



OPEN ACCESS

EDITED BY

Carlos Trenado,
Heinrich Heine University Düsseldorf,
Germany

REVIEWED BY

Maila Pentucci,
University of Studies G. d'Annunzio Chieti and
Pescara, Italy
María Del Rosario Bonilla-Sánchez,
Benemérita Universidad Autónoma de Puebla,
Mexico

*CORRESPONDENCE

Aldo Bazán-Ramírez
✉ abazanramirez@gmail.com

RECEIVED 24 September 2024

ACCEPTED 26 December 2024

PUBLISHED 28 January 2025

CITATION

Bazán-Ramírez A, Capa-Luque W,
Chávez-Nava R, Dávila-Navarro MC,
Ango-Aguilar H, Hervias-Guerra E and
Bello-Vidal C (2025) Self-assessment of
didactic performance of psychology and
education professors in Mexico and Peru.
Front. Educ. 9:1499598.
doi: 10.3389/feduc.2024.1499598

COPYRIGHT

© 2025 Bazán-Ramírez, Capa-Luque,
Chávez-Nava, Dávila-Navarro, Ango-Aguilar,
Hervias-Guerra and Bello-Vidal. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Self-assessment of didactic performance of psychology and education professors in Mexico and Peru

Aldo Bazán-Ramírez^{1*}, Walter Capa-Luque²,
Roberto Chávez-Nava³, Mónica C. Dávila-Navarro³,
Homero Ango-Aguilar⁴, Edmundo Hervias-Guerra² and
Catalina Bello-Vidal⁵

¹Escuela Profesional de Psicología, Universidad Nacional José María Arguedas, Andahuaylas, Peru, ²Facultad de Psicología, Universidad Nacional Federico Villarreal, Lima, Peru, ³Departamento de Educación, Instituto Tecnológico de Sonora, Ciudad Obregón, Mexico, ⁴Facultad de Ciencias Biológicas, Universidad Nacional de San Cristóbal de Huamanga, Ayacucho, Peru, ⁵Facultad de Medicina, Universidad Nacional Federico Villarreal, Lima, Peru

Introduction: The evaluation of teaching performance in didactic interactions has generally been approached from the student's perspective, with little literature on evaluation based on the teacher's own perspective. Having an instrument that allows self-evaluation of teaching performance during didactic interactions will contribute to improving the quality of teaching and learning at the university level.

Objectives: The validity and reliability evidence for the scores of the self-evaluation questionnaire of the teacher's didactic performance are examined. The self-assessment of didactic performance is analyzed and the general didactic performance of the teacher and of his interactive episodes is contrasted according to some socio-academic characteristics. Method: Methodological, instrumental, descriptive, comparative and predictive cross-sectional study. A total of 203 teachers participated, of whom 64 were professors from a public university in northwestern Mexico, and 139 professors from two public universities in Peru.

Results: Satisfactory evidence of content-based validity was found, Aiken's V estimates were higher than 0.70 for all items in the criteria: clarity, pertinence and relevance. As for the evidence of validity based on the internal structure of the construct, satisfactory results were found with the confirmatory factor analysis strategy for a multidimensional model of seven oblique factors (CFI = 0.99, TLI = 0.98, RMSEA = 0.05, SRMR = 0.06) and a second-order factor model made up of two general factors and seven specific first-order factors (CFI = 0.98, TLI = 0.98, RMSEA = 0.06, SRMR = 0.07). Evidence of convergent and discriminant validity was also acceptable. The reliability for the overall score of the questionnaire as well as for the interactive episodes evidenced McDonald's ordinal alpha and omega coefficients ≥ 0.94 and H coefficient ≥ 0.95 . With respect to the comparative analyses according to the professional training discipline and the sex factor, differences were found for a small effect size ($d > 0.20$) in favor of education teachers and women both in general didactic performance and in the teaching and formative evaluation factors. Likewise, being an education teacher and a woman is more than twice

as likely to have optimal teaching performance. As for the nationality of the teachers, no significant differences were found.

KEYWORDS

self-evaluation, professor, didactic, interaction, teaching, performance

1 Introduction

The didactic performance of the teacher in higher education (undergraduate and graduate) refers to two important elements in university instruction. The first includes teaching behaviors on the part of the teacher, and the second refers to the quality of teaching developed by the teacher in different contexts and different learning modalities in higher education (Aidoo and Shengquan, 2021; Díaz et al., 2015; López-Cámara et al., 2015). However, in contemporary literature there is no agreement as to what constitutes teacher performance in didactic interactions. That is, there is diversity both in the definition and in the evaluation of the performance of teachers during didactic interactions at the university level.

In some cases, teacher behavior, teaching-learning behavior, teaching approaches and content, feedback and formative evaluation have been characterized (Díaz et al., 2015; Chan, 2018; Gitomer, 2019; Bazán-Ramírez et al., 2023; Bazán-Ramírez et al., 2022). In other cases, it has been suggested that teachers' performance involves, Teaching methodology, Design of the course guide (theory and practice), Teachers' attitude, Internal coherence of teaching resources, and Information on evaluation systems (López-Cámara et al., 2015; González and López, 2010). On the other hand, other experts have proposed three levels of hierarchical development of didactic competencies of university teachers: teacher-centered practice, task-centered practice, and information on evaluation systems (López-Cámara et al., 2015; González and López, 2010), and skills focused on student learning (Noben et al., 2021). Also, the constructs of Thinking Skills Practice for Teacher Teaching: Teacher Thinking Effectiveness, Curriculum Loyalty, Teacher Dependence, and Thinking Fostering have been proposed and validated (Arce-Saavedra and Blumen, 2022; Dilekli and Tezci, 2018).

From the perspective of teaching in the health sciences, categories have also been proposed to analyze the teacher's performance in instructional planning, the evaluation of previous knowledge, as well as the relationships that are structured in teaching and learning situations, and the formative evaluation, application and solution of academic and professional problems (Melo and Calheiros, 2023; Pérez García et al., 2023).

On the other hand, for the study of such didactic performance, questionnaires or student self-reports on various categories or characteristics of their teachers' teaching performance have been the most common (López-Cámara et al., 2015; Chan, 2018; Feistauer and Richter, 2018; Neves-Balan et al., 2022). But in the last decade, research on teaching and teaching behaviors has focused on university professor self-evaluation processes (Max et al., 2022; Huang, 2022; Azevedo et al., 2023; Guirão et al., 2020). However, as can be seen in Table 1, research on teacher self-evaluation shows divergences on the categories or criteria of teaching performance (Max et al., 2022; Azevedo et al., 2023; Guirão et al., 2020; Wang et al., 2020; Torres-Delgado and Hernández-Gress, 2021). On one hand, there is a lack of studies focusing on didactic performance criteria per se, i.e., teacher behaviors during didactic interaction. On the other hand, teachers in

psychology teaching are not visualized, and studies on teachers in education teaching have been scarcely developed.

Taking as a theoretical framework the interbehavioral model of didactic performance (Kantor, 1959; Kantor, 1975; Carpio et al., 1998; Irigoyen et al., 2011; Silva et al., 2014), in this study we intend to answer the following research questions: (1) Is the evidence of content, construct, convergent, divergent and discriminant validity satisfactory for the self-assessment measures of the seven criteria of teacher didactic performance and reliability? (2) Is the self-assessment questionnaire of teacher performance invariant according to sex, nationality and professional training discipline? (3) Are there differences in overall teaching performance and in the seven criteria of teacher teaching performance according to professional training discipline, nationality and gender? (4) What is the impact of personal factors on the teaching performance of university teachers?

Derived from these questions, the objectives of the present article were: (1) To evaluate the evidence of content validity of the internal structure of the construct: convergent, divergent and discriminant for the scores of the self-evaluation questionnaire of the teacher's didactic performance. (2) To evaluate the measurement invariance for the teacher self-assessment questionnaire according to gender, nationality and professional training discipline. (3) To compare the overall didactic performance of the teacher and of his or her interactive episodes, according to professional training discipline, nationality and gender. (4) To evaluate the impact of personal factors on the didactic performance of university teachers.

Three general hypotheses were proposed and tested for this research: (1) There is content, convergent, divergent, discriminant and invariance validity in the self-evaluation of university teachers in seven constructs of the interbehavioral model of university teachers' didactic performance. (2) There are significant differences in the didactic performance of university teachers according to professional training discipline, nationality and gender. (3) There are differential effects of university teachers' personal factors on their self-evaluated didactic performance.

This research on the self-evaluation of teachers' didactic performance in the university context, specifically in psychology and education, is of utmost importance for two reasons. (1) It intends to apply teaching constructs under an interbehavioral theoretical model validated conceptually and empirically from the 1990s to date, for self-evaluation by the teachers themselves, meeting criteria of construct validity and measurement invariance. (2) To contribute through scientific research to the process of feedback and improvement of teaching practice in the context of teaching psychology and education at the university level.

Likewise, the results of this study will enrich the debate on how the didactic performance of teachers is conceived and from the different scientific disciplines interested in educational processes.

Also, it is of utmost relevance to expand the studies with a solid theoretical basis and consistent and adequate methods of analysis on the self-evaluation of university professors on their didactic performance.

TABLE 1 Examples of self-evaluation studies of university professors and performance variables.

References	Careers where taught	Teacher performance variables
Azizi et al. (2014)	Health and Nutrition.	Program, Teaching skill, Evaluation, Professional and skillful communication, Class management skill
Guirão et al. (2020)	Agrarian Sciences, Biological Sciences, Health Sciences, Exact and Earth Sciences, Human Sciences, Applied Human Sciences, Engineering and Linguistics, Letters and Arts	Behavior Pragmatic, Behavior Expressive, Behavior Affable, Behavior Analytical
Wang et al. (2020)	Medicine, Dental medicine, Pharmacy, Nursing, Health science, Life science, Humanities and Social science.	Teaching performance, Clarity of information Facilitation, Curriculum planning, Enthusiasm, Resource development
Torres-Delgado and Hernández-Gress (2021)	School of Government, School of Humanities and Education, School of Engineering and Science, School of Medicine and Health Sciences, Business School, School of Architecture, Art, and Design.	Transformation of society, Researcher training Originality and innovation, Teaching, Ethics and citizenship, Teachers' digital competence
Azevedo et al. (2023)	Professors at public university. No careers are specified where classes are taught.	Environmental control, Self-disclosure, Assertiveness, Interaction management, Availability to communication

1.1 Teaching performance from an interbehavioral perspective

Based on the interbehavioral perspective of psychology and educational psychology ([Kantor, 1959](#); [Kantor, 1975](#)), several classifications of teaching performance in didactic situations have been proposed. As a first theoretical proposal and based on the notion of language games to characterize the scientific practice of the psychologist ([Ribes, 1993](#); [Ribes et al., 1996](#)), it has been specified that there are five categories to describe the performance of the teacher during the practice of teaching scientific disciplines, including psychology: Cognitive exploration, Explanation of criteria, Illustration, Practice and Evaluation ([Carpio et al., 1998](#)). Subsequently, in the work of [Ibáñez and Ribes \(2001\)](#), a classification of teaching performance behaviors can be identified in four dimensions: (1) Clearly expressing to students how the established criterion is achieved, (2) Exemplifying to students the performance that leads to the achievement of the specified criterion (or criteria), (3) Arranging the necessary situations for the student to practice the ways in which the specified criterion is achieved, and (4) Informing students promptly about their successes or errors throughout the process of applying knowledge.

Later works of psychologists under an interbehaviorist perspective in educational psychology, were consolidating a model of didactic performance, mainly in the context of secondary and university education, of up to seven criteria or dimensions of the didactic performance of the teacher: Didactic Planning, Competency Exploration, Criteria Prescription/Explicitness, Illustration, Supervision/Supervised Practice, Feedback, and Evaluation ([Ribes, 1993](#); [Morales et al., 2013](#); [Morales et al., 2017](#); [Reyna and Hernández, 2017](#)). These specific criteria or categories of teacher didactic performance would imply two general categorical elements that define it, the first referring to the teaching modality of the referent object, and the second referring to the evaluation of student

learning progress. However, neither the seven specific criteria of teacher didactic performance nor the two generic categories had been empirically validated in classroom situations or didactic interactions.

A study with observational records and student self-reporting of teacher performance in high school science classes in Mexico ([Bazán-Ramírez et al., 2022](#)) constituted a first contribution of empirical evidence to validate the relevance of five criteria or dimensions of teacher didactic performance during classes: Exploration of competences, Explicitation of criteria, Illustration, Feedback and Evaluation. [Bazán-Ramírez et al. \(2023\)](#), observing Peruvian university students of biological sciences, obtained convergent and divergent validity of constructs of a self-report and confirmed the multidimensional structure of the teacher's didactic performance with six criteria: Competence exploration, Explicitness of criteria, Illustration, Practice supervision, Feedback and Evaluation.

With Peruvian graduate students in educational sciences, construct validity was obtained for a restricted model of five criteria of teacher didactic performance with in-variance by sex, age and level of studies ([Bazán-Ramírez et al., 2022](#)). Likewise, by means of second-order models, they confirmed the structure of the two generic components of Teaching and Evaluation. In the first case, Teaching performance was formed by two variables: Explicitation of Criteria and Illustration. In the second case, Formative evaluation was constituted by the variables: Monitoring (supervision), Feedback and Assessment. With Peruvian university students of professional careers in sciences, engineering, humanities and literature, social sciences and health sciences, at the beginning of in-person classes after the COVID-19 Pandemic, [Cuellar-Quispe et al. \(2023\)](#) validated these constructs of explanation of criteria, illustration, supervision of learning practices and activities, feedback, and evaluation. The research findings show that there is good virtual teaching performance,

and that university professors demonstrate their ability to teach in a hybrid way.

Based on the theoretical approaches and these empirical studies derived from them, Figure 1 characterizes an interactive model of didactic performance, thus extending the restricted models validated by Bazán-Ramírez and colleagues presented in the previous paragraphs. For the specific case of teacher performance, this model

includes two generic domains (second-order factors) related to each other: Teaching and Formative evaluation, which configure specific didactic teaching performance criteria. These domains also demarcate the possibility of structuring morphologically and functionally different but interrelated interactive episodes and involve identifying the organization of interactive didactic performances in two generic categories: teaching and feedback. Black bidirectional arrows mark the

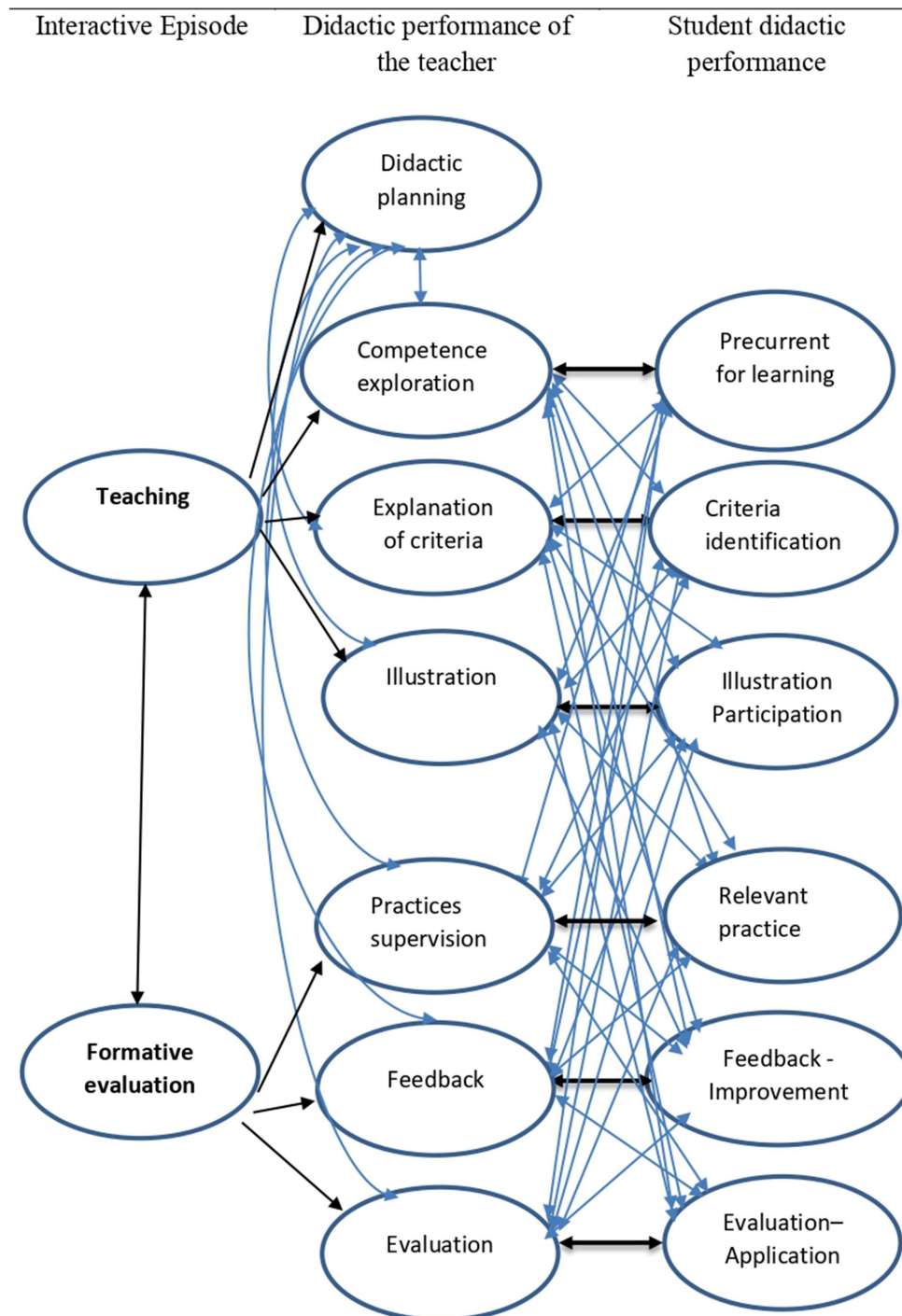


FIGURE 1 Interbehavioral model of teaching performance.

relationship of six didactic performance criteria that are directly functionally related to each other; however, blue arrows illustrate the establishment of other functional relationships between the criteria or areas of didactic teacher-student interaction.

It should be noted that the first criterion of the teacher's didactic performance, didactic planning, even though it is carried out prior to the development of instruction through the various didactic and interactive sequences, always influences them, and is functionally related to the other teacher's performances, both in teaching and formative evaluation. The results of these performances of the teacher in the didactic behavior of his students will also allow the teacher to make the necessary adjustments, oriented in the didactic interaction and in the learning needs of his students and their individual and group characteristics and capabilities.

1.2 Justification of the research questions

The first two questions on the content and construct validity of a scale of self-evaluation of university teachers on their own teaching performance, under a psychological model of teaching performance in the university context, justify the importance for the development of educational psychology, regardless of the theoretical root model, to recover in current research on measurement and evaluation of behaviors, the integration of substantive theory in psychological measurement (Mislevy, 1993; Cole, 1993; Bejar, 1993; Martínez and Moreno, 2002). Thus, obtaining content and construct validity in the self-assessment of teaching performance also implies testing the validity of the categories and criteria derived from a particular theory of university teachers' teaching performance.

On the other hand, question 3 on the invariance of the self-assessment of teaching performance is justified because a self-assessed behavioral measurement instrument must be able to withstand differences of origin, such as gender, nationality and initial professional training of the subjects, in order to explain differences in the teaching performance of university teachers. Likewise, the fourth question is justified due to the importance of personal factors, such as gender, initial training, and the geographic context where they carry out their teaching activity, in explaining the didactic performance of university teachers.

2 Materials and methods

2.1 Research design

Cross-sectional study of a methodological, instrumental (Ato et al., 2013), predictive and comparative nature (Ato and Vallejo, 2015).

2.2 Participants

A total of 203 university professors participated, of which 64 professors taught at a public university in northwestern Mexico; 38 taught in Education Sciences and 26 in Psychology. The other stratum of 139 professors, from two public universities in Peru, taught in Education (54) and Psychology (85). As shown in Table 2, of the total sample, 128 were women and the teachers' ages ranged from 27 to 76 years with an average age of around 50 years. In terms of academic training, 18 teachers only had a bachelor's degree, 99 had a master's degree and 86 had a doctorate.

Only those who agreed to give informed consent were included in the study and teachers from other professions who filled out the instrument virtually or presented inconsistent or atypical response patterns according to the multivariate Mahalanobis distance statistic were excluded.

2.3 Study variables

We took the seven categories or criteria of the interbehavioral model of teaching performance (Kantor, 1975; Carpio et al., 1998), and derived statements (items) that allow teachers to assess themselves with respect to these competency criteria of teaching performance. Each of these seven criteria is described below:

2.3.1 Didactic planning

In this area, the teacher, based on the disciplinary knowledge, the curriculum, and the course program, deploys skills and competencies to describe tasks, activities and situations in which the interaction with the learner will take place, and about what is going to be taught and what is going to be learned.

TABLE 2 Sociodemographic characteristics of the sample.

Age	Years	SD		<i>n</i>	%
Minimum male	29		Gender		
Maximum male	73		Males	75	36.9
			Females	128	63.1
Minimum females	27		Nationality		
Maximum females	76		Mexico	64	31.5
			Peru	139	68.5
Age mean male	50.97	± 11.58	Discipline		
Age mean female	51.44	± 9.92	Education	92	45.3
			Psychology	111	54.7

2.3.2 Competency exploration

Identifying and evaluating the learner's behavior at the beginning of a class (or a semester) in terms of potential skills and competencies and precurent to the didactic interaction. The purpose is to make adjustments to the elaborated plan; likewise, the criteria to explore come from didactic planning.

2.3.3 Explanation of criteria

To put the learner in contact with the disciplinary and didactic criteria that his performance must satisfy. The disciplinary criteria are all the behavioral requirements that are exclusive to the field (subject of the course) of what is being taught, and the didactic criteria are all the behavioral requirements linked to the organization of the daily didactic activities. The teacher makes explicit what will be taught, how it will be taught, what the students must do to achieve the course objectives, etc.

2.3.4 Illustration

Describe to the learner the characteristics of the skills of an expert in the discipline for the solution of a specific problem. The teacher must put the learner in contact with the elements of the behavior of another, pointing out in this description: (a) what he/she was doing, (b) the situation in which he/she was doing it, and (c) what he/she was doing it for. Illustrating consists of linguistically relating the learner to the skills of another, even when that other is the one who teaches (i.e., the one who teaches is used as an example). For better learning, teaching modalities should vary, for example: showing a video, telling experiences, giving a reading, etc.

2.3.5 Supervision of practice

The learner should be involved in controlled problem-solving situations. The control of these situations makes it possible to supervise and correct the learner's performance moment by moment. The skills and competencies to be developed are linked to the regulation of those who learn during the course of the didactic situation, reducing the possibilities of ineffective performance, and making the most of the situation for the benefit of their behavioral development. This moment is commonly identified as "practice" where students "practice what they learn."

2.3.6 Feedback

The learner must make contact with his or her performance in a previous moment. The teacher describes to the learner what he/she did, what he/she did it for, and in what situation he/she did it. Variants of the learner's doing (performance), which were possible at that time but did not materialize, can also be described. Feedback is more complex than saying "right" or "wrong"; it involves enabling the learner to get in touch with his own behavior and its possible variants. It is important to point out that the criteria on the basis of which performance is fed back come from the original didactic plan.

2.3.7 Evaluation

Contrasting the actual performance of the learner with an expected performance, proposed as a learning objective in the initial plan, requires an assessment system that allows determining the degree of similarity between the expected and actual performance. The verification of the student's learning (if he/she managed to reach or fulfill the criteria) must be immediate; timelessness is questionable.

The didactic purpose of this contrast lies not only in the possibility of extending a behavioral qualification, but above all in creating the possibility of making strategic adjustments to subsequent didactic interactions.

2.4 Instrument

Self-assessment of teacher didactic performance. Following the didactic performance model developed by interbehavioral psychologists (Carpio et al., 1998; Irigoyen et al., 2011; Silva et al., 2014; Ribes, 1993; Morales et al., 2017; Velarde-Corrales and Bazán, 2019; Bazán-Ramírez et al., 2022), a self-assessment questionnaire of the teacher's didactic performance was designed in seven categories (dimensions or performance criteria). This questionnaire is composed of 28 statements (items), with a response scale: 1 = Never, 2 = Almost never, 3 = Almost always and 4 = Always. Its consisting of 28 statements (items), with a response scale: 1 = Never, 2 = Almost never, 3 = Almost always and 4 = Always. Table 3 presents each item according to the didactic performance dimension, in which the teacher must choose one of the four response options, according to the frequency with which he/she behaved in correspondence with the statement, during his/her classes in the semester just concluded.

2.5 Procedure

The information collection process was carried out by means of an online form between March and May 2024, both in Mexico and Peru. The online form explained the objectives of the research, the ethical safeguards for the care of personal data, the no-penalty withdrawal and the need for voluntary participation expressed by selecting the corresponding option on the form. In this option the informed consent was provided online which after reading each participant could sign the commitment by marking the option I accept or the option I do not accept. Once the virtual applications were concluded at the end of May, the information from Google forms was retrieved in the EXCEL program and the database was prepared to be exported to the SPSS and R programs.

2.6 Data analysis

The descriptive analysis of the items, as well as the descriptive analyses of the self-assessment of the teacher's teaching performance for each criterion, and the comparative analyses of the general teaching performance and its second-order factors were performed with SPSS software version 25 for Windows. In the contrast of means with Student's t-test and one-way analysis of variance, effect sizes were assessed with Cohen's *d* (0.10 small effect, 0.50 medium effect, 0.80 large effect) and omega squared (0.01 small effect, 0.06 medium effect, 0.14 large effect) (Cohen, 1992).

As for the evidence of content-based validity of the questionnaire, it was quantified with the Aiken *V* coefficient from the assessment of 7 judges on the criteria of clarity, pertinence and relevance of the items (Sireci and Faulkner-Bond, 2014). The Aiken *V* values vary between 0 (total disagreement among the judges) and 1 (total agreement among the judges). A satisfactory evaluation of the items on each

TABLE 3 Self-assessment statements for seven dimensions of teacher performance.

Instructional planning
For each class, I planned the instruction by defining the achievement criteria that my students should reach. My lesson plan defined the activities that I as a teacher must develop in order for my students to achieve my expected accomplishments. My lesson plan precisely detailed the activities that students should perform in the process of learning in the classroom. My lesson plan included activities to evaluate student achievement.
Competence exploration
At the beginning of the course, I assessed my students' prior knowledge, either in writing or orally. At the beginning of a class, I explored my students' prior skills. I asked my students about concepts related to the topic before explaining it. I adjusted the course content according to my students' prior knowledge and skills.
Explanation of criteria
I explained the criteria needed to do some work in class. I described the learning objectives that my students were to achieve in the course. In each class, I stated what my students were expected to learn. I described the necessary skills that the students would develop in the course.
Illustration
I performed examples of how to develop a task or a practice. I solved problems in front of the students, based on the given topic. I described how an expert would solve a problem relevant to the course. About a solution to a problem, I described what the solution consists of, in what situations to apply it and why to apply that solution
Practice supervision
In the course activities, I corrected the activities performed by my students. I collaborated with my students in solving the problem they were solving in the course. I controlled the rules and conditions of the activities to ensure that the students learned effectively. All my students received my supervision during the course activities.
Feedback
I gave feedback on my students' performance in class activities, pointing out successes and mistakes. When giving feedback on my students' performance, I taught them the ways in which they should solve problems or tasks. I taught different ways in which the achievement criteria of my class activities could be met. I provided my students with suggestions on how to improve their skills in relation to the course.
Evaluation
I periodically review and evaluate the theoretical knowledge or basic fundamentals of the subject. I carried out applied evaluations and the solution of practical problems derived from the course. I evaluated the students according to the learning objectives that I showed at the beginning of the course, and that can be found in the syllabus. I evaluated my students' ability to integrate knowledge from other courses with the current course.

Source: Own elaboration.

criterion is assumed if the point estimate of Aiken's V is ≥ 0.70 and for the lower limit of the 95% confidence interval values greater than 0.59 (Penfield and Giacobbi, 2004).

For validity evidence based on the internal structure of the construct used the confirmatory factor analysis strategy, in which a 7-factor multidimensional model was initially examined, followed by a second-order factor model with two generic factors (teaching and formative assessment) containing in its configuration the specific teaching performance criteria (primary factors). As the items are categorical, the variance-adjusted weighted least squares (WLSMV) estimation method based on poly-choric correlation matrix was used (DiStefano and Morgan, 2014). For the evaluation of the fit of the models examined, the chi-square goodness-of-fit test (χ^2) was taken into account, which must offer a p -value greater than 0.05 to be considered a good fit, as well as the recommended robust fit indices such as CFI (Comparative Fit Index), TLI (Tucker Lewis Index),

RMSEA (Root Mean Square Error of Approximation) and SRMR (Standardized Root Mean Square Residual). If CFI and TLI present indexes ≥ 0.90 they indicate adequate fit and good fit if they are ≥ 0.95 , as for RMSEA it is considered adequate fit when its index is ≤ 0.08 and good fit when it is ≤ 0.05 , likewise if SRMR presents an index ≤ 0.08 it means that it presents adequate fit and if it is ≤ 0.06 it means good fit (Penfield and Giacobbi, 2004; DiStefano and Morgan, 2014). These analyses were run with R program version 4.0.2 and RStudio environment with Lavaan 0.6–7 and semTools 0.5–3 packages.

The evidence of internal consistency reliability was estimated by taking into consideration the nature of the categorical graded items. For the didactic performance criteria scores and the overall questionnaire score, the ordinal alpha, McDonald's omega and composite reliability coefficients were estimated. For the latent constructs (second-order factors and general factor) the H coefficient was also estimated (Hu and Bentler, 1999; Kline, 2015).

2.7 Ethical considerations

Both Mexican and Peruvian participants were informed about the objectives, benefits and negative risks of the research. They were asked to read the informed consent form, and if they wished to participate in the study, and if they wished to participate in the study, they were asked to sign it, as well as the authorization for the publication of the derived data, after an explanation of the ethical considerations of the study. In the case of the application in Mexico, the informed consent was printed and signed in person; in the case of the application in Peru, the consent was obtained online and signed virtually.

For both countries, the privacy and confidentiality of the data was protected. The data collected were used only for the purposes of this research. The dignity and rights of the participants were respected, avoiding any form of coercion or prejudice. Finally, the research was handled with integrity and honesty. The data are presented in an accurate, transparent and unbiased manner.

3 Results

3.1 Item analysis

As shown in [Table 4](#), according to the values of skewness and kurtosis, the items do not present normal distribution because there are values greater than ± 1.5 . The table also shows that the corrected homogeneity indexes (CHI) are ≥ 0.30 for all items, indicating that the items have a high discriminative capacity ([DiStefano and Morgan, 2014](#)).

3.2 Evidence of content-based validity of the performance self-assessment questionnaire

According to the results of Aiken's V ([Table 5](#)) all the items satisfy the criteria of clarity, pertinence and relevance because the point estimates are greater than 0.70 and the lower limits of the 95%CI are greater than 0.59, the critical values recommended as satisfactory ([Cohen, 1992](#)). Therefore, the data support the evidence of content-based validity of the teacher didactic performance self-assessment questionnaire.

3.3 Validity based on the internal structure of the teacher didactic performance construct

[Figure 2](#) presents the multifactorial model of teacher didactic performance examined with confirmatory factor analysis. The estimator used was WLSMV (Diagonally Weighted Least Squares with Mean and Variance corrected), which is recommended when there are categorical items.

The overall evaluation of the multifactor model was satisfactory according to the fit indices: $\chi^2(329) = 411.07$,

TABLE 4 Descriptive analysis of self-evaluation items of didactic teaching performance.

	Mean	SD	Skewness	Kurtosis	CHI
Plane01	3.50	0.63	-1.334	2.690	0.488
Plane02	3.61	0.54	-1.320	2.857	0.566
Plane03	3.59	0.58	-1.372	2.459	0.501
Plane04	3.48	0.65	-1.076	1.013	0.465
Explo01	3.35	0.83	-1.217	0.927	0.359
Explo02	3.18	0.82	-0.866	0.390	0.473
Explo03	3.48	0.63	-1.044	1.107	0.513
Explo04	3.34	0.67	-0.697	0.261	0.480
Expli01	3.69	0.60	-2.062	4.215	0.596
Expli02	3.69	0.54	-1.899	4.805	0.464
Expli03	3.48	0.63	-1.044	1.107	0.582
Expli04	3.40	0.59	-0.408	-0.670	0.478
Ilust01	3.59	0.53	-0.750	-0.661	0.296
Ilust02	3.43	0.57	-0.636	1.277	0.381
Ilust03	3.10	0.77	-0.797	0.765	0.412
Ilust04	3.34	0.71	-1.062	1.509	0.507
Super01	3.55	0.568	-0.830	-0.310	0.495
Super02	3.51	0.646	-1.178	1.206	0.547
Super03	3.47	0.570	-0.497	-0.730	0.499
Super04	3.61	0.490	-0.453	-1.828	0.345
Retro01	3.55	0.552	-0.681	-0.619	0.461
Retro02	3.52	0.570	-0.676	-0.540	0.558
Retro03	3.29	0.626	-0.304	-0.632	0.487
Retro04	3.61	0.490	-0.453	-1.828	0.450
Evalu01	3.51	0.602	-1.064	1.476	0.481
Evalu02	3.46	0.631	-0.978	1.003	0.508
Evalu03	3.65	0.517	-1.017	-0.122	0.459
Evalu04	3.34	0.625	-0.388	-0.647	0.530

$p = 0.001$, CFI = 0.96, TLI = 0.95, RMSEA = 0.05 [0.03, 0.06], SRMR = 0.07. Likewise, the evaluation of the individual parameters evidenced high standardized factor loadings (>0.50). These results support the existence of validity based on the internal structure of the construct.

It can also be seen in [Figure 1](#) that almost all the factor loadings are greater than the covariances between the factors, which means that there is divergent and convergent validity. It can also be seen in [Figure 1](#) that all the factor loadings are high (>0.50), implying the existence of convergent validity ([Kalkbrenner, 2021](#)). As can be seen in [Table 5](#), almost all of the values of the square root of Average Variance Extracted ($\sqrt{\text{AVE}} \geq 0.74$) are greater than the interfactor correlations of the multidimensional model (more than 80% are less than 0.74, see [Figure 1](#)), which means that the evidence of discriminant validity is acceptably met ([Muñiz et al., 2005](#)).

TABLE 5 Aiken's V coefficients with confidence intervals (7 judges).

	Clarity		Pertinence		Relevance	
	M	V [IC95%]	M	V [IC95%]	M	V [IC95%]
Plane01	3.71	0.90 [0.71, 0.97]	3.71	0.90 [0.71, 0.97]	3.71	0.90 [0.71, 0.97]
Plane02	3.86	0.95 [0.78, 0.99]	4.00	1.00 [0.85, 1.00]	4.00	1.00 [0.85, 1.00]
Plane03	3.71	0.90 [0.71, 0.97]	3.71	0.90 [0.71, 0.97]	3.71	0.90 [0.71, 0.97]
Plane04	3.71	0.90 [0.71, 0.97]	3.86	0.95 [0.78, 0.99]	4.00	1.00 [0.85, 1.00]
Explo01	4.00	1.00 [0.85, 1.00]	3.86	0.95 [0.78, 0.99]	4.00	1.00 [0.85, 1.00]
Explo02	3.71	0.90 [0.71, 0.97]	3.71	0.90 [0.71, 0.97]	3.86	0.95 [0.77, 0.99]
Explo03	4.00	1.00 [0.85, 1.00]	3.43	0.81 [0.60, 0.92]	4.00	1.00 [0.85, 1.00]
Explo04	3.71	0.90 [0.71, 0.97]	3.43	0.81 [0.60, 0.92]	3.86	0.95 [0.77, 0.99]
Expli01	3.86	0.95 [0.78, 0.99]	3.86	0.95 [0.78, 0.99]	3.86	0.95 [0.77, 0.99]
Expli02	4.00	1.00 [0.85, 1.00]	3.57	0.86 [0.65, 0.95]	3.86	0.95 [0.77, 0.99]
Expli03	4.00	1.00 [0.85, 1.00]	3.43	0.81 [0.60, 0.92]	3.43	0.81 [0.60, 0.92]
Expli04	4.00	1.00 [0.85, 1.00]	3.86	0.95 [0.78, 0.99]	3.71	0.90 [0.71, 0.97]
Ilust01	4.00	1.00 [0.85, 1.00]	3.57	0.86 [0.65, 0.95]	4.00	1.00 [0.85, 1.00]
Ilust02	3.86	0.95 [0.78, 0.99]	3.57	0.86 [0.65, 0.95]	3.57	0.86 [0.65, 0.95]
Ilust03	3.71	0.90 [0.71, 0.97]	3.43	0.81 [0.60, 0.92]	3.43	0.81 [0.60, 0.92]
Ilust04	3.86	0.95 [0.78, 0.99]	3.57	0.86 [0.65, 0.95]	4.00	1.00 [0.85, 1.00]
Super01	3.57	0.86 [0.65, 0.95]	3.57	0.86 [0.65, 0.95]	3.86	0.95 [0.77, 0.99]
Super02	3.43	0.81 [0.60, 0.92]	3.57	0.86 [0.65, 0.95]	3.71	0.90 [0.71, 0.97]
Super03	3.57	0.86 [0.65, 0.95]	3.71	0.90 [0.71, 0.97]	3.71	0.90 [0.71, 0.97]
Super04	3.57	0.86 [0.65, 0.95]	3.71	0.90 [0.71, 0.97]	3.71	0.90 [0.71, 0.97]
Retro01	4.00	1.00 [0.85, 1.00]	4.00	1.00 [0.85, 1.00]	4.00	1.00 [0.85, 1.00]
Retro02	3.86	0.95 [0.78, 0.99]	3.86	0.95 [0.78, 0.99]	3.86	0.95 [0.77, 0.99]
Retro03	3.71	0.90 [0.71, 0.97]	3.57	0.86 [0.65, 0.95]	3.86	0.95 [0.77, 0.99]
Retro04	4.00	1.00 [0.85, 1.00]	4.00	1.00 [0.85, 1.00]	4.00	1.00 [0.85, 1.00]
Evalu01	4.00	1.00 [0.85, 1.00]	3.57	0.86 [0.65, 0.95]	3.57	0.86 [0.65, 0.95]
Evalu02	3.86	0.95 [0.78, 0.99]	3.71	0.90 [0.71, 0.97]	3.71	0.90 [0.71, 0.97]
Evalu03	3.86	0.95 [0.78, 0.99]	3.86	0.95 [0.78, 0.99]	4.00	1.00 [0.85, 1.00]
Evalu04	3.43	0.81 [0.60, 0.92]	3.43	0.81 [0.60, 0.92]	3.43	0.81 [0.60, 0.92]

3.4 Second-order generic factors formative teaching and evaluation and specific teaching performance criteria

Figure 3 presents the model resulting from the interactive analysis between general second-order factors (Teaching and Formative Assessment Categories) that configure the specific didactic performance criteria. A strong covariance ($\phi = 0.96$) can be appreciated between the second-order factors.

In accordance with the theoretical approach, the interactive episode teaching (first second-order factor) contains four primary factors (didactic performance criteria: instructional planning, competency exploration, explicitness of criteria and illustration), according to the path coefficients, the criterion with the greatest variance explained corresponds to explicitness of criteria ($R^2 = 0.90$) and the criterion with the least variance explained corresponds to Illustration ($R^2 = 0.39$). Likewise, the second interactive formative evaluation episode (the other general factor of second order) contains in its structural configuration three primary factors (didactic performance criteria: supervision of practice, feedback and evaluation) highlighting the criterion supervision of practice with the greatest variance explained ($R^2 = 0.84$) and the criterion with the least variance explained corresponds to evaluation ($R^2 = 0.61$).

It is also observed in Figure 3 that all the primary factors are configured by four items as established by the theoretical design. The relationships between the primary factors and the items in all cases present high factor loadings ($\lambda > 0.50$). The interactive model presents indices that denote the presence of good fit: $\chi^2(342) = 446.804$, $p = 0.000$, CFI = 0.95, TLI = 0.94, RMSEA = 0.05 [0.038, 0.066], SRMR = 0.08. All these results support satisfactory evidence of validity based on the internal structure of the construct for a second-order model.

3.5 Measurement invariance

Having established the factor structure of the self-assessment questionnaire of teacher didactic performance, the measurement invariance was analyzed according to sex, nationality and professional discipline of origin. Table 6 shows that the configural invariance (unrestricted model for parameter estimation) presents a very satisfactory fit for the three sociodemographic conditions (CFI > 0.95 and RMSEA < 0.08). As for the nested models (metric, scalar and strict) the recommended values of ΔCFI (> -0.01) and $\Delta RMSEA$ (< 0.015) denote equally the fulfillment of invariance for the three conditions.

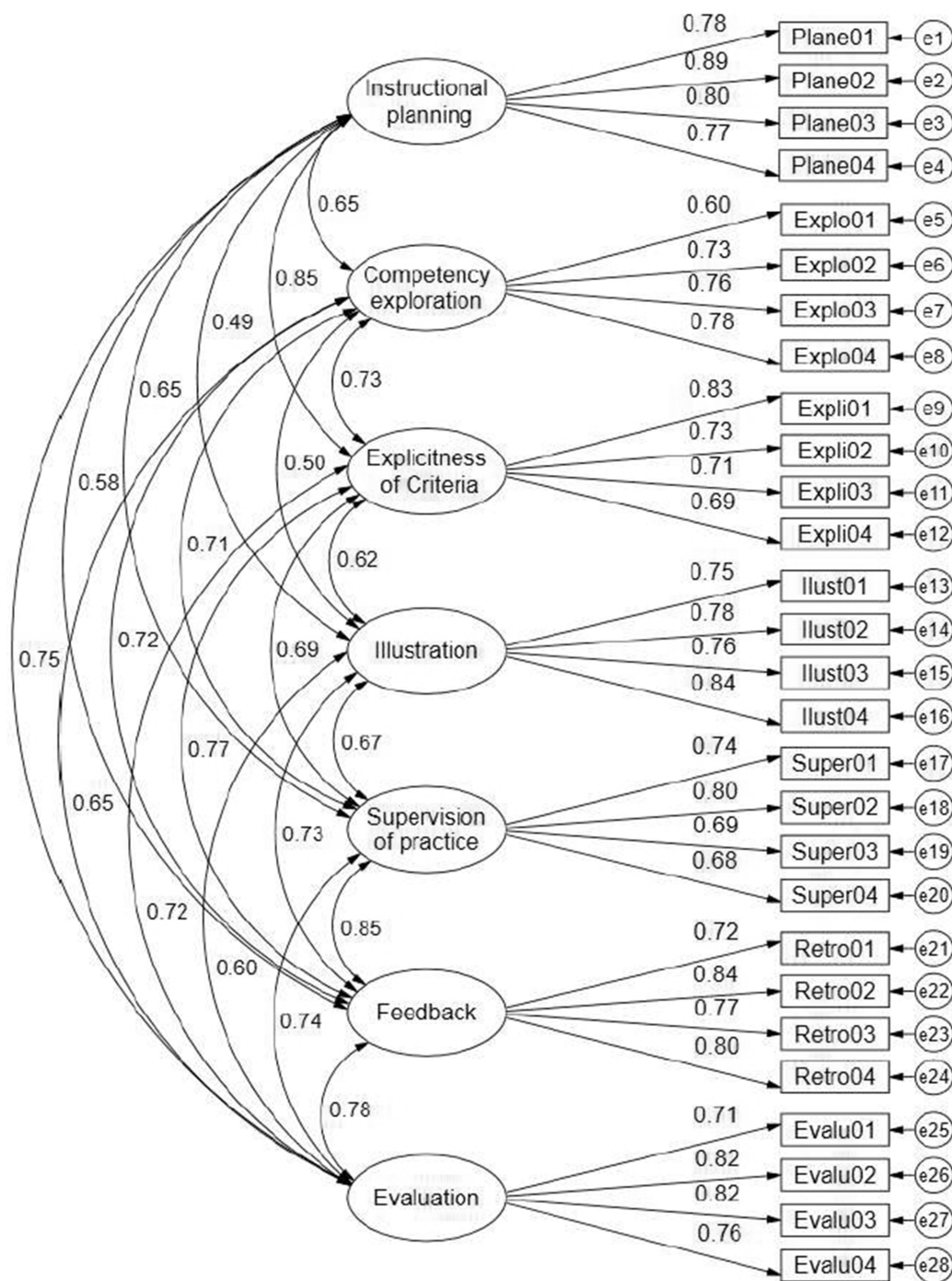
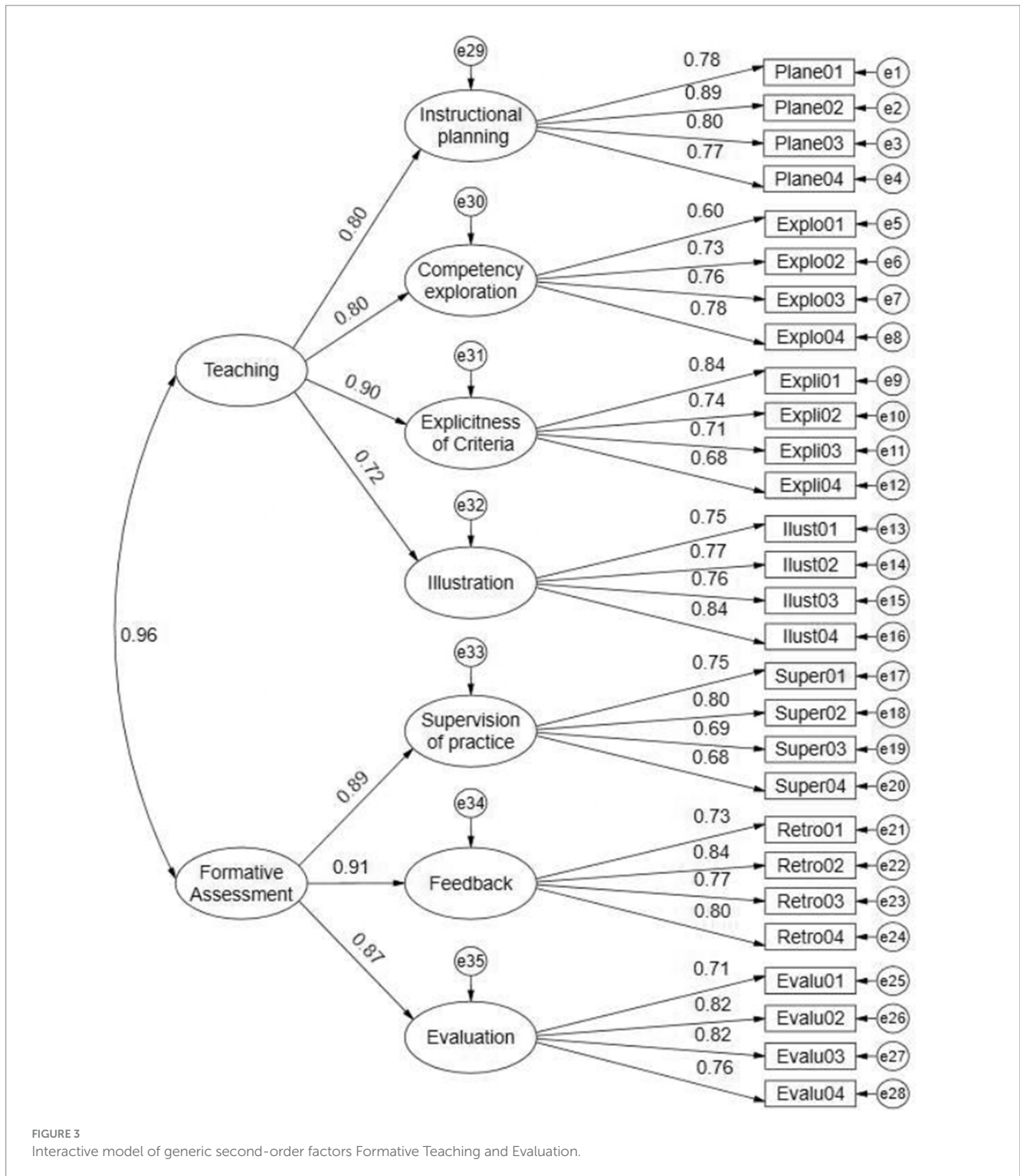


FIGURE 2 Multifactorial model of a self-evaluation questionnaire of teacher’s didactic performance.

3.6 Reliability of the self-evaluation questionnaire for teacher’s teaching performance

Table 7 shows the reliability coefficients estimated by internal consistency for categorical data. The reliability for the overall score of the questionnaire is higher than 0.93 according to Cronbach’s alpha ordinal and McDonald’s omega coefficients, and the re-reliability for the construct is of high precision ($H > 0.95$). On the

other hand, the reliability for the interactive episodes (second order factors) evidence high precision for the scores because the ordinal alpha and omega internal consistency coefficients are ≥ 0.90 ; likewise, these two general factors as latent constructs are of high reliability according to coefficient $H (\geq 0.94)$. Finally, the reliability for the seven factors [didactic performance (criteria) varies between 0.77 and 0.88 according to McDonald’s ordinal alpha and omega coefficients], denoting the presence of high precision for their scores.



3.7 Teachers’ self-assessment of their performance on the seven competency criteria

Table 8 identifies the self-assessment of didactic performance in the competency criteria taking into consideration the lower and upper limits of the CI95% of the mean, the coefficient of variation and the standardized Z scores. Lower values in the

coefficient of variation (CV) indicate greater homogeneity while higher positive Z-scores indicate greater presence of the measured attribute.

According to the context, the competency criteria in which teachers stand out are in the explicitness of criteria, instructional planning and supervision of practice. And the competencies in which strengthening is required are competency exploration, illustration, feedback and evaluation.

TABLE 6 Invariance of teacher didactic performance by sex, nationality and discipline.

Models	χ^2	df	p	CFI	Δ CFI	RMSEA	Δ RMSEA
Sex							
Configural	676.874	658	0.297	0.984		0.017	
Metric	702.683	679	0.257	0.980	-0.004	0.019	0.002
Scalar	724.721	700	0.251	0.980	-0.000	0.019	0.000
Strict	754.566	728	0.240	0.978	-0.002	0.019	0.000
Nationality							
Configural	682.519	658	0.246	0.981		0.019	
Metric	697.117	679	0.307	0.986	0.005	0.016	-0.003
Scalar	726.903	700	0.233	0.979	-0.007	0.020	0.004
Strict	760.249	728	0.198	0.975	-0.004	0.021	0.001
Discipline							
Configural	669.828	658	0.366	0.990		0.013	
Metric	704.423	679	0.242	0.978	-0.012	0.019	0.006
Scalar	727.823	700	0.226	0.976	-0.002	0.020	0.001
Strict	772.362	728	0.124	0.962	-0.014	0.025	0.005

TABLE 7 Reliability coefficients for each criterion of teacher performance.

	Items	$\alpha_{ordinal}$	ω	FC	H	AVE	\sqrt{AVE}
Instructional planning	4	0.88	0.90	0.88	0.90	0.66	0.81
Competence exploration	4	0.81	0.83	0.81	0.82	0.52	0.72
Explanation of criteria	4	0.83	0.87	0.83	0.84	0.55	0.74
Illustration	4	0.86	0.90	0.86	0.87	0.61	0.78
Practice supervision	4	0.82	0.83	0.82	0.83	0.53	0.73
Feedback	4	0.86	0.89	0.86	0.87	0.61	0.78
Evaluation	4	0.85	0.86	0.86	0.87	0.60	0.77
Teaching	16	0.96	0.96	0.96	0.96		
Formative evaluation	12	0.94	0.94	0.94	0.95		
General scale	28	0.98	0.98	0.98	0.98		

$\alpha_{ordinal}$, ordinal alpha; ω , McDonald's omega, composite reliability; H, H coefficient; AVE, average variance extracted; \sqrt{AVE} , square root of AVE.

3.8 Comparison of general didactic performance, teaching and formative assessment

Table 9 presents the results of comparative analyses with the analysis of Student's *t*-test. In the comparison according to the professional training discipline and the gender factor, significant differences ($p < 0.05$) were observed for a small effect size ($d > 0.20$), with higher self-evaluation scores for teachers of education and women in general didactic performance, as well as in the teaching and formative evaluation factors. On the other hand, in the comparison by

TABLE 8 Descriptive analysis of self-assessment of teaching performance for each criterion.

	$n = 203$	Min	Max	Mean [CI 95%]	SD	CV	Z
Instructional planning	4	4	16	14.40 [14.15, 14.64]	1.770	12.29	0.192
Competence exploration	5	5	16	13.59 [13.29, 13.89]	2.142	15.76	-0.244
Explanation of criteria	6	6	16	14.47 [14.24, 14.70]	1.663	11.49	0.229
Illustration	6	6	16	13.63 [13.34, 13.91]	2.029	14.89	-0.226
Practice supervision	7	7	16	14.12 [13.88, 14.36]	1.728	12.24	0.040
Feedback	6	6	16	14.04 [13.80, 14.28]	1.760	12.54	-0.002
Evaluation	8	8	16	14.06 [13.80, 14.32]	1.883	13.39	0.011

SD, standard deviation; CI, confidence interval; CV, coefficient of variation; Z, Z score.

nationality, no differences were observed ($p > 0.05$, $d < 0.20$) in general didactic performance or in the general factors of teaching or formative evaluation.

3.9 Impact of gender and professional training discipline factors on teaching performance

Table 10 presents an ordinal regression model to evaluate the impact of the factors gender and professional training discipline on the probability that university teachers have a regular, acceptable or optimal teaching performance.

TABLE 9 Comparative analysis of self-assessment of overall teaching performance and its factors.

	n	Didactic performance (general)			Teaching			Formative Assessment		
		M ± SD	t	d	M ± SD	t	d	M ± SD	t	d
Discipline of professional training										
Psychology	111	96.1 ± 10.04	-3.56**	-0.50	54.8 ± 5.93	-3.40**	-0.48	41.3 ± 4.73	-3.25**	-0.45
Education	92	100.9 ± 8.65			57.5 ± 5.27			43.3 ± 4.06		
Nationality										
Perú	139	98.4 ± 10.07	0.27	0.04	56.4 ± 5.84	1.23	0.18	42.0 ± 4.74	-0.99	-0.14
México	64	98.0 ± 8.95			55.3 ± 5.66			42.6 ± 4.10		
Gender										
Female	128	99.7 ± 8.74	2.49*	0.38	56.9 ± 5.12	2.74**	0.42	42.6 ± 4.21	1.91	0.27
Male	75	95.9 ± 10.85			54.5 ± 6.54			41.4 ± 5.00		

*p < 0.05, **p < 0.01; n, sample; M, mean; SD, standard deviation; F, F for analysis of variance; t, Student's t; ω², omega squared; d, Cohen's d.

TABLE 10 Ordinal logistic regression model for the effect of personal factors on teaching performance.

Didactic performance	B	Standard error	p	OR	95% CI for OR	
					Lower	Upper
Gender = Female	0.72	0.29	0.012	2.05	1.17	3.60
Training discipline = Education	0.96	0.28	0.001	2.62	1.51	4.55

The estimated logistic model is statistically significant ($\chi^2 = 18.840$, $gl = 2$, $p < 0.001$), which means that personal factors predict the occurrence of didactic performance categories. Likewise, Pearson's Chi-square goodness-of-fit tests ($\chi^2 = 6.976$, $gl = 4$, $p = 0.137$) and the Deviance ($\chi^2 = 7.050$, $gl = 4$, $p = 0.133$) evidence good fit of the data. On the other hand, the parallel lines test presented satisfactory results for the fulfillment of the proportionality assumption ($\chi^2 = 1.356$, $gl = 2$, $p = 0.508$).

Nagelkerke's Pseudo R-squared provides a 10.2% explanation for the overall logistic model of two predictors. Among university professors, specifically women are 2.05 times more likely to have an optimal didactic performance compared to men. Regarding the discipline of professional training of origin, professors of education are 2.62 times more likely to achieve optimal didactic performance in contrast to professors of psychology.

4 Discussion

Overall, this study and the derived results are consistent with Gitomer (2019) and Feistauer and Richter (2018) assumptions for the evaluation of teaching and teaching practice under a substantive theory guiding the assessment of these domains. In the foreground, this questionnaire is linked to a substantive theory of teachers' criteria and competencies in didactic inter-actions, which have been conceptually delimited (Carpio et al., 1998; Irigoyen et al., 2011; Ibáñez and Ribes, 2001; Morales et al., 2013), and which have been validated through observational procedures (direct or indirect), experimental (Morales et al., 2017) and through validation of measures with self-reports from the student's perspective (Reyna and Hernández, 2017). But, the evaluation of the teacher's didactic performance, from the teacher's own perspective, was lacking. In this

sense, both the evidence of content validity and construct validity in two hierarchically different confirmatory models theoretically constitute forms of internal coherence in accordance with the substantive theory on which it is based.

These results of the self-assessment of university teachers showed content validity, construct validity, convergent validity, divergent validity, discriminant validity, invariance of the measure and high reliability for the interpretation of the instrument scores. This means that the data found support the first hypothesis.

As Muñoz (Hair et al., 2014) refers, the basis of any validation process must start with the "task of checking the relevance of the contents; if this fails, everything else, no matter how technically sophisticated it may be, has feet of clay" (p. 103). Content validity evidence conducted through analytical and rational expert judgment is of utmost importance to assess the construct definition and its correct representation (Elosua and Zumbo, 2008). In this sense, the Self-assessment of teacher didactic performance is a questionnaire with satisfactory evidence of content validity given that the point estimates and the lower bounds of the 95% confidence intervals of the Aiken V support the existence of clarity, relevance and pertinence of the items. This implies that the questionnaire is very well represented in quality and sufficiency to assess the construct of teaching performance, which in turn leads to a correct interpretation of the instrument scores (Bazán-Ramírez et al., 2022).

Another important evidence to highlight is the results of the confirmatory factor analysis that presented excellent or very good global assessment indexes of fit for the multidimensional model as well as the individual parameters expressed in the high factor loadings. These results empirically support the evidence of validity based on the internal structure of the construct, which is configured in seven primary factors that correspond to the competency performance

criteria established by the theory on which the design of the questionnaire is based. Likewise, the findings support the existence of a second model whose internal structure is configured with the presence of two highly correlated second-order general factors that contain the primary factors. This means that it is correct or valid to obtain scores for each of the competency criteria, as well as for the second-order general factors, to make inferences and interpretations of the scores (Fornell and Larcker, 1981). For practical purposes the existence of high covariance among the general second-order factors supports the possibility of obtaining an overall score and deriving inferences from it for appropriate decision making (Muñiz, 2018).

Therefore, these findings constitute an important contribution to the understanding of teachers' self-evaluation of their teaching practice. This study constitutes an important practical contribution to the teacher to obtain valuable information about his or her teaching in various didactic performance criteria. These data also provide important theoretical evidence in understanding teacher performance in didactic interaction, providing empirical evidence for the interbehavioral model of didactic performance (Carpio et al., 1998; Irigoyen et al., 2011; Silva et al., 2014). To date, within the interbehavioral perspective on didactic interaction performances, teachers' self-evaluation had not been measured, as had been reported with laboratory observations (Raykov and Marcoulides, 2011; Flora and Flake, 2017), students' self-reporting of their teachers' didactic performances (Bazán-Ramírez et al., 2023; Bazán-Ramírez et al., 2022; Galindo et al., 2017). Another important contribution is that this study confirms the possibility of obtaining simultaneously with empirical evidence in teaching in two different disciplines, such as psychology and educational sciences, the theoretical assumptions of a model of didactic performance outlined from psychology (Carpio et al., 1998; Ribes, 1993; Ribes et al., 1996; Ibáñez and Ribes, 2001; Morales et al., 2013; Morales et al., 2017).

The flexibility of the interbehavioral model of didactic interactions to generate self-evaluations of the performance of teachers from two different disciplines is noteworthy.

Another new and important contribution of this study has to do with the measurement invariance for Self-assessment of teacher didactic performance according to gender, nationality and professional discipline of origin. The factorial invariance results for the three sociodemographic conditions were not affected by the gradually imposed restrictions because the fit of the restricted models did not deteriorate. The base model (configural invariance) by showing satisfactory fit indices implies the existence of the same multidimensional model for the internal structure of the construct. For the re-stricted models in the progressive stages the recommended fit indices ($\Delta CFI > -0.01$ and $\Delta RMSEA < 0.015$) were found within the expected threshold (Chen, 2007; Cheung and Rensvold, 2002), thus denoting that the construct (didactic performance and teacher competence criteria) have the same meaning (metric invariance) for the groups constituted by the three sociodemographic conditions, also the scores for the construct and the didactic criteria are equivalent in all groups (scalar invariance), and finally it should be noted that the questionnaire items measure with the same precision in each of the groups (strict invariance). In summary, the invariance shows that the instrument can present the same internal structure (multidimensional with primary and secondary factors) to evaluate the didactic performance of teachers, regardless of their country of origin (Mexico or Peru), sex (male or female) and profession (psychologist or

educator). The understanding of the items and the interpretation of the scores is similar for these three conditions.

It will be necessary to contrast these preliminary findings in applications with a larger number of teachers, and to test the invariance of the questionnaire itself and its derived constructs, in comparisons between groups, considering for example: the discipline of origin, academic qualification, length of service, gender, discipline or area in which they teach, rank or prestige, age, level of demand, number of students in charge, type of educational institution (public or private), educational modality (face-to-face, blended, virtual and distance) on the part of the teachers (Feistauer and Richter, 2018; Neves-Balan et al., 2022; Scherer and Gustafsson, 2015). Similarly, it should be contrasted with the students' assessment of their teacher's performance in these same categories in which the teacher self-evaluated, as well as the evaluation of the teacher's supervisors, under some of the seven established parameters (Max et al., 2022; Huang, 2022; Azevedo et al., 2023; Guirão et al., 2020; Wang et al., 2020; Torres-Delgado and Hernández-Gress, 2021).

While the validity evidence presented above is important for the inferences and decisions that can be made from the Self-assessment of teacher didactic performance scores, it is also necessary that the scores are not affected by the inaccuracies of the instrument but that the instrument shows reliability for the scores. In this sense, given that the self-assessment questionnaire is designed with categorical graded response items, recommended internal consistency coefficients (Sireci and Faulkner-Bond, 2014) such as ordinal alpha and ordinal McDonald's omega estimated from a polychoric correlation matrix were estimated, which showed high precision for the instrument scores. In the framework of structural equation modeling the H coefficient which is suggested as the most accurate indicator for latent variable reliability exceeds the cutoff values of 0.70, which means that the construct is very well represented by the items (Dominguez-Lara, 2016).

Another important contribution of the study refers to the comparison of didactic performance and the general factors of second order teaching and formative assessment, in which the findings partially support the second and third hypotheses. According to the discipline of professional training, education teachers are the ones who mostly achieve optimal performance in contrast to psychology teachers. In part, a reasonable explanation could be that teachers with professional training must have developed didactic performance competencies more homogeneously, while psychologists do not have this training; they learn didactic competencies after graduation when the teaching vocation arises.

Regarding gender, there were differences in didactic performance scores and second-order factors, in favor of female teachers. Although other studies have found no differences according to the teacher's gender (Minaya et al., 2022; Vargas, 2023; Al Khazaleh and Hawamdeh, 2023), self-assessments of performance should continue to be studied in relation to gender and other educational and demographic variables. An important fact is that the nationality of being Mexican or Peruvian is not a discriminating factor for didactic performance in university teachers.

Finally, it should be noted that, teachers' self-assessment of their own didactic and behavioral competencies should allow them to analyze and make adjustments and improvements in their didactic performance as teachers. In the same way, based on this information, teachers can reflect on the performance of their students against each

of the teacher performance criteria, identifying student factors that may hinder their learning, and the teaching of both professional and research competencies in their discipline (Vargas, 2023; Al Khazaleh and Hawamdeh, 2023). The self-assessment questionnaire (Dilekli and Tezci, 2018; Melo and Calheiros, 2023; Pérez García et al., 2023; Feistauer and Richter, 2018; Neves-Balan et al., 2022; Scherer and Gustafsson, 2015; Azizi et al., 2014) can be used in a complementary way to the analysis of observational records of didactic interaction to improve the training process and to provide feedback to the teacher's practice (Díaz et al., 2015; Bazán-Ramírez et al., 2023; Bazán-Ramírez et al., 2022; Flora and Flake, 2017; Galindo et al., 2017).

The most important limitation of the study would correspond to the external validity due to the type of non-probabilistic sampling used, in this sense, the possibility of generalization should be made with caution; however, despite the limitation, the pre-sent study is considered important in view of the lack of existing knowledge regarding studies that address evaluations of teaching performance from the perspective of university teachers. As already stated, these preliminary findings leave open the need for further research with samples of both teachers and students.

5 Conclusion

We have a self-assessment questionnaire of the teacher's didactic performance with robust evidence of construct validity and reliability. It is invariant for its application, obtaining and interpreting its scores in a similar way for university teachers in Peru and Mexico, in the professions of Psychology and Education, without distinction of sex. These facts are very important for decision making to optimize the quality of education and, above all, to facilitate the understanding of teachers' self-assessment of their teaching practice.

There are differences in general teaching performance and in specific teaching and formative assessment performances, depending on the training discipline of university teachers. There are also differences according to the sex of the professors, but not by nationality. Nationality is not a differentiating factor between teachers from Mexico and Peru in their teaching performances in the university context.

Regarding the effect of university teachers' personal factors on their self-evaluated didactic performance, both gender (Female) and the discipline of their training of origin (Education) are significant predictors of the didactic performance of the university teachers evaluated in this study.

In this group of Mexican and Peruvian university professors who teach psychology and education, self-assessed performances were obtained at a moderately high level, between 13.59 and 14.47 in a range from 0 to 16 (Maximum average achievement = 16). However, it was performance in the explanation of criteria and performance in instructional planning that had the highest average presence according to the teachers themselves during their classes or practices. Likewise, the exploration of competencies was the teaching performance criterion with the lowest self-assessment in terms of its presence in didactic interactions.

This study has an imminent and growing potential in the generation and application of knowledge regarding teaching performance, regardless of the theoretical perspective, and the results can be used to generate interventions with teachers and institutions

that contribute to the improvement of the quality of teaching and learning at the university level, particularly in Psychology and Educational Sciences. Likewise, the refresher and training courses for these university teachers should focus not only on the teaching competencies in the teaching performances with lower evaluation, but also on the seven criteria or areas of didactic performance studied here.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The application of self-report scales to teachers was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all subjects involved in the study.

Author contributions

AB-R: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. WC-L: Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. RC-N: Data curation, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. MD-N: Investigation, Methodology, Supervision, Visualization, Project administration, Writing – review & editing. HA-A: Conceptualization, Data curation, Investigation, Resources, Visualization, Supervision, Validation, Writing – review & editing. EH-G: Data curation, Formal analysis, Investigation, Resources, Validation, Visualization, Supervision, Writing – review & editing. CB-V: Data curation, Investigation, Software, Validation, Visualization, Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aidoo, A., and Shengquan, L. (2021). The conceptual confusion of teaching quality and teacher quality, and a clarity pursuit. *IJAE* 2, 98–119. doi: 10.46966/ijae.v2i2.168
- Al Khazaleh, T. H., and Hawamdeh, M. F. (2023). Jordanian Arabic language teachers' self-assessment of their language teaching approach practices. *Theory Pract. Lang. Stud.* 13, 2830–2840. doi: 10.17507/tpls.1311.13
- Arce-Saavedra, B. J., and Blumen, S. (2022). Critical thinking, creativity, self-efficacy, and teaching practice in Peruvian teacher trainers. *Rev. Psicol. (Peru)* 40, 603–633. doi: 10.18800/psico.202201.020
- Ato, M., López, J. J., and Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología [a classification system for research designs in psychology]. *Anales Psicol.* 29, 1038–1059. doi: 10.6018/analesps.29.3.178511
- Ato, M., and Vallejo, G. (2015). Diseños de investigación en psicología [research designs in psychology]. 1st Edn. Madrid: Ediciones Pirámide.
- Azevedo, C., Mesquita, A., and Caldas, L. (2023). University professors' communicative competence and its relationship with interpersonal communication and voice symptoms. *J. Voice* 37:25. doi: 10.1016/j.jvoice.2023.08.025
- Azizi, K., Aghamolaei, T., Parsa, N., and Dabbaghmanesh, T. (2014). Comparison of faculty Members' self-assessment and their evaluation by students. *J. Adv. Med. Educ. Prof.* 2, 108–113.
- Bazán-Ramírez, A., Ango-Aguilar, H., Cárdenas-López, V., Anaya-González, R. B., Capa-Luque, W., and Bazán-Ramírez, M. A. (2023). Self-reporting of teacher–student performance in virtual class interactions in biological sciences during the SARS-CoV-2/COVID-19 pandemic. *Sustain. For.* 15:16198. doi: 10.3390/su152316198
- Bazán-Ramírez, A., Capa-Luque, W., Bello-Vidal, C., and Quispe-Morales, R. (2022). Influence of teaching and the teacher's feedback perceived on the didactic performance of peruvian postgraduate students attending virtual classes during the covid-19 pandemic. *Front. Educ.* 7:818209. doi: 10.3389/feduc.2022.818209
- Bazán-Ramírez, A., Quispe-Morales, R., De La Cruz, C., and Henostroza-Mota, C. (2022). Teacher–student performance criteria during online classes due to COVID-19: self-report by postgraduate students in education. *Eur. J. Educ. Res.* 11, 2101–2114. doi: 10.12973/eu-jer.11.4.2101
- Bejar, I. I. (1993). "A generative approach to psychological and educational measurement." In *N* in Test theory for a new generation of test. eds. N. Frederiksen, R. J. Mislevy and I. I. Bejar (New Jersey: Lawrence Erlbaum), 323–357.
- Carpio, C., Pacheco, V., Canales, C., and Flores, C. (1998). Comportamiento inteligente y juegos de lenguaje en la enseñanza de la psicología [intelligent behavior and language games in the teaching of psychology]. *Acta comportamental* 6, 47–60. doi: 10.32870/ac.v6i1.18244
- Chan, W. M. (2018). Teaching in higher education: students' perceptions of effective teaching and good teachers. *Soc. Sci. Educ. Res. Rev.* 5, 40–58.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model.* 14, 464–504. doi: 10.1080/10705510701301834
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255. doi: 10.1207/S15328007SEM0902_5
- Cohen, J. A. (1992). Power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Cole, N. S. (1993). "Comment on chapters 1-3." In *N* in Test theory for a new generation of test. eds. N. Frederiksen, R. J. Mislevy and I. I. Bejar (New Jersey: Lawrence Erlbaum), 72–77.
- Cuellar-Quispe, S., Huaman-Romani, Y. L., Sarmiento-Campos, N. V., Silvera-Alarcon, E. N., and Nolasco-Carbajal, E. (2023). Perspectives on teaching performance after COVID-19. *Int. J. Innov. Res. Sci. Stud.* 6, 826–834. doi: 10.53894/ijirss.v6i4.2038
- Díaz, M. C., Borges, A., Valadez, M. D., and Zambrano, R. (2015). Valoración de buenas prácticas docentes a través de observación sistemática [assessing of good teaching practices through systematic observation]. *Univ. Psychol.* 14, 913–922. doi: 10.11144/Javeriana.upsy14-3.vbpd
- Dilekli, Y., and Tezci, E. (2018). Adaptation of teachers' Self-efficacy towards teaching thinking skills scale into English. *J. Educ. Train. Stud.* 6, 260–270. doi: 10.11114/jets.v6i11.3630
- DiStefano, C., and Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Struct. Equ. Model.* 21, 425–438. doi: 10.1080/10705511.2014.915373
- Dominguez-Lara, S. A. (2016). Evaluation of the construct reliability through the index H: brief conceptual review and applications. *Psychologia* 10, 87–94. doi: 10.21500/19002386.2134
- Elosua, P., and Zumbo, B. D. (2008). Coeficientes de fiabilidad Para escalas de respuesta categórica ordenada [reliability coefficients for ordered categorical response scales]. *Psicothema* 20, 896–901.
- Feistauer, D., and Richter, T. (2018). Validity of students' evaluations of teaching: biasing effects of likability and prior subject interest. *Stud. Educ. Eval.* 59, 168–178. doi: 10.1016/j.stueduc.2018.07.009
- Flora, D. B., and Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological re-search: decisions for scale development and validation. *Can. J. Behav. Sci.* 49, 78–88. doi: 10.1037/cbs0000069
- Fornell, C., and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.1177/002224378101800104
- Galindo, L., Silva, H., Serrano, V., Rocha, E., and Galguera, R. (2017). Aprendizaje por observación de interacciones didácticas de ilustración y retroalimentación [Interbehavioral survey about learning by observation in didactic interactions of illustration and feedback]. *Interacciones* 3, 131–140. doi: 10.24016/2017.v3n3.71
- Gitomer, D. H. (2019). Evaluating instructional quality. *Sch. Eff. Sch. Improv.* 30, 68–78. doi: 10.1080/09243453.2018.1539016
- González, I., and López, A. B. (2010). Sentando las bases Para la construcción de un Modelo de evaluación a las competencias docentes del profesorado universitario [foundation of an assessment model of teaching competences among university teachers]. *Rev. Invest. Educ.* 28, 403–423.
- Guirão, I., Ribeiro, N., Caldas, L., and Mesquita, A. (2020). Self-referred personal behavior profile of university professors: association with communicative and vocal self-evaluation. *CoDAS* 32:e20180141. doi: 10.1590/2317-1782/20192018141
- Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2014). Multivariate data analysis. 7th Edn. London: Pearson.
- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118
- Huang, Q. (2022). Influence of EFL teachers' self-assessment on their self-regulation and self-efficacy. *Front. Psychol.* 13:891839. doi: 10.3389/fpsyg.2022.891839
- Ibáñez, C., and Ribes, E. (2001). Un análisis interconductual de los procesos educativos [an Interbehavioral analysis of educational processes]. *Rev. Mex. Psicol.* 18, 359–371.
- Irigoyen, J., Acuña, F., and Jiménez, M. Y. (2011). "Interacciones didácticas en educación superior. Algunas consideraciones sobre la evaluación de desempeños [didactic interactions in higher education. Some considerations on performance evaluation]" in Evaluación de desempeños académicos. eds. J. Irigoyen, F. Acuña and M. Jiménez (Hermosillo: Interactum), 73–96.
- Kalkbrenner, M. T. (2021). Alpha, omega, and H internal consistency reliability estimates: reviewing these options and when to use them. *Couns. Outcome Res. Eval.* 14, 77–88. doi: 10.1080/21501378.2021.1940118
- Kantor, J. R. (1959). *Interbehavioral psychology*. Chicago: The Principia Press.
- Kantor, J. R. (1975). Education in psychological perspective. *Psychol. Rec.* 25, 315–323. doi: 10.1007/BF03394321
- Kline, R. (2015). *Principles and practice of structural equation modeling*. 4th Edn. DC., EEUU: The Guilford Press, Washington.
- López-Cámara, A. B., González-López, I., and de León-Huertas, C. (2015). Exploratory factor analysis to construct a model of university teaching evaluation indicators. *Cult. Educ.* 27, 337–371. doi: 10.1080/11356405.2015.1035546
- Martínez, R., and Moreno, R. (2002). "Integración de teoría sustantiva, diseño de pruebas y modelos de análisis en la medición psicológica [integration of substantive theory, test design and models of analysis in measurements of the psychological]" in Estrategias de evaluación y medición del comportamiento en Psicología / evaluation and measurement strategies in psychology. eds. A. Bazán and A. Arce (Ciudad Obregón: ITSON-UADY), 87–119.
- Max, A. L., Lukas, S., and Weitzel, H. (2022). The relationship between self-assessment and performance in learning TPACK: are self-assessments a good way to support preservice teachers' learning? *J. Comput. Assist. Learn.* 38, 1160–1172. doi: 10.1111/jcal.12674
- Melo, L. M., and Calheiros, D. S. (2023). Didactic sequences as an educational product to facilitate teaching learning processes in Lato Sensu graduate courses in the area of health Management in Primary Care. *Health* 15, 495–506. doi: 10.4236/health.2023.156033

- Minaya, M., Requena, G., Mamani-Benito, O., and Landa-Barzola, M. (2022). Predictores de la autoeficacia profesional en docentes universitarios de salud durante la Pandemia COVID-19 [predictors of professional self-efficacy in university health teachers during the COVID-19 pandemic]. *Rev. Cubana Enferm.* 38:e4560.
- Mislevy, R. J. (1993). "Foundations of a new test theory" in Test theory for a new generation of tests. eds. N. Frederiksen, R. J. Mislevy and RJB. II (New Jersey: Lawrence Erlbaum), 19–39.
- Morales, G., Alemán, M., Canales, C., Arroyo, R., and Carpio, C. (2013). Las modalidades de las interacciones didácticas: entre los disensos esperados y las precisiones necesarias [the modalities of didactic interactions: between the expected disagreements and the necessary clarifications]. *Conductual* 1, 73–89. doi: 10.59792/DSKU8151
- Morales, G., Peña, B., Hernández, A., and Carpio, C. (2017). Competencias didácticas y competencias de estudio: su integración funcional en el aprendizaje de una disciplina [didactic competencies and study competencies: their functional integration in the learning of a discipline]. *Rev. Altern. Psicol.* 37, 24–35.
- Muñiz, J. (2018). Introducción a la Psicometría. Teoría clásica y TRI [introduction to psychometrics. Classical theory and IRT] (Madrid: Pirámide).
- Muñiz, J., Fidalgo, A. M., García-Cueto, E., Martínez, R., and Moreno, R. (2005). Análisis de los ítems [item analysis]. Madrid: La Muralla.
- Neves-Balan, R., Bender-Haydu, V., Henrique de Almeida, J., Oliveira-Henklaín, M. H., and Zacyntho-Zacarin, M. R. (2022). Teacher behavior checklist e implicit relational assessment procedure na avaliação de professores [teacher behavior checklist and implicit relational assessment procedure in teacher evaluation]. *Rev. Psicol. (Peru)* 40, 553–577. doi: 10.18800/psico.202201.018
- Noben, I., Maulana, R., Deinum, J. F., and Hofman, W. H. A. (2021). Measuring university teachers' teaching quality: a Rasch modelling approach. *Learn. Environ. Res.* 24, 87–107. doi: 10.1007/s10984-020-09319-w
- Penfield, R. D., and Giacobbi, P. R. Jr. (2004). Applying a score confidence interval to Aiken's item content-relevance index. Measurement in physical education and exercise science. *Meas. Phys. Educ. Exerc. Sci.* 8, 213–225. doi: 10.1207/s15327841mpee0804_3
- Pérez García, S., Cala Peguero, T. Y., Herrera Miranda, G. L., and Díaz, C. M. (2023). Retos y perspectivas del proceso enseñanza aprendizaje de la disciplina Farmacología en la Carrera medicina [challenges and perspectives of the teaching-learning process of the pharmacology discipline in the medical career]. *Salud Ciencia Tecnol* 2:491. doi: 10.56294/sctconf2023491
- Raykov, T., and Marcoulides, G. A. (2011). Introduction to psychometric theory. New York: Routledge.
- Reyna, W., and Hernández, M. (2017). Enseñanza-aprendizaje de la psicología: reflexiones desde la matriz científica interconductual [teaching-learning of psychology: reflections from interbehavioral scientific matrix]. *Interacciones* 3, 171–182. doi: 10.24016/2017.v3n3.67
- Ribes, E. (1993). La práctica de la investigación científica y la noción de juego del lenguaje [the practice of scientific research and the notion of language game]. *Acta Comportamental* 1, 63–82. doi: 10.32870/ac.v1i1.18207
- Ribes, E., Padilla, A., and Moreno-Rodríguez, R. (1996). Un análisis funcional de la práctica científica: extensiones de un Modelo psicológico [a functional analysis of scientific practice: extensions of a psychological model]. *Acta Comportamental* 4, 205–235. doi: 10.32870/ac.v4i2.18284
- Scherer, R., and Gustafsson, J. E. (2015). Student assessment of teaching as a source of information about aspects of teaching quality in multiple subject domains: an application of multilevel bifactor structural equation modeling. *Front. Psychol.* 6:1550. doi: 10.3389/fpsyg.2015.01550
- Silva, H., Morales, G., Pacheco, V., Camacho, A., Garduño, H., and Carpio, C. (2014). Didáctica Como conducta: una propuesta Para la descripción de las habilidades de enseñanza [didactic as behavior: a proposal for the description of teaching skills]. *Rev. Mex. Anal. Conducta* 40, 32–46. doi: 10.5514/rmac.v40.i3.63679
- Sireci, S. G., and Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema* 1, 100–107. doi: 10.7334/psicothema2013.256
- Torres-Delgado, G., and Hernández-Gress, N. (2021). Research professors' self-assessment of competencies. *Future Internet* 13:41. doi: 10.3390/fi13020041
- Vargas, I. (2023) Análisis comparado de las competencias docentes promovidas en pregrado y posgrado en la ECEDU-UNAD durante el periodo comprendido entre 2018–2022 [comparative analysis of teaching competencies promoted at undergraduate and postgraduate levels at ECEDU-UNAD during the period 2018–2022]. [Master's thesis] [Bogotá]: Universidad Nacional Abierta y a Distancia
- Velarde-Corrales, N., and Bazán, A. (2019). Sistema observacional Para analizar interacciones didácticas en clases de ciencias en bachillerato [observational system to analyze didactic interactions in science classes in bachelor]. *Rev. investig. psicol.* 22, 197–216. doi: 10.15381/rinvp.v22i2.16806
- Wang, C. F., Wang, R. H., Yen, P. Y., Huang, S. L., Huang, Y. B., and Lin, Y. C. (2020). Construction and validation of a teacher self-evaluation scale to measure teaching performance in a medical university. *J. Med. Educ.* 24, 195–208. doi: 10.6145/jme.202012_24(4).0001