das metodologias ativas como recurso didático na formação crítica do estudante do ensino superior. **Cairu em Revista**, Visconde de Cairu, v. 3, n. 4, p. 119-214, jul./ago. 2014. Disponível em: [https://www.cairu.br/revista/arquivos/artigos/2014\\_2/08%20METODOLOGIAS%20ATIVAS%20NA%20PROMOCAO%20DA%20FORMACAO%20CRITICA%20DO%20ESTUDANTE.pdf](https://www.cairu.br/revista/arquivos/artigos/2014_2/08%20METODOLOGIAS%20ATIVAS%20NA%20PROMOCAO%20DA%20FORMACAO%20CRITICA%20DO%20ESTUDANTE.pdf). Acesso em: 23 maio 2020.

BRANCO, Carla Castello; ALVES, Marcia Maria. Complexidade e sala de aula invertida: considerações sobre o método. *In*: CONGRESSO NACIONAL DE EDUCAÇÃO, 12., 2015, Curitiba. **Educere XII**. Curitiba: PUC, 2015. Disponível em: [https://educere.bruc.com.br/arquivo/pdf2015/20881\\_9548.pdf](https://educere.bruc.com.br/arquivo/pdf2015/20881_9548.pdf). Acesso em: 19 ago. 2020.

BRASIL. Resolução nº 466, de 12 de dezembro de 2012. Resolução 466 do CNS que trata de pesquisas em seres humanos e atualiza a resolução 196. **Diário Oficial da União**, Brasília, DF, 14 jun. 2013. Disponível em: [https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466\\_12\\_12\\_2012.html](https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466_12_12_2012.html) Acesso em: 12 maio 2020.

BRASIL. Conselho Nacional de Educação. Câmara de Educação Superior. Resolução CNE/CES 3/2001. Diretrizes Nacionais Curso Graduação Enfermagem. **Diário Oficial da União**, Brasília, DF, 9 Seção 1, p. 37, nov. 2001. Disponível em: [http://www.cofen.gov.br/wp-content/uploads/2012/03/resolucao\\_CNE\\_CES\\_3\\_2001Diretrizes\\_Nacionais\\_Curso\\_Graduacao\\_Enfermagem.pdf](http://www.cofen.gov.br/wp-content/uploads/2012/03/resolucao_CNE_CES_3_2001Diretrizes_Nacionais_Curso_Graduacao_Enfermagem.pdf). Acesso em: 18 jan. 2021.

DE BEM, Amilton Barreto *et al.* Validade e confiabilidade de instrumento de avaliação da docência sob a ótica dos modelos de equação estrutural. **Avaliação**, Campinas; Sorocaba, v. 16, n. 2, p. 375-401, jul. 2011. Disponível em: <https://www.scielo.br/pdf/aval/v16n2/a08v16n2.pdf>. Acesso em: 18 jan. 2021.

FERRAZ, Ana Paula do Carmo Marcheti; BELHOT, Renato Vairo. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. **Revista Gestão & Produção**, São Carlos, v. 17, n. 2, p. 421-431, 2010. <https://doi.org/10.1590/S0104-530X2010000200015>

FREITAS, André Luis Policani; RODRIGUES, Sidilene Gonçalves. Avaliação da confiabilidade de questionário: uma análise utilizando o coeficiente alfa de Cronbach *In*: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 2005, Bauru. **Anais...** Bauru: Unesp, 2005. <https://doi.org/10.13140/2.1.3075.6808>

FREITAS, Sirley Leite; COSTA Michele Gomes Noé da; MIRANDA Flavine Assis de. Avaliação educacional: formas de uso na prática pedagógica. **Meta**, Rio de Janeiro, v. 6, n. 16, p. 85-98, jan./abr. 2014. Disponível em: <https://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/217/pdf>. Acesso em: 18 jan. 2021.

HORA, Henrique Rego Monteiro da; MONTEIRO, Gina Torres Rego; ARICA, José. Confiabilidade em questionários para qualidade: um estudo com o coeficiente alfa de Cronbach. **Produto & Produção**, Porto Alegre, v. 11, n.2, p.85-103, 2010. Disponível em: <https://www.seer.ufrgs.br/ProdutoProducao/article/view/9321/8252> Acesso em: 18 jan. 2021.

IFPR. Instituto Federal do Paraná. **Projeto pedagógico do curso de enfermagem**. Palmas: IFPR, 2019. Disponível em: [https://palmas.ifpr.edu.br/wp-content/uploads/2020/11/PPC\\_ENFERMAGEM\\_2020.pdf](https://palmas.ifpr.edu.br/wp-content/uploads/2020/11/PPC_ENFERMAGEM_2020.pdf) Acesso em: 18 jan. 2021.

LIKERT, Rensis. A technique for the measurement of attitudes. **Archives of Psychology**, New York, n. 140, p. 44-53, 1932. Disponível em: [https://legacy.voteview.com/pdf/Likert\\_1932.pdf](https://legacy.voteview.com/pdf/Likert_1932.pdf). Acesso em: 18 jan. 2021.

LOSTADA, Lauro Roberto. Resenha - Sala de Aula Invertida: uma metodologia ativa de aprendizagem. Do livro de BERGMANN, J.; SAMS, A. **Revista Contexto & Educação**, Rio de Janeiro, v. 32, n. 102, p. 205-209, maio/ago. 2017. <https://doi.org/10.21527/2179-1309.2017.102.205-209>

MARTINS, Gilberto de Andrade. Sobre Confiabilidade e Validade. **Revista Brasileira de Gestão de Negócios**, São Paulo, v. 8, n. 20, p. 1-12, jan./abr. 2006. Disponível em: <https://rbgn.fecap.br/RBGN/article/view/51/44> Acesso em: 18 jan. 2021.

MEDEIROS, Rosana Kelly da Silva *et al.* Modelo de validação de conteúdo de Pasquali nas pesquisas em enfermagem. **Revista de Enfermagem Referência**, Coimbra, Série IV, n. 4, jan./mar. 2015. <http://dx.doi.org/10.12707/RIV14009> Acesso em: 18 jan. 2021.

MELLO, Carolina de Castro Barbosa; ALVES, Renato Oliveira; LEMOS, Stela Maris Aguiar. Metodologias de ensino e formação na área da saúde: revisão de literatura. **Revista CEFAC**, Belo Horizonte, v. 16, n. 6, p. 2015-2028, nov./dez. 2014. <https://doi.org/10.1590/1982-0216201416012>

MOURA, Escolástica Rejane Ferreira *et al.* Validação de jogo educativo destinado à orientação dietética de portadores de diabetes mellitus. **Revista de APS**, Juiz de Fora, v. 11, n. 4, p. 435-443, out./dez. 2008. Disponível em: <https://periodicos.ufff.br/index.php/aps/article/view/14153> Acesso em: 18 jan. 2021.

PAIVA, Marlla Rúbya Ferreira *et al.* Metodologias ativas de ensino-aprendizagem: revisão integrativa. **Revista Sanare**, Sobral, v. 15, n. 2, p. 145-153, jun./dez. 2016. Disponível em: <https://sanare.emnuvens.com.br/sanare/article/viewFile/1049/595> Acesso em: 18 jan. 2021.



PARANÁ. Governo do Estado. Secretaria da Educação. Os desafios da escola pública paranaense na perspectiva do professor PDE: produções didático-pedagógicas. **Cadernos PDE**, v. 2, 2013. Disponível em: [http://www.diaadiaeducacao.pr.gov.br/portals/cadernospde/pdebusca/producoes\\_pde/2013/2013\\_unicentro\\_port\\_pdp\\_serli\\_rech\\_moleta.pdf](http://www.diaadiaeducacao.pr.gov.br/portals/cadernospde/pdebusca/producoes_pde/2013/2013_unicentro_port_pdp_serli_rech_moleta.pdf). Acesso em: 18 jan. 2021.

PASQUALI, Luiz. **Instrumentação psicológica: fundamentos e práticas**. Porto Alegre, Brasil: Artmed, 2010.

PASQUALI, Luiz. Psicometria. **Revista da Escola de Enfermagem USP**, v. 43, n. esp., p. 992-999, 2009. <http://dx.doi.org/10.1590/S0080-62342009000500002>

PASQUALI, Luiz. **Teoria e métodos de medida em ciências do comportamento**. Brasília, DF: Laboratório de Pesquisa em Avaliação e Medida: Instituto de Psicologia: UnB: INEP, 1996. Disponível em: [https://www.faepr.edu.br/site/documentos/teoria\\_metodos\\_ciencias\\_comportamento.pdf](https://www.faepr.edu.br/site/documentos/teoria_metodos_ciencias_comportamento.pdf). Acesso em: 18 jan. 2021.

PILATTI, Luiz Alberto; PEDROSO, Bruno; GUTIERREZ, Gustavo Luis. Propriedades psicométricas de instrumentos de avaliação: um debate necessário. **Revista Brasileira de Ensino de Ciência e Tecnologia**, Ponta Grossa, v. 3, n. 1, p. 81-91, jan./abr. 2010. <https://doi.org/10.3895/S1982-873X2010000100005>

ROMAN, Cassiela *et al.* Metodologias ativas de ensino-aprendizagem no processo de ensino em saúde no Brasil: uma revisão narrativa. **Clinical & Biomedical Research**, Porto Alegre, v. 37, n. 4, 2017. <http://dx.doi.org/10.4322/2357-9730.73911>

VELOSO, Raphael de Mello; SHIMODA, Eduardo; SHIMOYA, Aldo. A confiabilidade em uma pesquisa sobre a qualidade em serviços bancários: um estudo com o coeficiente Alpha de Cronbach. **Revista Científica Linkania Master**, Maringá, v. 5, n. 1, p. 27-51, 2015. Disponível em: <https://linkania.org/master/article/view/293> Acesso em: 18 jan. 2021.

WEBER, Lidia Catarina. **Metodologias ativas no processo de ensino da enfermagem: revisão integrativa**. 2018. Dissertação (Mestrado em Ensino) – Universidade do Vale do Taquari, Lajeado, 2018. Disponível em: <https://www.univates.br/bdu/bitstream/10737/2494/1/2018LidiaCatarinaWeber.pdf>. Acesso em: 18 jan. 2021.

*Received on January 25, 2021*

*Reviewed on May 23, 2022*

*Approved on July 04, 2022*

**Dionara Guarda** holds a degree in nursing from the Federal Institute of Paraná (IFPR) – Palmas campus. Master's student in nursing at the Federal University of Santa Catarina (UFSC) with an emphasis on women's and newborn health.

**Graciela Cabreira Gehlen** has a degree in nursing and midwifery (2001) and a degree in nursing (2003) from the Federal University of Rio Grande (FURG). She holds a master's degree in nursing from FURG (2003) and a doctorate in nursing from the Federal University of Rio

Grande do Sul (UFRGS) 2012. She is currently a professor of the undergraduate nursing course and director of teaching, research, extension and innovation at the IFPR – *campus* Clap.

**Gimene Cardozo Braga** holds a master's degree in practical social sciences in nursing and health, and a bachelor's degree in nursing from the Federal University of Pelotas (UFPEl). She is a specialist in mental health from the Pontifical Catholic University of Rio Grande do Sul (PREMUS / PUCRS). She is a professor of the undergraduate nursing course at IFPR – Palmas *campus*.

**Albimara Hey** has a degree in nursing and is a specialist in contemporary people management from Unicentro. She specializes in medical urgencies and emergencies at the Makro Postgraduate and Extension Institute, master's degree in nursing at UFPR and doctoral candidate at Unicentro. Professor of the undergraduate nursing course at IFPR – Palmas *campus*.