Student Evaluation of Teachers' Effectiveness (SETE) scale: translation, cross-cultural adaptation and psychometric properties in a Latin American sample

Miluska Villar-Guevara[1,2]*, José Hector Livia-Segovia[3], Elizabeth Emperatriz García-Salirrosas[4] and Israel Fernández-Mallma[5]

[1]UPG de Ciencias Humanas y Educación, Escuela de Posgrado, Universidad Peruana Unión, Lima, Peru, [2]EP de Administración, Facultad de Ciencias Empresariales, Universidad Peruana Unión, Juliaca, Peru, [3]Universidad Nacional Federico Villarreal, Lima, Peru, [4]Faculty of Management Science, Universidad Autónoma del Perú, Lima, Peru, [5]EP de Ingeniería Ambiental, Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Juliaca, Peru

Introduction: Teacher effectiveness is a burgeoning field. Those responsible for educational policies seem increasingly committed to this fact, since it is one of the most important factors that influence the success of a university student; for this reason, the study of this topic has gained relevance in recent years. Therefore, an instrument with adequate psychometric properties is needed to measure this construct in Spanish-speaking countries.

Objective: The study had the purpose of translating, adapting, and evaluating the internal structure, providing evidence of reliability and validity of the Student Evaluation of Teachers' Effectiveness (SETE) Scale in a Latin American sample.

Methods: An instrumental study was carried out and through convenience sampling, data were collected from 1,000 university students from South America belonging to a private educational network. The participants were between 18 and 40 years old (M = 21.25, SD = 2.99). Analyzes used Structural Equation Modeling (SEM) with AMOS 24 statistical software.

Results: Confirmatory Factor Analysis provided a 4-factor, 28-item fit model (CMIN/DF = 4.359; CFI = 0.956; SRMR = 0.030; RMSEA = 0.058). The results demonstrated good internal consistency (α = between 0.927 and 0.961; CR = between 0.927 and 0.962; AVE = between 0.646 and 0.799). Evidence of validity and reliability was obtained for the total sample.

Discussion: This adaptation and validation of the SETE scale makes it a valid, useful, reliable, and necessary tool that can be considered to evaluate teachers' effectiveness from the perspective of university students.

KEYWORDS

teaching effectiveness, higher education, Latin America, psychometric properties, SETE

1 Introduction

Teaching-learning is fundamental in higher education (Knol et al., 2016) and plays a crucial role in developing individuals, societies, and communities (Al Kuwaiti et al., 2021). Its importance lies in several aspects: Acquisition of knowledge, development of skills, socialization, economic development, innovation and progress, social mobility, autonomy and decision-making, improvement of quality of life, and social and cultural change (dos Santos et al., 2018; Martínez-Huamán et al., 2022; Oweis et al., 2022). Unfortunately, university reality reveals that many undergraduate students do not learn meaningfully, and scientific evidence confirms that the reasons are born in a culture and methodology of teaching and evaluation, where teachers and students are key pieces (Tadesse et al., 2021). On the other hand, longitudinal studies have shown that intensive programs on the professional development of teaching practice may not be effective in achieving changes in favor of teaching practice, even in this era where knowledge has grown (Hobbiss et al., 2021).

Recent studies suggest that, to improve teaching practice, qualitative feedback is invaluable. This is similar to feedback offered to students, arising from a teacher's careful observation of students during cooperative learning activities. Thoughtful, critical commentary from students can help educators identify the root cause of joy or frustration in a class session and use it to improve the next episode of teaching and learning (Harrison, 1987; Holland, 2019; Bardach and Klassen, 2020; Imron, 2024). Consequently, some academics have developed teaching effectiveness models to understand this topic's behavior better. The models have included characteristics such as teacher personality, content knowledge, communicative competence, organizational and preparation skills, and performance evaluation. These models have been used to support various empirical studies. The theoretical models of Stronge (2007) and Danielsons (2007) have been applied in school environments, while Witcher et al. (2003) and Faranda and Clarke (2004) designed theoretical models for higher education.

There is a debate in academia about whether teachers are more effective as their classroom experience increases (Coady et al., 2020; Galmes-Panades et al., 2021). A teacher's teaching ability comes not only from classroom experience, but from other sources such as a healthy student-teacher relationship (Kyrgiridis et al., 2014; Shahzad and Mehmood, 2019), by a conscious attitude with students' emotions (Kuzmanovic et al., 2012; Shahzad and Mehmood, 2019), and through the enthusiasm reflected by the teacher in and out of class (Rocha, 2013), among others. However, some highly experienced teachers do not do their job effectively, while other novice teachers can become more dynamic, innovative, and effective (Debets et al., 2020; Hoque et al., 2020; Musodza et al., 2020; Saeeda et al., 2021; Shin and Bolkan, 2021; Wulandari et al., 2021). In general, however, if you have a teaching team with significant experience, the team can provide a variety of benefits to their students (Podolsky et al., 2019).

The effectiveness of an educational system depends mainly on the effectiveness of the teaching staff, which in turn has a significant influence on student learning (Avalos, 1980). As a result, measuring teacher effectiveness is important in promoting educational quality and improving students' quality (Sánchez and Craig, 2007), which means that, in the education sector, evaluating teacher effectiveness is similar to assessing student learning (Ayaneh et al., 2021). Considering the role of teaching effectiveness, various investigations have analyzed

its importance in educational environments. Some researchers have analyzed its importance with a review of the literature (Yamamoto, 1963; Avalos, 1980; Reynolds, 1998; Podolsky et al., 2019; Sofyan et al., 2021; Bardach et al., 2022; Mastrokoukou et al., 2022) and others empirically. Because teaching tasks demand certain activities that occupy a large part of the time, their association with other constructs has been investigated. This includes their association to behavior management and support training (Monzalve-Macaya et al., 2023), the interaction between emotional intelligence (Anwar et al., 2021), habit formation (Hobbiss et al., 2021), school management practices and academic performance (Arop et al., 2020), neurolinguistic programming and teacher identity (Javadi and Asl, 2020), preparation of students to learn (Kearney and Garfield, 2019), personality and teaching support (Kim et al., 2019), structured collaboration (Graham, 2007), teacher preparation (Darling-Hammond et al., 2005); the role it plays in the influence of the minimum wage and the timely payment of wages (Adekanmbi and Ukpere, 2021), and on the impact of contextual distractors (Aslantas, 2020).

Teacher effectiveness is closely related to providing quality education, which is Sustainable Development Goals (SDGs) established by the United Nations in its 2030 Agenda (Bantekas, 2023; Pedraja-Rejas et al., 2023). It is known that these are a set of 17 integrated goals that seek to address global challenges, such as poverty, gender equality, health, education, environmental sustainability, and other areas (Ocaña-Zúñiga et al., 2023; Tomasella et al., 2023). Teaching effectiveness in higher education is associated in various ways with the SDGs (Miranda-Gonçalves, 2023; Bray, 2024; Rose and Sayed, 2024). For example, SDG 4 is based on ensuring inclusive, equitable, and quality education for all without exception. This translates into the development of high-performing teachers and educational institutions, aiming at the promotion and achievement of these objectives through its educational work and its impact on society (Leal et al., 2023; McCowan, 2023; Morris et al., 2023).

For many years now, various studies affirm that teaching effectiveness in higher education entails numerous benefits for students, educational institutions, and society. It is a crucial component to the success of higher education institutions and student development. Some of the key benefits include quality learning, motivation and engagement, developing critical skills, individualized feedback and support, improving retention and completion rates of academic programs, fostering diversity and inclusion, research and development, prestige and institutional reputation, and significant contributions to society through specialized areas or programs (Bridgwater, 1982; Kyriacou and Newson, 1982; Ngala and Odebero, 2010; Welsh, 2011; Gabriel and Allington, 2012; Darling-Hammond et al., 2013; Grant et al., 2013; Skourdoumbis, 2013; Joyce and Magesh, 2016). Therefore, teaching effectiveness in higher education positively impacts students, the institution, and society (Darza and Tesfaye, 2020; Latif et al., 2021).

According to the background mentioned, there is evident interest in developing scales to measure this construct, therefore, previous research has disclosed its contributions (Gusthart et al., 1997; Rocha, 2013; Mohebbi et al., 2022). Empirical studies on teacher effectiveness have been conducted in countries such as the United States, the United Kingdom, Australia, Cyprus, India, Iran, Nigeria, South Africa, Canada, and China. Therefore, this bibliometric review demonstrates the need to make greater efforts for its study and implementation in Latin America, since there is no contextualized metric in the scientific

literature with evidence of validity and reliability where teaching effectiveness is evaluated from the university student's perspective. To fill this knowledge gap, an instrumental study was considered appropriate to adapt the Student Evaluation of Teachers' Effectiveness (SETE) scale of Ethiopian origin, to be applicable to higher education students, given that the joint evaluation of the reliability and validity of measurement scales is classified as "psychometric properties," considered the most important characteristics for the evaluation of any scale; guaranteeing the quality and integrity of a measurement scale (Mohajan, 2017; Asiamah et al., 2021). In that sense, the present study aimed to translate, adapt, and evaluate the validity and reliability of the SETE scale in a sample of university students from Latin America.

2 Literature review

2.1 Teaching effectiveness

Some scholars have made efforts to distinguish the difference between "teaching effectiveness," "teacher's quality," and "teaching quality" (Al Ansari et al., 2020; Cai and Wang, 2022; Cherng et al., 2022). Teaching quality refers to a teacher's innate qualities, skills, and competencies (Chan, 2002), while when considering practices and instruction, emphasis is placed on the quality of teaching (Bradney, 1996; Hansen, 2023). Teaching effectiveness is analyzed in light of student outcomes, ensuring students learn and achieve specific outcomes (Sofyan et al., 2021). The latest studies on this topic have been used to offer new knowledge and concepts about performance evaluation, effectiveness (Al Kuwaiti et al., 2021), and the pedagogical skills of teachers in higher education (Hansen, 2023; Monzalve-Macaya et al., 2023; Pham et al., 2023). On the other hand, the review of the literature shows that various theoretical models measure this construct (Sánchez-Cabrero et al., 2021; Sofyan et al., 2021; Matosas-López, 2023); such as the one proposed by Rocha (2013) that evaluates teaching effectiveness from 4 approaches: teacher-student relationship, teacher's personality, student evaluation performed by the teacher, and his teaching method. His proposal is close to that of other specialists who in turn analyze the variable from 4 perspectives (Calaguas, 2012; Shahzad and Mehmood, 2019; Ayaneh et al., 2021). However, Aleamoni and Hexner (1980) suggest a 3-component theoretical model: Instructor evaluation, course-specific teaching procedure, and student perceptions of learning outcomes. Resembling other studies that propose the study of this construct from 3 factors (Marshall et al., 2016).

However, for this study, the four dimensions proposed by the Ministry of Science and Higher Education of Ethiopia (MOE) and the latest scientific contributions of Ayaneh et al. (2021), who consider a better theoretical model of four dimensions: subject matter knowledge (SK), professional competence (PC), ethical competence (EC), and time management (TM). Subject matter knowledge (SK) is the ability of the teacher to meet the content of the course, demonstrating preparation and using common examples to achieve the objectives of each session. Professional competence (PC) refers to the degree to which the teacher uses his or her knowledge, skills and good judgment related to professional skills to perform his or her main task with acceptable quality. Ethical competence (EC) is understood as the teaching competence that reconsiders promoting ethical education in classroom activities, conferences and through extracurricular activities. And time management (TM), known as the ability to make effective use of teaching time to efficiently fulfill their academic responsibilities.

2.2 Instruments to evaluate teaching effectiveness

Most of what has been reported on this construct has been of great contribution to improving the quality of education at all educational levels. In this sense, a review of previous research confirms the importance of providing valid instruments that can measure teaching efficacy in the context of higher education. These measurement instruments must comply with valid psychometric properties in order to be used in different realities. However, it is important to note that so far, teaching efficacy remains one of the most difficult constructs to measure. Various researchers have carried out studies on teaching effectiveness using different instruments (Adekanmbi and Ukpere, 2021; Anwar et al., 2021; Tadesse et al., 2021; Monzalve-Macaya et al., 2023). These studies can be separated into two groups. The first group of studies reviewed present reliability and validity.

In Ethiopia, a previous study presents the validity of the Student Evaluation of Teachers' Effectiveness (SETE) scale that was applied to a sample of university students, which presents a 2-factor model, 18 items and $\alpha = 0.79$, and a second 4-factor model, 20 items and $\alpha = 0.80$ (Ayaneh et al., 2021). In 2019, Shahzad and Mehmood designed the Teaching Effectiveness Scale (TES), which was applied to university students in Pakistan; it has 32 items and 4 dimensions ($\alpha = 0.71$ to 0.87). Marshall et al. (2016) developed and validated the Teacher Intentionality of Practice Scale (TIPS) in secondary school teachers in the United States; the scale confirmed 22 items and 3 dimensions ($\alpha = 0.96$). Kyrgiridis et al. (2014) developed the Self-Evaluation of Teacher Effectiveness in Physical Education (SETEQ-PE) questionnaire and applied it to Greek physical education teachers. This questionnaire has 25 items and 6 dimensions ($\alpha = 0.87$). Moreover, in Rocha (2013) study, he designed a Student Opinion about Teacher Effectiveness (SOTES) questionnaire and applied it to Mexican undergraduate students. It has 17 items and 4 dimensions ($\alpha = 0.947$). Calaguas (2012) developed and evaluated the psychometric properties of the Teacher Effectiveness Scale in Higher Education (TESHE) in university students in the Philippines; it has 67 items and 4 dimensions ($\alpha = 0.972$). In addition, the Student Evaluation of Educational Quality (SEEQ) presented 35 items discriminated in 8 dimensions ($\alpha = $ between 0.88 and 0.97): (1) Learning, (2) Enthusiasm, (3) Organization, (4) Interaction with the group, (5) Updated presentation of the subject, (6) Interaction of the teacher with the students individually, (7) Evaluation, (8) Feedback (Marsh, 1983; Marsh and Roche, 1997).

A second group was identified, which was made up of some studies that did not describe the validity or reliability of the instrument. This is the case with Mohebbi et al. (2022), who investigated the factors contributing to the effectiveness of language teachers in Iran and designed an instrument (EFL Language Teachers' Effectiveness) with 18 items and 6 dimensions. On the other hand, in the study by Nema et al. (2023) they used the Student Evaluations of Teachers (SET) as a measurement tool, which has been widely used for students in colleges and universities in India (Aleamoni and Hexner, 1980). The metric

consists of 33 items subdivided into 3 factors: instructor evaluation, teaching procedure in a specific course, and students' perceptions of learning outcomes. Based on the above, the purpose of this research is to translate, adapt and evaluate the validity and reliability of the SETE scale in a sample of Latin American university students. Several specialists argue that the SETE scale captures multiple aspects of a university teacher's good practices and that its periodic revision in different cultural contexts could help educational leaders to improve their educational and teaching management policies and strategies.

3 Materials and methods

3.1 Study design and participants

The work responds to psychometric research, defined as the science of evaluating the characteristics of tests designed to measure psychological attributes (Price, 2017). The study population was composed of university students from four South American countries, Peru, Bolivia, Chile, and Colombia, belonging to a private educational network. A condition to be part of the study was that the university students were studying the academic semester in person at the time of the survey. It should be noted that this study was approved by the Ethics Committee of the Postgraduate School of a private university (2023-CE-EPG-00071) and was conducted under the ethical standards of the Declaration of Helsinki (Manzini, 2000; Puri et al., 2009). The study was applied from September to December 2023. Non-probabilistic convenience sampling was applied (Otzen and Manterola, 2017), and the survey was carried out through a virtual link; the questionnaire was hosted on a Google form, and a Likert-type response format was utilized, ranging from (1) never to (5) very frequently. The questionnaire was self-administered, and informed consent was obtained from each participant before its administration ('I acknowledge that by completing this questionnaire, I am giving my consent to participate in the study'). The questionnaire was shared virtually (via Email, WhatsApp, Messenger, Instagram) and in a personalized way. The total sample was 1,000 university students, who provided their answers anonymously and voluntarily. Table 1 shows the frequencies and percentages by categories (sex, age range, marital status, country of origin, university campus, year of study, and religious inclination).

3.2 Instrument

The SETE scale is a harmonized instrument used to measure teacher effectiveness. Highly qualified experts originally developed it by the Ministry of Science and Higher Education (previously, Ministry of Education) of Ethiopia (MOE, 2018). Ayaneh et al. (2021) later evaluated its psychometric properties analyzing 2 models (4 and 2 factors), managing to confirm 2 factors (CFI = 0.999, TLI = 0.999, SRMR = 0.056; RMSEA = 0.008). The scale demonstrated good internal consistency (α = between 0.87 and 0.93) in a sample of Ethiopian university students.

3.3 Translation process

The original version of the SETE scale required translation from its original English language to Spanish using a bilingual trial

TABLE 1 Sociodemographic characteristics (n = 1,000).

Characteristic	Category	Frequency	Percentage (%)
Sex	Female	474	47.4
	Male	526	52.6
Age range	18–20 years	478	47.8
	21–30 years	508	50.8
	31–40 years	14	1.4
Marital status	Single	956	95.6
	Married	24	2.4
	Cohabitant	10	1.0
	Divorced	5	0.5
	Widowed	5	0.5
Country of origin	Peru	367	36.7
	Colombia	124	12.4
	Chile	245	24.5
	Brazil	2	0.2
	Bolivia	234	23.4
	Argentina	3	0.3
	Ecuador	8	0.8
	Venezuela	8	0.8
	Other	9	0.9
University campus	Peru	366	36.6
	Bolivia	250	25.0
	Chile	250	25.0
	Colombia	134	13.4
Year of study	First	351	35.1
	Second	237	23.7
	Third	168	16.8
	Forth	186	18.6
	Fifth	32	3.2
	Sixth	10	1.0
	Seventh	16	1.6
Religious inclination	Adventist	452	45.2
	Catholic	339	33.9
	Evangelical	58	5.8
	Other Christian denomination	49	4.9
	Not religious	102	10.2

back-translation method. Three bilingual (Spanish-English) Spanish-speaking individuals completed English-to-Spanish translations of the SETE scale individually. The translations were compared, discussed, and reviewed in a focus group of six university students (Peru, Bolivia, Chile, and Colombia) who met the study's inclusion criteria to obtain the first complete version in Spanish of the scale and its contextual application to the four selected South American countries. The English and Spanish versions of the SETE scale were tested on a target

group of bilingual individuals before some final changes were made and distributed to the study sample.

3.4 Data collection and analysis

Two statistical software were used to analyze the data: (1) SPSS software version 25 for descriptive analysis (sociodemographic profile of the participants, among others.) and Exploratory Factor Analysis (EFA), and (2) then Structural Equation Modeling of covariance (CB-SEM) to perform Confirmatory Factor Analysis (CFA), evaluate convergent and discriminant validity, and adjust the measurement model. This required AMOS version 24 software. This method is highly recommended to evaluate the psychometric properties of measurement models (Fornell and Larcker, 1981). Likewise, reliability was evaluated using Cronbach's Alpha coefficient and composite reliability.

4 Results

4.1 Content validity

From the validation of Ayaneh et al. (2021), the proposal of 20 items was taken as a basis, where a group of educational specialists evaluated the saturated items. The scale was organized according to the content validity (Table 2) process by expert judgment (30 items), structured in 4 dimensions (SK = subject knowledge, PC = professional competence, EC = ethical competence, and TM = time management). For this analysis, six experts in university higher education with a minimum of 10 years of experience were recruited to serve as judges. At the expert committee's suggestion, 2 items were removed (16 and 19). To analyze the results, Aiken's V Coefficient (V > 0.50; 95% CI) was used, considering the criterion value for deciding which items should be eliminated, revised, or withdrawn. Finally, the scale was left with 28 items to be included in the virtual questionnaire. All the items

TABLE 2 Content validity in the SETE scale.

Measurement items	V for Aiken			CI 95%		
	Clarity	Pertinent	Relevance	Clarity	Pertinent	Relevance
SK1	0.88	0.96	0.96	[0.69–0.96]	[0.80–0.99]	[0.80–0.99]
SK2	0.96	0.88	0.88	[0.80–0.99]	[0.69–0.96]	[0.69–0.96]
SK3	0.96	0.96	0.96	[0.80–0.99]	[0.80–0.99]	[0.80–0.99]
SK4	0.88	0.88	0.92	[0.69–0.96]	[0.69–0.96]	[0.74–0.98]
SK5	1	1	1	[0.86–1]	[0.86–1]	[0.86–1]
SK6	0.92	0.96	0.92	[0.74–0.98]	[0.80–0.99]	[0.74–0.98]
PC1	0.88	0.88	0.96	[0.69–0.96]	[0.69–0.96]	[0.80–0.99]
PC2	0.96	0.92	0.92	[0.80–0.99]	[0.74–0.98]	[0.74–0.98]
PC3	0.92	0.96	0.96	[0.74–0.98]	[0.80–0.99]	[0.80–0.99]
PC4	0.96	0.92	0.96	[0.80–0.99]	[0.74–0.98]	[0.80–0.99]
PC5	0.96	0.96	0.96	[0.80–0.99]	[0.80–0.99]	[0.80–0.99]
PC6	0.92	0.92	0.96	[0.74–0.98]	[0.74–0.98]	[0.80–0.99]
PC7	0.96	0.96	0.92	[0.80–0.99]	[0.80–0.99]	[0.74–0.98]
PC8	1	1	1	[0.86–1]	[0.86–1]	[0.86–1]
PC9	1	1	1	[0.86–1]	[0.86–1]	[0.86–1]
PC10	1	1	1	[0.86–1]	[0.86–1]	[0.86–1]
PC11	1	1	1	[0.86–1]	[0.86–1]	[0.86–1]
PC12	1	1	1	[0.86–1]	[0.86–1]	[0.86–1]
PC13	0.92	1	1	[0.74–0.98]	[0.86–1]	[0.86–1]
PC14	0.96	0.96	1	[0.80–0.99]	[0.80–0.99]	[0.86–1]
EC1	0.88	0.96	0.96	[0.69–0.96]	[0.80–0.99]	[0.80–0.99]
EC2	1	1	1	[0.86–1]	[0.86–1]	[0.86–1]
EC3	1	1	1	[0.86–1]	[0.86–1]	[0.86–1]
EC4	0.92	0.96	0.96	[0.74–0.98]	[0.80–0.99]	[0.80–0.99]
TM1	0.96	1	1	[0.80–0.99]	[0.86–1]	[0.86–1]
TM2	0.92	0.92	0.92	[0.74–0.98]	[0.74–0.98]	[0.74–0.98]
TM3	0.92	0.92	0.92	[0.74–0.98]	[0.74–0.98]	[0.74–0.98]
TM4	1	1	0.96	[0.86–1]	[0.86–1]	[0.80–0.99]

were evaluated regarding clarity, relevance, and relevance to the construct (Table 2).

On the other hand, indicators such as mean, standard deviation, skewness, and kurtosis were obtained for each of the items applied in this study. The skewness and kurtosis results are nearly zero, meaning the distributions are symmetrical. Furthermore, variability is denoted in the mean, which shows the diversity of responses regarding perceptions among the study participants, as shown in Table 3.

4.2 Exploratory factor analysis

To identify the factorial condition of the scale, an Exploratory Factor Analysis (EFA) was carried out on each element, observing that the items were distributed into four factors according to the construct analyzed (Table 4). The difference is quite clear between the four factors. The KMO and Bartlett test (Kaiser-Meyer-Olkin correlation coefficient = 0.975) has a value greater than 0.7 and the Bartlett test (Sig = 0.000) is very significant for performing factor analysis. The total variance explained in the model is 73.66%, which is greater than 50%, with Subject Knowledge (SK) = 57.15%, Professional Competence (PC) = 7.59%, Ethical Competence (EC) = 5.21%, and Time Management (TM) = 3.71%. All items have been grouped according to their original dimensions. Next, Confirmatory Factor Analysis (CFA) was performed.

The validation of the final measurement model is shown in Table 5 along with the convergent reliability and validity. Cronbach's Alpha (α) values range between 0.927 and 0.961, considered satisfactory values since all levels of this coefficient must be above 0.70 for the model to be valid (Agbo, 2010). Furthermore, the reliability values (CR) were found between 0.927 and 0.962, which is favorable because this value must be greater than 0.70 to be considered a perfect model (Bagozzi and Yi, 1988). Likewise, the AVE values are between 0.646 and 0.799, which are considered acceptable since this index must be equal to or greater than 0.50 (Hair et al., 2014). In that sense, these values translate as an acceptable measurement model that meets favorable levels of reliability and convergent validity.

TABLE 3 Descriptive analysis of the items (*n* = 1,000).

Code	Mean ± Standard Deviation	Skewness	Kurtosis
SK1	3.9490 ± 0.95567	−0.870	0.543
SK2	4.0460 ± 0.88582	−0.878	0.661
SK3	3.9970 ± 0.90764	−0.775	0.285
SK4	3.9380 ± 0.98241	−0.764	0.051
SK5	3.9350 ± 0.94640	−0.829	0.468
SK6	3.9700 ± 0.93379	−0.857	0.540
PC1	3.6740 ± 1.06624	−0.553	−0.339
PC2	4.0240 ± 0.95201	−0.892	0.380
PC3	3.9680 ± 0.96535	−0.832	0.264
PC4	4.0310 ± 0.96279	−0.945	0.542
PC5	3.8590 ± 1.01101	−0.781	0.181
PC6	3.9700 ± 0.96230	−0.804	0.255
PC7	4.0290 ± 0.97113	−0.893	0.239
PC8	4.0110 ± 0.93689	−0.834	0.325
PC9	3.9270 ± 0.94476	−0.725	0.078
PC10	3.7530 ± 1.10146	−0.666	−0.277
PC11	3.8020 ± 1.05447	−0.743	0.017
PC12	3.8290 ± 0.99436	−0.650	−0.081
PC13	3.9630 ± 0.96673	−0.859	0.363
PC14	4.0080 ± 0.95438	−0.854	0.294
EC1	4.1600 ± 0.97845	−1,128	0.771
EC2	4.1410 ± 0.96023	−1,026	0.499
EC3	4.1610 ± 0.96748	−1,097	0.620
EC4	4.1360 ± 1.03758	−1,228	0.997
TM1	3.9610 ± 0.97387	−0.795	0.200
TM2	3.9910 ± 0.98686	−0.840	0.146
TM3	3.8630 ± 1.05610	−0.726	−0.104
TM4	3.9370 ± 1.00898	−0.811	0.135

Table 6 shows each indicator of model fit that measures teaching effectiveness, showing acceptable and excellent measures.

Regarding discriminant validity (Table 7), the results show that the confidence intervals, in none of the cases, reach unity. Additionally, the quantile covariances do not exceed the AVE; therefore, there is clear evidence of discrimination between the constructs subjected to evaluation (Fornell and Larcker, 1981).

To provide further strength in evaluating discriminant validity as a requirement for analyzing the relationships between the latent variables, Table 8 presents the results regarding discriminant validity using the heterotoit-monotrait relationship criterion (HTMT); in this case, the values are less than 0.90. Therefore, it is stated that the discriminant validity between two reflective constructs has been established (Henseler et al., 2015). Additionally, Figure 1 shows the factor structure of the SETE scale in a sample of university students from Latin America.

Finally, the final version of the instrument, which underwent rigorous content validity, EFA, and CFA processes to ensure reliable psychometric properties for use, is described (Table 9). It is made up of four factors: 06 items for Subject matter Knowledge (SK), 14 items for Professional Competency (PC), 04 items for Ethical Competence (EC), and 04 items for Time Management (TM).

5 Discussions

5.1 Discussion of findings

The objective of the present study was to evaluate the validity and reliability of the SETE scale (Ayaneh et al., 2021) in the Latin American context. This is the first study in which the evidence and reliability of this scale are published in this context. Other validations of the same construct

TABLE 4 Exploratory factor analysis (EFA) pattern matrix.

	Factor			
	1	2	3	4
PC13	0.878			
PC9	0.825			
PC7	0.818			
PC6	0.815			
PC8	0.812			
PC14	0.788			
PC12	0.784			
PC4	0.782			
PC11	0,773			
PC3	0.758			
PC10	0.727			
PC2	0.715			
PC5	0.706			
PC1	0.513			
SK4		0.838		
SK5		0.825		
SK6		0.802		
SK3		0.743		
SK1		0.742		
SK2		0.734		
EC2			0.918	
EC3			0.914	
EC1			0.889	
EC4			0.763	
TM3				0.875
TM4				0.852
TM2				0.758
TM1				0.720

Extraction method: maximum authenticity. Rotation method: Promax with Kaiser normalization. ªThe rotation has converged in 7 iterations.

TABLE 5 Validation of the final measurement model with reliability and convergent validity.

Predictor	Items	Estimate	Alpha	CR	AVE
SK	SK1	0.813 ***	0.928	0.928	0.682
	SK2	0.817 ***			
	SK3	0.803 ***			
	SK4	0.808***			
	SK5	0.851***			
	SK6	0.861 ***			
PC	PC1	0.732***	0.961	0.962	0.646
	PC2	0.830***			
	PC3	0.845 ***			
	PC4	0.877 ***			
	PC5	0.637***			
	PC6	0.802***			
	PC7	0.817 ***			
	PC8	0.842 ***			
	PC9	0.810 ***			
	PC10	0.703 ***			
	PC11	0.775 ***			
	PC12	0.818 ***			
	PC13	0.854 ***			
	PC14	0.869 ***			
EC	EC1	0.918 ***	0.938	0.940	0.799
	EC2	0.927 ***			
	EC3	0.933***			
	EC4	0.788 ***			
TM	TM1	0.847 ***	0.927	0.927	0.762
	TM2	0.898***			
	TM3	0.855 ***			
	TM4	0.890***			

Cronbach's alpha (α) for all variables is > 0.9, the composite reliability (CR) > 0.90, and the mean–variance extracted (AVE) > 0.60; ***p < 0.001 (significance level), indicating a significant validity of the model.

TABLE 6 Statistical goodness-of-fit indices of the SETE scale.

Measure	Threshold	Estimate	Interpretation
CMIN	–	1,499,529	–
DF	–	344	–
CMIN/DF	Between 1 and 3	4,359	Acceptable
CFI	>0.95	0.956	Excellent
SRMR	<0.08	0.030	Excellent
RMSEA	<0.06	0.058	Excellent

CMIN, Chi-square; DF, Degrees of Freedom; SRMR, Standardized Root Means square Residual; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index.

TABLE 7 Validation of the discriminant validity of the measurement model (Fornell-Lacker Criteria).

	CR	AVE	SK	PC	EC	TM
SK	0.928	0.682	**0.826**			
PC	0.962	0.646	0.786***	**0.803**		
EC	0.940	0.799	0.657***	0.672***	**0.894**	
TM	0.927	0.762	0.705***	0.710***	0.750***	**0.873**

***$p < 0.001$ (significance level). The square root of AVEs is shown diagonally in bold.

TABLE 8 Discriminant validity of the model using the heterotroit-monotrait (HTMT) relationship criterion.

	SK	PC	EC	TM
SK				
PC	0.787			
EC	0.653	0.664		
TM	0.702	0.705	0.753	

have been carried out in other latitudes (Calaguas, 2012; Rocha, 2013; Kyrgiridis et al., 2014; Marshall et al., 2016; Shahzad and Mehmood, 2019; Adekanmbi and Ukpere, 2021; Mohebbi et al., 2022). The scales review found a diversity of factors associated with this construct, ranging from single-factor scales to scales with 10 factors (Shahzad and Mehmood, 2019; Li et al., 2024). The dimensions presented in these scales share characteristics with others that measure the same construct, although to date there is no consensus on their dimensionality.

Originally the SETE scale was validated in university students from Ethiopia with 20 items, however, in this new version for Latin America the items went through a rigorous adaptation process, resulting in 30 items. This is because the original scale items were found to address more than one specific topic, saturating them. To take an example, the factor "Professional Competence" (PC), originally had the item "Follows the continuous evaluation approach and gives feedback on continuous evaluations on time." This item, when going through the process of semantic validation by expert judgment, it was considered that the item structure should be reformulated to make it clearer and more specific (PC9: Permanently evaluates students, PC10: Resolves tests when an evaluation is concluded). Based on the opinion of psychometric specialists who argue that by analyzing, eliminating or reformulating saturated items, the validity and reliability of the scale can be improved, ensuring that the instrument measures what it is intended to measure accurately and consistently (Matas, 2018). This explains why the 4-factor model lacked discriminant validity and

probably better explains why the CFA was deficient. The adaptation carried out in the present study showed that the factors SK = Subject knowledge, PC = Professional competence, EC = Ethical competence, and TM = Time management, fit the model. Although two items (items 16 and 19) had to be eliminated from the initial 30, the results demonstrated good internal consistency (α = between 0.927 and 0.961; CR = between 0.927 and 0.962; AVE = between 0.646 and 0.799). The final version of the scale consists of 28 items.

5.2 Theoretical and practical implications

The SETE scale is now ready to be used as a valid and reliable tool to evaluate the effectiveness of teachers in the university teaching process. Teachers can use this measurement tool to understand what students perceive as their strengths and weaknesses. College students can also use this information to make more informed course selections. This can be applied in future research to find correlations, predictors, and moderators of this construct. Institutions can also use it to make decisions about their teachers. Although the SETE scale has good psychometric properties, it is recommended that its revision be reconsidered in the future. It needs to be revised and updated periodically to meet the rapidly changing needs of university students, teachers, and educational institutions. In addition, these results can be the platform for the creation of innovative policies, strategies and educational programs that increase university students' perception of teaching effectiveness, thus contributing to an increase in the quality of education in Latin American countries.

5.3 Limitations and future research

The results of this study should be considered taking into account the following limitations: One limitation is that, although the scale was adapted and harmonized to be applied to all private universities in the Latin American countries of South America (whose languages derive from Latin, mainly Spanish and Portuguese), this analysis only used data from some countries (4/10), which may not be generalizable to other private universities in South America. In this sense, this study highlights the need to obtain a large amount of data from multiple universities and representative samples from each South American country to strengthen existing findings further. On the other hand, the questionnaire assumed that students' evaluations of teachers were free of prejudices or stereotypes. However, university participants with high grades in their subjects are likely to give higher scores on the questionnaire. In contrast, participants who obtained low grades in their subjects consider this evaluation a form of retaliation against their teachers.

In addition, the physical attractiveness of the teacher, the time taken to complete the survey, the proximity to a teacher, the difficulty of the course, and the teacher's age, performance, and personality can become determining factors in students' evaluations (Calaguas, 2012; Kim et al., 2019; Shahzad and Mehmood, 2019; Sofyan et al., 2021). It is also recommended to consider the semester for the application of the questionnaire, given that taking it in the first (students are not entirely familiar with the teacher's methods) or last (students with failing grades, in a spirit of retaliation, among others) weeks of class can create study bias. In addition, in future studies, it is recommended that some indication or filter be included in the questionnaire to determine whether the subjects considered are all the subjects of the academic period or only the compulsory ones, since this could influence the evaluation of the students.
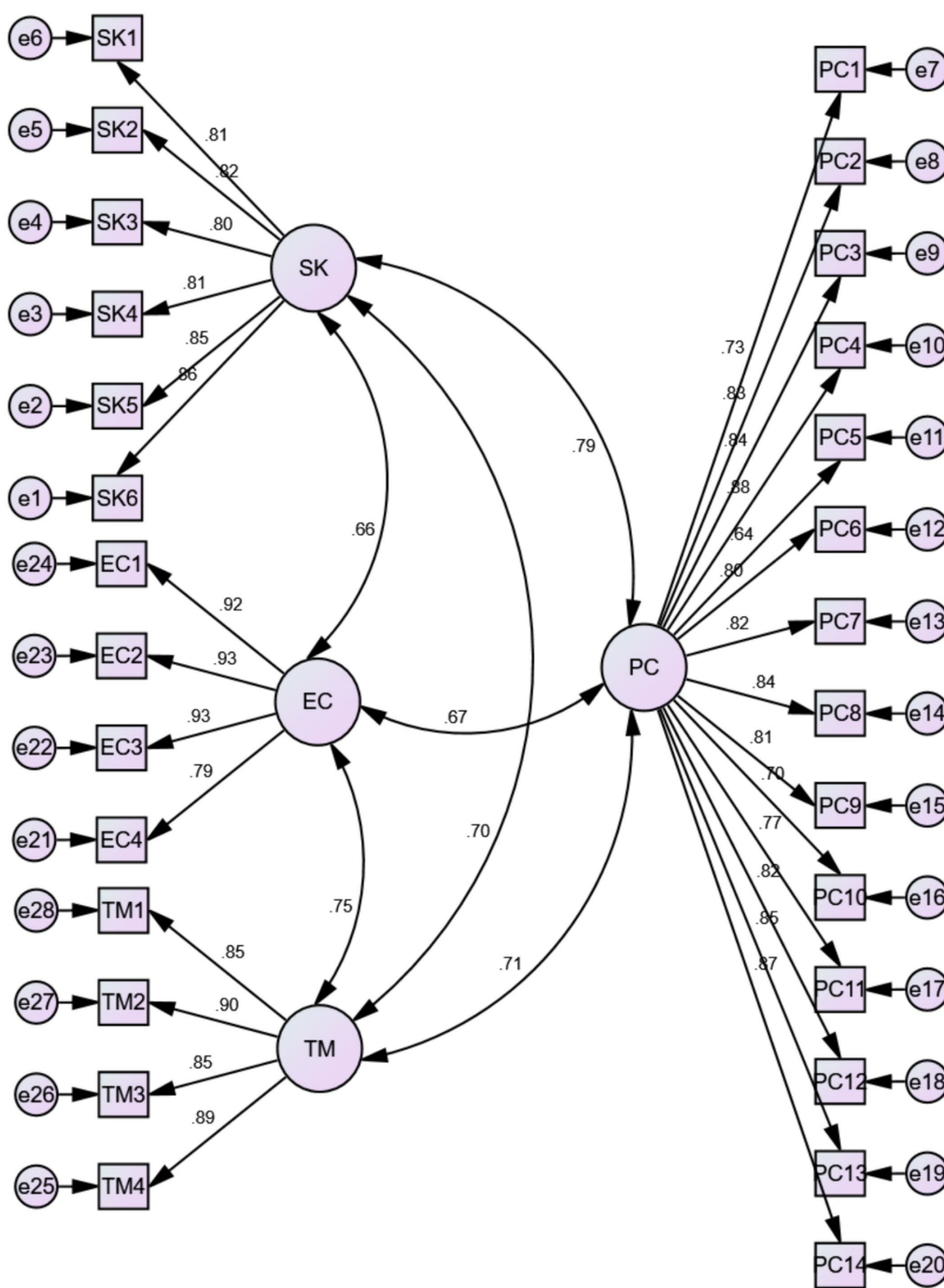
FIGURE 1
Factor structure of the SETE scale.

Finally, the "Professional Competence" (PC) factor was validated with 14 items, the broadest dimension of the scale, which may be another limitation of this study. However, despite these limitations, the study's findings will significantly help university education professionals address teacher effectiveness problems.

6 Conclusion

The SETE scale's translation, adaptation, analysis of validity, and reliability were performed on a sample of 1,000 South American university students of both sexes. After having passed through a diligent process of content validity, EFA, and CFA, the results presented reliable psychometric properties for the application. The scale confirmed 28 items distributed in 4 factors (SK, PC, EC, and TM). In terms of convergent and discriminant validity, the factors showed acceptable values. In addition, the internal consistency of the scores was determined by observing adequate values for Cronbach's Alpha coefficient (α = between 0.927 and 0.961). In this sense, the SETE scale is considered a scientific tool with accessible language for practical, useful, reliable, and necessary applications.

TABLE 9 28-item instrument (Spanish version).

Predictor	Measurement items	Affirmations
Subject matter Knowledge (SK)	SK1	Explica los objetivos y las unidades del curso/materia a tiempo.
	SK2	Demuestra dominio del curso/materia.
	SK3	Proporciona apuntes y material de lectura del curso/materia.
	SK4	Considera en el sílabo del curso, libros disponibles en la biblioteca y direcciones web accesibles.
	SK5	Enseña en función de la naturaleza del curso/materia e imparte sesiones prácticas.
	SK6	Imparte el curso/materia de tal manera que los estudiantes lo entienden.
Professional Competency (PC)	PC1	Utiliza materiales didácticos adicionales.
	PC2	Responde a las preguntas planteadas en el salón de clase.
	PC3	Asigna trabajos para desarrollar en el salón de clase.
	PC4	Plantea preguntas en el salón de clase.
	PC5	Deja tareas para la casa.
	PC6	Los alumnos exponen como parte del desarrollo de la clase.
	PC7	Los alumnos trabajan en grupos.
	PC8	Prepara exámenes según el contenido del curso/materia, incluyendo varios modos de evaluación.
	PC9	Evalúa permanentemente a los estudiantes.
	PC10	Resuelve los exámenes cuando concluye una evaluación.
	PC11	Brinda orientación a los estudiantes, sobre todo a los que tienen necesidades especiales y de bajo rendimiento.
	PC12	Da retroalimentación.
	PC13	Permite que los estudiantes interactúen en determinados momentos de la clase.
	PC14	Demuestra compromiso para la transferencia del conocimiento.
Ethical Competence (EC)	EC1	Respeta a los estudiantes.
	EC2	Demuestra una conducta ética.
	EC3	Demuestra buen comportamiento.
	EC4	No discrimina por motivos étnicos, religiosos o de sexo.
Time Management (TM)	TM1	Llega a tiempo durante el horario de clase.
	TM2	Usa el tiempo de clase apropiadamente.
	TM3	Establece horarios de atención a los estudiantes.
	TM4	Resuelve a tiempo los problemas académicos de los estudiantes.

Data availability statement

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

Ethics statement

The studies involving humans were approved by Comité de Ética de la Escuela de Posgrado de la Universidad Peruana Unión, Lima, Perú. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MV-G: Conceptualization, Investigation, Project administration, Resources, Visualization, Writing – original draft, Writing – review &

editing. JHL-S: Methodology, Resources, Supervision, Writing – review & editing. EEG-S: Data curation, Formal analysis, Software, Supervision, Validation, Writing – review & editing. IF-M: Investigation, Visualization, Writing – review & editing.

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.

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

Adekanmbi, F. P., and Ukpere, W. (2021). Influence of minimum wage and prompt salary payment on teachers' effectiveness in public secondary schools. *Probl. Perspect. Manag.* 19, 116–126. doi: 10.21511/ppm.19(1).2021.10

Agbo, A. A. (2010). Cronbach's alpha: review of limitations and associated recommendations. *J. Psychol. Afr.* 20, 233–239. doi: 10.1080/14330237.2010.10820371

Al Ansari, A., Strachan, K., Al Balooshi, S., Al-Qallaf, A., and Otoom, S. (2020). Influence of student feedback on the quality of teaching among clinical teachers in Bahrain. *Med Sci Educ* 30, 253–262. doi: 10.1007/s40670-019-00892-1

Al Kuwaiti, A., Subbarayalu, A. V., Kamalasanan, A., Prabaharan, S., and Raman, V. (2021). A critical appraisal of students evaluating teaching effectiveness to manage the quality of higher education institutions. *J. Legal, Ethical and Regulatory Issues* 24, 1–11.

Aleamoni, L. M., and Hexner, P. Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instr. Sci.* 9, 67–84. doi: 10.1007/BF00118969

Anwar, R. H., Zaki, S., Memon, N., and Thurasamy, R. (2021). Exploring the interplay of trait emotional intelligence and ESL teacher effectiveness: is self-efficacy the mechanism linking them? *SAGE Open* 11, 215824402110613–215824402110619. doi: 10.1177/21582440211061378

Arop, F. O., Mbon, U. F., Ekanem, E. E., Ukpabio, G. E., Uko, E. S., and Okon, J. E. (2020). School management practices, teachers effectiveness, and students' academic performance in mathematics in secondary schools of cross river state, Nigeria. *Humanities Soc. Sci. Letters* 8, 298–309. doi: 10.18488/JOURNAL.73.2020.83.298.309

Asiamah, N., Kouveliotis, K., Eduafo, R., and Borkey, R. (2021). Psychometric properties of a new scale measuring neglect and abuse of older adults in the community: implications for social activity. *Int. Q. Community Health Educ.* 41, 163–172. doi: 10.1177/0272684X20915384

Aslantas, I. (2020). Impact of contextual predictors on value-added teacher effectiveness estimates. *Educ Sci (Basel)* 10, 1–16. doi: 10.3390/educsci10120390

Avalos, B. (1980). Teacher effectiveness: research in the third world—highlights of a review. *Comp. Educ.* 16, 45–54. doi: 10.1080/0305006800160106

Ayaneh, M. G., Dessie, A. A., and Fetene, D. M. (2021). Psychometric properties of student evaluation of teachers' performance scale: evidence from Debre Markos university students' evaluation dataset. *Educ Res Int* 2021, 1–15. doi: 10.1155/2021/5543317

Bagozzi, R. R., and Yi, Y. (1988). On the evaluation of structural equation models. *J. Acad. Mark. Sci.* 16, 74–94. doi: 10.1007/BF02723327

Bantekas, I. (2023). Aligning the right to education with the sustainable development goals: is it just about finance? *Hum. Rights Q.* 45, 62–87. doi: 10.1353/hrq.2023.0002

Bardach, L., and Klassen, R. M. (2020). Smart teachers, successful students? A systematic review of the literature on teachers' cognitive abilities and teacher effectiveness. *Educ. Res. Rev.* 30, 100312–100321. doi: 10.1016/j.edurev.2020.100312

Bardach, L., Klassen, R. M., and Perry, N. E. (2022). Teachers' psychological characteristics: do they matter for teacher effectiveness, teachers' well-being, retention, and interpersonal relations? An integrative review. *Educ. Psychol. Rev.* 34, 259–300. doi: 10.1007/s10648-021-09614-9

Bradney, A. (1996). The quality of teaching quality assessment in English law schools. *Law Teacher* 30, 150–167. doi: 10.1080/03069400.1996.9992948

Bray, M. (2024). Still in need of confronting: shadow education and its implications in the sustainable development goals. *Int. J. Educ. Dev.* 104, 102967–102965. doi: 10.1016/j.ijedudev.2023.102967

Bridgwater, C. A. (1982). Personality characteristics of ski instructors and predicting teacher effectiveness using the PRF. *J. Pers. Assess.* 46, 164–168. doi: 10.1207/s15327752jpa4602_11

Cai, B., and Wang, D. (2022). Teaching quality of college students' mental health based on mathematical programming algorithm. *Math. Probl. Eng.* 2022, 1–11. doi: 10.1155/2022/6384369

Calaguas, G. M. (2012). Teacher effectiveness scale in higher education: development and psychometric properties. *Int. J. Res. Stud. Educ.* 2, 1–18. doi: 10.5861/ijrse.2012.108

Chan, B. (2002). A study of the relationship between Tutor's personality and teaching effectiveness: does culture make a difference? *Int. Rev. Res. Open Dist. Learn.* 3, 47–67. doi: 10.19173/irrodl.v3i2.110

Cherng, H. Y. S., Halpin, P. F., and Rodriguez, L. A. (2022). Teaching Bias? Relations between teaching quality and classroom demographic composition. *Am. J. Educ.* 128, 171–201. doi: 10.1086/717676

Coady, M., Miller, M. D., Jing, Z., Heffington, D., Lopez, M., Olszewska, A., et al. (2020). Can English learner teacher effectiveness be observed? Validation of an EL-modified framework for teaching. *TESOL Q.* 54, 173–200. doi: 10.1002/tesq.544

Danielsons, C. (2007). Enhancing professional practice: A framework for teaching. *2nd* Edn. Alexandria, VA: ASCD.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., and Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, teach for America, and teacher effectiveness. *Educ Policy Anal Arch* 13, 1–51. doi: 10.14507/epaa.v13n42.2005

Darling-Hammond, L., Newton, S. P., and Wei, R. C. (2013). Developing and assessing beginning teacher effectiveness: the potential of performance assessments. *Educ. Assess. Eval. Account.* 25, 179–204. doi: 10.1007/s11092-013-9163-0

Darza, D., and Tesfaye, M. (2020). Evaluation of teachers' competency in higher education: the case of evaluation by students in Arba Minch University, Ethiopia. *European J. Educ. Stud.* 7, 161–178. doi: 10.5281/zenodo.3817503

Debets, M. P. M., Scheepers, R. A., Boerebach, B. C. M., Arah, O. A., and Lombarts, K. M. J. M. H. (2020). Variability of residents' ratings of faculty's teaching performance measured by five-and seven-point response scales. *BMC Med. Educ.* 20, 1–9. doi: 10.1186/s12909-020-02244-9

dos Santos, J., de Souza, C., Tourinho, F., Sebold, L., Kempfer, S., and Linch, G. (2018). Didactic strategies in the teaching-learning process of nursing management. *Texto e Contexto Enfermagem* 27, 1–11. doi: 10.1590/0104-070720180001980016

Faranda, W. T., and Clarke, I. (2004). Student observations of outstanding teaching: implications for marketing educators. *J. Mark. Educ.* 26, 271–281. doi: 10.1177/0273475304268782

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

Gabriel, R., and Allington, R. (2012). Constructing and measuring teacher effectiveness in global education reform. *Int. Perspect. Educ. Soc.* 16, 203–228. doi: 10.1108/S1479-3679(2012)0000016014

Galmes-Panades, A. M., Palou-Sampol, P., Borras, P. A., and Muntaner-Mas, A. (2021). Reliability and feasibility of an evaluation tool of teacher effectiveness and the quality of physical education sessions. *J. Human Sport Exercise* 18, 166–179. doi: 10.14198/JHSE.2023.181.14

Graham, P. (2007). Improving teacher effectiveness through structured collaboration: a case study of a professional learning community. *RMLE Online* 31, 1–17. doi: 10.1080/19404476.2007.11462044

Grant, L. W., Stronge, J. H., and Xu, X. (2013). A cross-cultural comparative study of teacher effectiveness: analyses of award-winning teachers in the United States and China. *Educ. Assess. Eval. Account.* 25, 251–276. doi: 10.1007/s11092-013-9170-1

Gusthart, J. L., Kelly, I. M., and Rink, J. E. (1997). The validity of the qualitative measures of teaching performance scale as a measure of teacher effectiveness. *J. Teach. Phys. Educ.* 16, 196–210. doi: 10.1123/jtpe.16.2.196

Hair, J. F., Sarstedt, M., Hopkins, L., and Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): an emerging tool in business research. *Eur. Bus. Rev.* 26, 106–121. doi: 10.1108/EBR-10-2013-0128

Hansen, T. I. (2023). Quality of teaching quality: contribution to a phenomenologically based concept of quality in teaching. *Acta Didactica Norden* 17, 1–19. doi: 10.5617/adno.9113

Harrison, J. (1987). A review of the research on teacher effectiveness and its implications for current practice. *Quest* 39, 36–55. doi: 10.1080/00336297.1987.10483855

Henseler, J., Ringle, C. M., and Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* 43, 115–135. doi: 10.1007/s11747-014-0403-8

Hobbiss, M., Sims, S., and Allen, R. (2021). Context and implications document for: habit formation limits growth in teacher effectiveness: a review of converging evidence from neuroscience and social science. *Rev. Educ.* 9, 24–26. doi: 10.1002/rev3.3225

Holland, E. P. (2019). Making sense of module feedback: accounting for individual behaviours in student evaluations of teaching. *Assess. Eval. High. Educ.* 44, 961–972. doi: 10.1080/02602938.2018.1556777

Hoque, K. E., Bt Kenayathulla, H. B., D/O Subramaniam, M. V., and Islam, R. (2020). Relationships between supervision and teachers' performance and attitude in secondary schools in Malaysia. *SAGE Open* 10, 215824402092550–215824402092511. doi: 10.1177/2158244020925501

Javadi, Y., and Asl, S. A. (2020). Neuro-linguistic programming, teacher's identity, and teachers' effectiveness. *J. Lang. Teach. Res.* 11, 389–398. doi: 10.17507/jltr.1103.07

Joyce, S., and Magesh, R. (2016). Emotional intelligence and teacher effectiveness of arts and science college - an empirical study in Chennai. *Int J Econ Res* 13, 1323–1328.

Kearney, W. S., and Garfield, T. (2019). Student readiness to learn and teacher effectiveness: two key factors in middle grades mathematics achievement. *RMLE Online* 42, 1–12. doi: 10.1080/19404476.2019.1607138

Kim, L. E., Jörg, V., and Klassen, R. M. (2019). A Meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. *Educ. Psychol. Rev.* 31, 163–195. doi: 10.1007/s10648-018-9458-2

Knol, M. H., Dolan, C. V., Mellenbergh, G. J., and Van der Maas, H. L. J. (2016). Measuring the quality of university lectures: development and validation of the instructional skills questionnaire (ISQ). *PLoS One* 11, 1–21. doi: 10.1371/journal.pone.0149163

Kuzmanovic, M., Savic, G., Popovic, M., and Martic, M. (2012). A new approach to evaluation of university teaching considering heterogeneity of students' preferences. *Procedia Soc. Behav. Sci.* 64, 402–411. doi: 10.1016/j.sbspro.2012.11.047

Kyrgiridis, P., Derri, V., Emmanouilidou, K., Chlapoutaki, E., and Kioumourtzoglou, E. (2014). Development of a questionnaire for self-evaluation of teacher effectiveness in physical education (SETEQ-PE). *Meas. Phys. Educ. Exerc. Sci.* 18, 73–90. doi: 10.1080/1091367X.2013.866557

Kyriacou, C., and Newson, G. (1982). Teacher effectiveness: a consideration of research problems. *Educ Rev (Birm)* 34, 3–12. doi: 10.1080/0013191820340101

Latif, K., Machuca, M., Marimon, F., and Sahibzada, U. (2021). Servant leadership, career, and life satisfaction in higher education: a cross-country study of Spain, China, and Pakistan. *Appl. Res. Qual. Life* 16, 1221–1247. doi: 10.1007/s11482-019-09809-x

Leal, W., Salvia, A. L., and Eustachio, J. H. P. P. (2023). An overview of the engagement of higher education institutions in the implementation of the UN sustainable development goals. *J. Clean. Prod.* 386, 135694–135699. doi: 10.1016/j.jclepro.2022.135694

Li, Y., Lin, M., Lin, X., Chen, Q., Wang, F., Chen, W., et al. (2024). Development and psychometric testing of the narrative nursing teaching effectiveness scale: a methodological study. *Nurse Educ. Today* 133, 106060–106151. doi: 10.1016/j.nedt.2023.106060

Manzini, J. L. (2000). Declaración de Helsinki: Principios éticos para la investigación médica sobre sujetos humanos. *Acta Bioeth* 6, 321–334. doi: 10.4067/S1726-569X2000000200010

Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *J. Educ. Psychol.* 75, 150–166. doi: 10.1037/0022-0663.75.1.150

Marsh, H. W., and Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective the critical issues of validity, Bias, and utility. *Am. Psychol.* 52, 1187–1197. doi: 10.1037/0003-066X.52.11.1187

Marshall, J. C., Smart, J., and Alston, D. M. (2016). Development and validation of teacher intentionality of practice scale (TIPS): a measure to evaluate and scaffold teacher effectiveness. *Teach. Teach. Educ.* 59, 159–168. doi: 10.1016/j.tate.2016.05.007

Martínez-Huamán, E. L., Landa, J., Mendoza, V., and Plata, C. (2022). Formative assessment: component of teaching-learning process in the university context in post-COVID-19. *Int. J. Learn., Teach. Educ. Res.* 21, 300–314. doi: 10.26803/ijlter.21.11.17

Mastrokoukou, S., Kaliris, A., Donche, V., Chauliac, M., Karagiannopoulou, E., Christodoulides, P., et al. (2022). Rediscovering teaching in university: a scoping review of teacher effectiveness in higher education. *Front Educ (Lausanne)* 7, 1–17. doi: 10.3389/feduc.2022.861458

Matas, A. (2018). Diseño del formato de escalas tipo Likert: Un estado de la cuestión. *Revista Electronica de Investigacion Educativa* 20, 38–47. doi: 10.24320/redie.2018.20.1.1347

Matosas-López, L. (2023). Measuring teaching effectiveness with behavioral scales: a systematic literature review. *Int. J. Educ. Organiz. Leadership* 30, 43–58. doi: 10.18848/2329-1656/CGP

McCowan, T. (2023). The crosscutting impact of higher education on the sustainable development goals. *Int. J. Educ. Dev.* 103, 102945–102944. doi: 10.1016/j.ijedudev.2023.102945

Miranda-Gonçalves, R. (2023). Quality education and human rights in the 21st century: deciphering SDG 4. *Law Justice J.* 37, 264–293. doi: 10.5335/rjd.v37i2.15295

MOE (2018). Ethiopian education development roadmap.

Mohajan, H. K. (2017). Two criteria for good measurements in research: validity and reliability. Annals of Spiru Haret university. *Econ. Series* 17, 59–82. doi: 10.26458/1746

Mohebbi, H., Nayernia, A., and Nosrati, R. (2022). The factors contributing to language teachers' effectiveness in an EFL learning context: a questionnaire validation study. *Profile: Issues Teachers' Professional Develop.* 24, 63–79. doi: 10.15446/profile.v24n2.93571

Monzalve-Macaya, M., Sánchez-Contreras, S., and Díaz-Levicoy, D. (2023). Examining classroom behavioral management and support training and future teachers'

effectiveness sense. *Formación Universitaria* 16, 1–10. doi: 10.4067/S0718-50062023000100001

Morris, E. M., Qargha, G. O., and Winthrop, R. (2023). Elevating the purpose of education to achieve the spirit of SDG 4. *Int. J. Educ. Dev.* 103, 102926–102924. doi: 10.1016/j.ijedudev.2023.102926

Musodza, B. R., Runhare, T., Mpeta, M., and Cishe, E. N. (2020). Influence of timing in introducing teacher performance evaluation on effective outcomes: the case of one education district in Zimbabwe. *AJIS* 9, 18–33. doi: 10.36941/AJIS-2020-0108

Nema, P., Srivastava, R., Bhalla, R., and Chakarboty, A. (2023). Impact of social media distraction on student evaluation of teacher effectiveness. *Int. J. Educ. Manag.* 37, 300–313. doi: 10.1108/IJEM-10-2022-0389

Ngala, F. B. J. A., and Odebero, S. O. (2010). Teachers' perceptions of staff development programmes as it relates to teachers' effectiveness: a study of rural primary schools in Kenya. *Educ. Res. Rev.* 5, 1–009.

Ocaña-Zúñiga, C. L., Tineo, M., Fernandez-Zarate, F. H., Quiñones-Huatangari, L., Huaccha-Castillo, A. E., Morales-Rojas, E., et al. (2023). Implementing the sustainable development goals in university higher education: a systematic review. *Int. J. Sustain. Dev. Plan.* 18, 1769–1776. doi: 10.18280/ijsdp.180612

Otzen, T., and Manterola, C. (2017). Sampling techniques on a population study. *Int. J. Morphol.* 35, 227–232. doi: 10.4067/S0717-95022017000100037

Oweis, T., Al Natour, A., Muhaisen, B., and Bader, S. (2022). The effect of Covid-19 pandemic on the teaching-learning process: a theoretical overview. *J. Higher Educ. Theory Prac.* 22, 8–17. doi: 10.33423/jhetp.v22i12.5458

Pedraja-Rejas, L., Rodríguez-Ponce, E., Muñoz-Fritis, C., and Laroze, D. (2023). Sustainable development goals and education: a bibliometric review—the case of Latin America. *Sustainability (Switzerland)* 15, 1–19. doi: 10.3390/su15129833

Pham, P. T., Thi Phan, T. T., Nguyen, Y. C., and Hoang, A. D. (2023). Factor associated with teacher satisfaction and online teaching effectiveness under adversity situations: a case of Vietnamese teachers during COVID-19. *J. Educ.* 203, 605–615. doi: 10.1177/00220574211039483

Podolsky, A., Kini, T., and Darling-Hammond, L. (2019). Does teaching experience increase teacher effectiveness? A review of US research. *J. Professional Capital and Community* 4, 286–308. doi: 10.1108/JPCC-12-2018-0032

Price, L. R. (2017). Psychometric methods: Theory into practice. New York: The Guilford Press.

Puri, K. S., Suresh, K. R., Gogtay, N. J., and Thatte, U. M. (2009). Declaration of Helsinki, 2008: implications for stakeholders in research. *J. Postgrad. Med.* 55, 131–134. doi: 10.4103/0022-3859.52846

Reynolds, D. (1998). Schooling for literacy: a review of research on teacher effectiveness and school effectiveness and its implications for contemporary educational policies. *Educ Rev (Birm)* 50, 147–162. doi: 10.1080/0013191980500206

Rocha, R. (2013). Escala de opinión de los estudiantes sobre la efectividad de la docencia (EOEED) en educación superior. *Formación Universitaria* 6, 5–6. doi: 10.4067/S0718-50062013000600003

Rose, P., and Sayed, Y. (2024). Assessing progress in tracking progress towards the education sustainable development goal: global citizenship education and teachers missing in action? *Int. J. Educ. Dev.* 104, 102936–102934. doi: 10.1016/j.ijedudev.2023.102936

Saeeda, B., Noor, T., Ullah, A., Aznaoui, H., and Ullah, A. (2021). A critical study of the teaching attitudes of teachers at primary school level in Pakistan. *Int. J. Knowledge Learn.* 14, 385–403. doi: 10.1504/ijkl.2021.10041958

Sánchez, E., and Craig, R. (2007). Strengthening teaching effectiveness through cooperative learning activities. *HortTechnology* 17, 254–257. doi: 10.21273/horttech.17.2.254

Sánchez-Cabrero, R., Estrada-Chichón, J. L., Abad-Mancheño, A., and Mañoso-Pacheco, L. (2021). Models on teaching effectiveness in current scientific literature. *Educ Sci (Basel)* 11:18. doi: 10.3390/educsci11080409

Imron, A. (2024). Integrative approaches to internal quality Assurance in Junior High Schools: exploring impacts on student learning, teacher effectiveness, and school management. *Int. J. Religion* 5, 90–102. doi: 10.61707/4wesnw44

Shahzad, S., and Mehmood, N. (2019). Development of teaching effectiveness scale for university teachers. *J. Res. Soc. Sci.* 7, 1–14. doi: 10.52015/jrss.7i2.74

Shin, M., and Bolkan, S. (2021). Intellectually stimulating students' intrinsic motivation: the mediating influence of student engagement, self-efficacy, and student academic support. *Commun. Educ.* 70, 146–164. doi: 10.1080/03634523.2020.1828959

Skourdoumbis, A. (2013). Classroom teacher effectiveness research and inquiry, and its relevance to the development of public education policy: an Australian context. *Int. J. Qual. Stud. Educ.* 26, 967–985. doi: 10.1080/09518398.2012.724465

Sofyan, M., Barnes, M., and Finefter-Rosenbluh, I. (2021). Teacher effectiveness in Asian higher education contexts: a systematic review. *Teach. High. Educ.* 28, 2135–2159. doi: 10.1080/13562517.2021.1952567

Stronge, J. (2007). Qualities of effective teachers. *2nd* Edn. Alexandria, VA: ASCD.

Tadesse, T., Gillies, R. M., and Manathunga, C. (2021). The effects of informal cooperative learning pedagogy on teaching effectiveness, task orientation, and learning

satisfaction in undergraduate classrooms in Ethiopia. *High. Educ. Res. Dev.* 40, 627–645. doi: 10.1080/07294360.2020.1765320

Tomasella, B., Wylie, A., and Gill, D. (2023). The role of higher education institutions (HEIs) in educating future leaders with social impact contributing to the sustainable development goals. *Soc. Enterp. J.* 19, 329–346. doi: 10.1108/SEJ-03-2022-0027

Welsh, M. E. (2011). Measuring teacher effectiveness in gifted education: some challenges and suggestions. *J Adv Acad* 22, 750–770. doi: 10.1177/1932202X11424882

Witcher, A. E., Onwuegbuzie, A., Collins, K., Filer, J., Wiedmaier, C., and Moore, C. (2003). Students' perceptions of characteristics of effective college teachers. *Institute of Education Sciences* 1–49.

Wulandari, S. S., Suratman, B., Trisnawati, N., and Narmaditya, B. S. (2021). Teacher's performance, facilities and students' achievements: does principal's leadership matter? *Pedagogika* 142, 71–88. doi: 10.15823/p.2021.142.4

Yamamoto, K. (1963). Evaluating teacher effectiveness: a review of research. *J. Sch. Psychol.* 2, 60–71. doi: 10.1016/0022-4405(63)90017-7