JUTLP
Journal of University Teaching & Learning Practice

Teacher Assessment in Greek Higher Education Development (TAGGED): Conceptualisation, scale development and validation

Sofia Mastrokoukou[a,c†], Antonia Seremetaki[b†], Nansy Karali[c], Michele Settanni[b], Ali Ijaz[d], Claudio Longobardi[d]

[a] Universitas Mercatorum, Italy; [b] University of Patras, Greece; [c] University of Piraeus, Greece; [d] Università degli studi di Torino, Italy

[†]Shared co-first authorship.

Abstract

Student evaluations of teaching inform quality assurance, but existing instruments vary in length, content, and psychometric quality. This study developed and validated the Teacher Assessment in Greek Higher Education Development (TAGGED), a brief student-reported questionnaire assessing two distinct constructs: perceived Instructional Practices and Student Study Effort. After expert content validation, an 11-item pilot version was administered to 2,233 Greek undergraduate students. Exploratory factor analysis yielded a seven-item instrument with five Instructional Practices items and two Student Study Effort items. Independent validation with 15,133 undergraduates supported the two-factor measurement model (CFI = .998, TLI = .996, RMSEA = .028, SRMR = .007), satisfactory reliability, construct validity, and measurement invariance across gender. Student Study Effort must not be interpreted as an indicator of teacher effectiveness or combined with Instructional Practices into an overall teacher-effectiveness score. TAGGED offers focused, course-level feedback on observable instructional practices alongside a separate measure of students' behavioural engagement.

Editors

Section: Educational Psychology
Senior Editor: Prof Louise Taylor
Associate Editor: Dr Sarah Rose

Publication

Submission: 4 April 2026
Revised: 31 May 2026
Accepted: 16 November 2026
Published: 17 November 2026

Copyright © by the authors, in its year of first publication. This publication is an open access publication under the Creative Commons Attribution CC BY-ND 4.0 license.

Practitioner Notes

1. Brief teaching-evaluation questionnaires may reduce respondent burden and survey fatigue.
2. TAGGED assesses instructional practices and student study effort as separate constructs.
3. Its concise format may support consistent course-level evaluation across Greek universities.
4. Feedback on organisation, clarity, interest, and participation can inform teaching enhancement.
5. Student evaluations should complement peer review and other evidence of teaching quality.

Keywords

Student evaluation of teaching; instructional practices; scale development; student engagement; Greece

Citation:

Mastrokoukou, S., Seremetaki, A., Karali, N., Settanni, M., Ijaz, A., & Longobardi, C. (2026). Teacher Assessment in Greek Higher Education Development (TAGGED): Conceptualisation, scale development and validation. *Journal of University Teaching and Learning Practice, 23*(4). https://doi.org/10.53761/jhk2ae85

Introduction

In developing human capital and supporting innovation within knowledge-based economies (Blaskova et al., 2015; Mastrokoukou & Crawford-Lee, 2023), higher education (HE) is essential. Over the past decades, HE systems have expanded and diversified in response to growing societal and economic demands for advanced knowledge and skills (Mastrokoukou et al., 2022; Milenios et al., 2021). As the HE system continues to evolve, the assessment of teaching performance has become a significant aspect of the review process and should be examined for quality assurance and accountability at the institutional level.

Student evaluations of teaching (SETs) are widely used to assess instructional quality and support institutional improvement efforts (Flora & Flake, 2017; Seldin, 1999). However, existing evaluation instruments are often institution-specific, lengthy, and inconsistently designed, limiting comparability across higher education institutions (HEIs) (Fernandes & Flores, 2022). Establishing reliable and standardised evaluation frameworks may enhance institutional quality assurance and facilitate benchmarking across contexts (Johnson et al., 2016; Zuñiga et al., 2010).

SETs typically collect information regarding teaching practices, course delivery, curriculum organisation, and students' learning experiences (Revilla & Ochoa, 2017; Van Mol, 2017). Despite their widespread adoption, concerns remain regarding their validity, susceptibility to bias, and ability to capture the multidimensional nature of teaching effectiveness (Quansah et al., 2024).

Online surveys have become increasingly common in teaching evaluation research because they enable efficient data collection from large student populations (Van Mol, 2017). While web-based surveys offer practical advantages, declining response rates remain a persistent challenge (Saleh & Bista, 2017). Questionnaire length has consistently been identified as a key factor influencing participation, with shorter instruments generally associated with higher response rates and lower respondent burden (Rolstad et al., 2011; Sahlqvist et al., 2011). Consequently, the development of concise yet psychometrically sound evaluation instruments represents an important methodological priority.

Parsimony was therefore treated as a practical design constraint rather than an end in itself. The aim was to identify a brief set of non-redundant and interpretable indicators that could be used in repeated institutional evaluation while limiting respondent burden. Brevity was not intended to substitute for construct coverage; rather, the study sought an explicit balance between measurement efficiency and the breadth of the targeted domains (Rolstad et al., 2011; Sahlqvist et al., 2011).

Greece offers a useful setting for examining these challenges because national quality-assurance reforms coexist with institutionally developed student-evaluation instruments. Despite reforms over the past two decades, Greek HEIs continue to use course-evaluation measures that vary in content, length, and design (Drosos et al., 2023; Zmas, 2015). This variation may limit comparability across institutions and complicate the use of feedback for monitoring and improving instructional quality.

The present study addresses this gap through the development and validation of the Teacher Assessment in Greek Higher Education Development (TAGGED), a concise student-reported instrument designed to assess observable instructional practices and student study effort. Drawing on literature concerning instructional quality and student engagement, the study

evaluates the factorial structure, reliability, validity, and measurement invariance of the proposed instrument. Rather than capturing the full multidimensional construct of teaching effectiveness, TAGGED provides a parsimonious operationalisation of core instructional practices and engagement-related study behaviours that can support institutional evaluation processes.

Teacher Effectiveness and Instructional Quality

Teacher effectiveness plays a central role in promoting student learning and achievement, a relationship consistently supported by empirical research (Ammigan & Jones, 2018; Bardach & Klassen, 2020). Despite its importance, teacher effectiveness continues to be defined and assessed in various ways, making comparisons difficult because definitions of high-quality teaching vary across educational contexts and theoretical perspectives (Behera et al., 2019; Senyametor et al., 2020). One common way to evaluate teaching effectiveness in HE research is through student evaluations, which provide a systematic means of measuring students' perceptions of the quality and effectiveness of instruction and the learning environment (Kim et al., 2019). Recent evidence, however, supports a multidimensional conceptualisation of teacher effectiveness and indicates that such evaluations should be interpreted as informative, but not exhaustive, indicators of instructional quality (Quansah et al., 2024; Şahin Kürşad et al., 2026; Woodzicka et al., 2025).

For the purposes of this study, teacher effectiveness is defined narrowly rather than as a simplistic or all-encompassing construct: the focus is on student-observable aspects of instructional practice that can be assessed consistently across a course, specifically organization of teaching, clarity and quality of delivery, engagement of interest, and promotion of participation. These aspects are supported by previous research on teaching practices and teaching effectiveness in HE (Klassen & Tze, 2014; López & Dávila, 2019; Sofyan et al., 2021), and the operationalization is further consistent with recent evidence indicating that student assessments should concentrate on specific teaching aspects that students can observe and evaluate directly (Quansah et al., 2024; Woodzicka et al., 2025). This narrower definition does not, however, imply that other aspects of teacher effectiveness, such as subject-matter knowledge, assessment methods and fairness, mentoring, and relationship building, are unimportant; rather, the Instructional Practices variable should be interpreted as a narrow characteristic of perceived teacher effectiveness rather than a comprehensive measure of all its aspects.

More broadly, effective teaching is typically associated with a range of pedagogical practices that facilitate student learning, including clear course organization, effective communication, and the use of instructional strategies that encourage active participation and critical thinking. Research has shown that teachers who foster collaborative learning and stimulate intellectual engagement contribute to deeper understanding and sustained motivation among students (López & Dávila, 2019; Sofyan et al., 2021). Accordingly, teaching effectiveness can be partially examined through core instructional practices that shape the classroom environment, often operationalised in evaluative research through students' perceptions of clarity, organisation, and engagement. Consistent with this interpretation, recent meta-analytic evidence indicates a modest association between student evaluations of teaching and academic achievement, supporting their use as one indicator of instructional quality while cautioning against treating them as comprehensive measures of teaching effectiveness (Şahin Kürşad et al., 2026).

Teaching performance encompasses a broad set of pedagogical behaviours, including instructional methods, classroom interaction, and strategies used to support student participation and comprehension (Klassen & Tze, 2014; Tickell & Klassen, 2024). These practices are embedded within broader educational philosophies that shape conceptions of teaching and learning. For example, Paulo Freire's critique of the traditional "banking model" of education highlights the limitations of instructional approaches that treat students as passive recipients of knowledge (Freire, 1970). In contrast, student-centred pedagogies emphasize dialogue, active participation, and critical engagement with knowledge, thereby promoting deeper and more meaningful learning experiences (Baeten et al., 2016; Beausaert et al., 2013).

While these theoretical perspectives highlight the multidimensional nature of teaching effectiveness, empirical measurement in HE often focuses on observable instructional practices that can be consistently evaluated across courses and institutional contexts. This reflects a necessary balance between theoretical complexity and the practical requirements of measurement in large-scale evaluation systems. Such parsimonious approaches are particularly valuable in large-scale student evaluation systems, where brevity and clarity are needed to sustain participation and preserve data quality.

Student Study Effort and Learning Engagement

Student study effort represents a behavioural dimension of learning engagement that is conceptually distinct from teacher effectiveness, although it may be associated with the instructional conditions students experience (Parsons & Taylor, 2011). Although the terms engagement and involvement are sometimes used interchangeably, some scholars distinguish between them (Strayhorn, 2018; Wolf-Wendel et al., 2009). Wolf-Wendel et al. (2009) suggest that engagement is shaped largely by the institutional environment that students perceive as supportive and inclusive, whereas involvement refers more specifically to the actions and behaviours through which students participate in academic activities.

From this perspective, learning engagement can be understood as the interaction between students and their educational environment through the investment of time, effort, and related resources that support learning and personal development (Fredricks et al., 2004; Kahu, 2013; Krause & Coates, 2008; Lamborn et al., 1992; Trowler, 2010; Wladis et al., 2014; Zhoc et al., 2019). Research has consistently shown that stimulating teaching practices and supportive learning environments encourage higher levels of student participation and engagement (Markwell, 2007).

Instructional quality may influence the extent to which students invest effort in their studies. When teaching practices are perceived as clear, engaging, and well organised, students are more likely to participate actively in the learning process and devote greater time and effort to their academic work (Cinches et al., 2017). Within this framework, student study effort is conceptualised as a specific, behavioural component of learning engagement. Specifically, it refers to the behavioural dimension of engagement, capturing students' investment in their studies through the frequency and intensity of study-related activities. Although theoretically associated with perceived teaching quality, study effort is treated as conceptually and analytically distinct from teaching performance, reflecting a related but separate aspect of the teaching–learning process. Including this separate construct allows the instrument to capture not only students' perceptions of instructional practices

but also a behavioural indicator of engagement. This provides a more integrated perspective on the teaching–learning process.

This conceptual distinction enables the examination of how instructional practices relate to students' behavioural engagement without conflating teaching performance with student outcomes.

Importantly, Student Study Effort is treated as a separate construct in TAGGED and is not considered either a measure of effectiveness or a contributing factor in assessing teacher effectiveness. Instead, it is included in TAGGED as a variable indicating student engagement in the learning process and enabling the relationship between teaching and learning to be studied. In view of this, it must be emphasized that Student Study Effort scores do not serve as direct measures of teaching effectiveness.

Institutional Setting: The Greek Higher Education System

Over the past two decades, the Greek HE system has undergone a series of reforms aimed at strengthening governance structures, improving quality assurance mechanisms, and aligning national policies with the European Higher Education Area (EHEA) (European Association for Quality Assurance in Higher Education [ENQA], 2015; Zmas, 2015). Legislative initiatives such as Laws 3374/2005 and 4009/2011 introduced formal procedures for institutional evaluation and monitoring (Hellenic Republic, 2005, 2011). Law 3374/2005 established national quality assurance framework in HE, while Law 4009/2011 further strengthened accreditation procedures and aligned evaluation processes with European standards. As part of these reforms, the Hellenic Quality Assurance and Accreditation Agency (HQAA) was established as the central body responsible for coordinating quality assurance processes in higher education.

Further structural changes were introduced through the Athena Plan following the Higher Education Act of 2011, aiming to consolidate departments, improve efficiency, and align HE with national and regional development priorities (OECD, 2011). These reforms were partly driven by broader European policy influences and intensified by the economic pressures of the 2009 financial crisis, which highlighted inefficiencies in the Greek HE system and the need for structural rationalisation (OECD, 2011).

Additional legislative amendments in subsequent years continued institutional restructuring and governance reforms. Despite these developments, concerns remain regarding the effectiveness of quality assurance mechanisms and the extent to which they support consistent and comparable evaluation practices across institutions (OECD, 2011; Zmas, 2015). In particular, the implementation of evaluation processes has been uneven, limiting their impact on institutional improvement and accountability. More recent evidence shows that student evaluation is now part of the procedures for ensuring quality in education in Greece. However, this evidence also reveals shortcomings related to the psychological framework and the use of existing evaluation forms (Papadogiannis et al., 2024).

Compared with the growing international emphasis on student-centred learning and systematic teaching evaluation frameworks, Greek HE has largely retained a more teacher-centred instructional tradition (Zmas, 2015). This contrasts with more established evaluation systems in other national contexts, such as the Course Experience Questionnaire (CEQ), the National Survey of Student Engagement (NSSE), and the National Student Survey (NSS), which rely

extensively on student feedback for quality monitoring (Alzafari & Ursin, 2019; Jingura & Kamusoko, 2019). Nonetheless, these frameworks differ from TAGGED in their coverage and measurement function. While NSS and NSSE cover a range of dimensions, including students' educational experiences, their learning environment, and their engagement in the learning process, TAGGED focuses on observable instructional practices at the course level.

TAGGED is not intended to replace broader national or institutional student-experience frameworks such as the NSS or NSSE. Its contribution is complementary: students are positioned to report on instructional practices that they directly and repeatedly experience within individual courses, while their own study effort can be assessed separately as an engagement-related behaviour (Quansah et al., 2024; Woodzicka et al., 2025). Accordingly, TAGGED distinguishes perceived Instructional Practices from Student Study Effort rather than treating the latter as evidence of teacher effectiveness.

Although Greek HEIs are permitted to develop their own evaluation instruments within the national quality assurance framework (Hellenic Republic, 2011), variation in course-evaluation practices may constrain the comparability and consistent interpretation of student feedback across institutional contexts (OECD, 2011; Papadogiannis et al., 2024).

Against this background, TAGGED was developed to address a specific measurement gap by providing a concise, psychometrically evaluated course-level instrument that assesses student-observable instructional practices while retaining Student Study Effort as a conceptually distinct engagement-related construct.

Aims of the Study

Taken together, the literature identifies three gaps that motivate the present study. First, teacher effectiveness is multidimensional, whereas student evaluations are most appropriate for assessing instructional practices that students can directly observe (Quansah et al., 2024; Woodzicka et al., 2025). Second, Greek HE lacks a standardized course-level instrument that supports consistent measurement and comparison across institutions. Third, existing evaluation tools vary considerably in scope and length, potentially increasing respondent burden and obscuring the specific construct being assessed. TAGGED addresses these limitations through a concise, context-sensitive measure of student-observable instructional practices, with Student Study Effort retained as a distinct engagement-related construct rather than an indicator of teacher effectiveness. This conceptual separation is consistent with evidence that student evaluations provide informative but bounded evidence of instructional quality (Şahin Kürşad et al., 2026).

The present study aims to develop and validate a concise student-reported instrument for use in Greek HE. Specifically, it seeks to address the absence of a standardised and psychometrically tested questionnaire that can be used consistently across Greek HEIs to evaluate students' perceptions of instructional practices. In doing so, the study also responds to broader methodological challenges related to the development of efficient and reliable student evaluation instruments.

To this end, the study introduces the Teacher Assessment in Greek Higher Education Development (TAGGED), a newly developed instrument designed to assess core observable instructional practices associated with teaching effectiveness together with the separate

behavioural-engagement construct of student study effort. The instrument is designed to provide a concise operationalisation of these constructs rather than a comprehensive representation of the full multidimensional construct of teaching effectiveness. Using data collected from undergraduate students across multiple academic departments, the study evaluates the psychometric properties of the instrument by examining its factorial structure, internal consistency, and construct validity. More specifically, the study tests the factor structure of the instrument and assesses its reliability and validity within a large and diverse student sample.

In addition, the study investigates whether the measurement structure of the instrument remains stable across different student groups, including gender categories. By examining measurement invariance across these groups, the study seeks to determine whether the TAGGED questionnaire can provide reliable and comparable assessments of the two measured constructs within the broader Greek HE context. This step is critical for establishing the generalisability and practical applicability of the instrument across student populations.

Method

Study Design

The study used a two-stage scale-development and validation design. Stage 1 used the pilot sample (*N* = 2,233) for exploratory factor analysis (EFA) and item refinement following content validation. Stage 2 used an independent validation sample (*N* = 15,133) for confirmatory factor analysis (CFA), reliability and construct validity evaluation, and measurement invariance testing. The pilot data were not included in the validation analyses; thus, the exploratory and confirmatory analyses used non-overlapping samples.

Participants

The pilot phase included 2,233 undergraduate students from several universities across Greece. Participants were between 18 and 24 years old (M_{age} = 20.31, *SD* = 1.42), with 57.81% identifying as female and 42.14% as male. No participants reported other gender identities.

The validation phase involved a substantially larger sample of 15,133 undergraduate students drawn from multiple Greek universities and academic departments. Participants were also between 18 and 24 years old (M_{age} = 20.48, *SD* = 1.37), with 57.42% female and 42.55% male students.

These demographic characteristics broadly reflect the composition of the undergraduate student population in Greece, where most students fall within the 18–24 age range and female students slightly outnumber male students (Hellenic Statistical Authority, 2025). However, given the use of non-probability sampling, the samples should not be considered statistically representative of the Greek undergraduate population.

Departmental information was available for 14,672 participants in the validation sample. These participants were distributed across seven academic departments representing different disciplinary fields, as presented in Table 1.

Table 1

Distribution of Participants With Available Departmental Information in the Validation Sample

Department	Frequency	Percentage (%)	Cumulative Percentage (%)
International & European Studies	2,638	18.0	18.0
Economics	2,497	17.0	35.0
Maritime Studies	2,449	16.7	51.7
Business Administration	2,373	16.2	67.9
Statistics & Insurance Science	2,116	14.4	82.3
Industrial and Management Technology	1,451	9.9	92.2
Digital Systems	1,148	7.8	100.0
Total	14,672	100.0	

Ethical Approval

The study received ethical approval from the Institutional Research Ethics Committee (Protocol No. 2024014/19.09.2024). All procedures were conducted in accordance with the ethical principles of the Declaration of Helsinki. Participation was voluntary and anonymous, and informed consent was obtained from all participants prior to data collection. Participants were informed of their right to withdraw from the study at any time without penalty.

The study complied with the requirements of the General Data Protection Regulation (GDPR; EU Regulation 2016/679). No personally identifiable information was collected, and all responses were analysed in aggregated form. Data were stored securely and used exclusively for research purposes. Access to the data was restricted to the research team.

Development of the New Instrument

The Teacher Assessment in Greek Higher Education Development (TAGGED) questionnaire was developed following established guidelines for psychometric scale construction (DeVellis, 2017; Hinkin, 1998). The development process began with a focused review of 18 papers on instructional practices associated with teaching effectiveness, student engagement, and student evaluations of teaching in HE. The review was conducted to identify recurrent conceptual domains relevant to item generation and was not intended as a systematic evidence synthesis.

Based on this review an *a priori* preliminary conceptual framework was specified before item generation to represent a broad range of course-level teaching and learning characteristics, including instructional practices and student engagement-related behaviors. The literature review identified key dimensions of instructional quality, including organisation, clarity of explanation, stimulation of student interest, and encouragement of participation, alongside behavioural indicators of student engagement reflected in study effort. Drawing on these domains, the research team generated an initial pool of candidate items designed to capture observable teaching practices and student learning behaviours. Item development involved iterative discussions among the researchers to ensure conceptual relevance, clarity of wording, and applicability across different disciplinary contexts within Greek HE. Items judged to be conceptually redundant, ambiguous, or insufficiently aligned with the intended constructs were revised or removed during this stage. This process resulted in an initial pool of 21 items.

Consistent with scale-development guidance emphasising a priori domain specification and transparent item generation (Gupta et al., 2025), the candidate pool was generated primarily through a deductive process. The principal sources informing domain specification supported organisation and clarity/effective delivery (Kim et al., 2019; Klassen & Tze, 2014), stimulation of interest and participation (López & Dávila, 2019; Sofyan et al., 2021), and behavioural study effort as a component of student engagement (Fredricks et al., 2004; Kahu, 2013). At the item-generation stage, candidate items were considered suitable when they were student-observable, conceptually aligned with one of the intended domains, applicable across disciplinary contexts, clearly worded, and non-redundant.

For transparency, the original 21-item pool (see Appendix 2) sampled a broader set of course- and teaching-evaluation content areas than the final scale. These content areas were informed by the literature reviewed during domain specification and were intended to provide broad initial coverage rather than an exhaustive representation of teacher effectiveness. The pool comprised: (1) course aims, curriculum alignment, and content organisation (3 items; Q1-Q3); (2) learning materials and resources (4 items; Q4-Q7); (3) prerequisites, prior knowledge, course difficulty, and workload (4 items; Q8-Q11); (4) core instructional practices, including organisation of delivery, stimulation of interest, effective delivery of concepts, and encouragement of participation (4 items; Q12-Q15); (5) teacher professional/accessibility practices, including fulfilment of obligations, use of educational technology, and accessibility to students (3 items; Q16-Q18); and (6) student attendance and study effort (3 items; Q19-Q21).

Particular attention was given to ensuring conceptual clarity, contextual relevance to the Greek HE context, and concise item wording to minimize respondent fatigue in survey-based evaluations. Most candidate items were rated using a five-point agreement scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Items assessing behavioural frequency or study effort used response categories appropriate to the behaviour assessed.

For the agreement-based items, the five-category response format was selected as a pragmatic means of maintaining a straightforward response task while providing an ordered range of response options. This choice does not imply that five response categories are psychometrically superior to seven. Because response-scale format can influence response styles and measurement characteristics, alternative category structures should be examined in future applications of TAGGED (Weijters et al., 2010).

Following item generation, the preliminary item set was evaluated by a panel of subject matter experts (SMEs) with expertise in HE pedagogy and student engagement. The experts reviewed the items in terms of relevance, clarity, and conceptual alignment with the intended constructs. Content validity was assessed using established criteria for expert agreement. This process resulted in an 11-item version that was subsequently administered during the pilot phase.

Pilot-stage item refinement resulted in the final seven-item version of the questionnaire, which was subsequently administered to the independent validation sample for psychometric evaluation.

Scale Composition

The following section describes the final seven-item version obtained after expert review and pilot-stage item refinement; it does not describe the original 21-item candidate pool.

The TAGGED questionnaire consists of seven items organised into two subscales measuring two separate constructs: five items assessing Instructional Practices and two items assessing Student Study Effort.

The five Instructional Practices items assess students' perceptions of the organisation of course delivery, clarity and effectiveness of instructional delivery, stimulation of student interest, and encouragement of participation during class. They therefore capture a focused set of student-observable instructional practices rather than the broader construct of teacher effectiveness.

The two Student Study Effort items assess students' own study behaviour, specifically how frequently students study for the course and the amount of time they typically devote to studying each week. Student Study Effort is scored and interpreted separately from Instructional Practices and is not included as an indicator of teacher effectiveness.

Procedure

Stage 1: Expert content validation

Following the development of the TAGGED scale, its content validity was evaluated through consultation with a panel of SMEs. Based on their feedback, several revisions were made to improve the clarity and coherence of the instrument. The revised version of the questionnaire was then used in the pilot phase of the study.

Stage 2: Pilot EFA and item refinement

During the pilot phase, 2,240 responses were collected using the 11-item questionnaire. Following data screening, seven responses were excluded, resulting in a final pilot sample of 2,233 undergraduate students. This stage constituted the first empirical phase of instrument development and was designed to examine item performance and explore the latent structure of the scale. The results of the pilot EFA informed subsequent item refinement, with retention decisions based on both psychometric evidence and theoretical interpretability.

Stage 3: Independent validation

Following pilot-stage item refinement, the final seven-item version of the instrument was administered during the validation phase. A total of 15,252 responses were collected. Following data screening, 119 responses were excluded, resulting in a final independent validation sample of 15,133 undergraduate students. This phase was used to test the final measurement model and evaluate reliability, construct validity, and measurement invariance.

Data collection

Data were collected during the 2024–2025 academic year in the middle of the semester, because by that point students should have passed the "honeymoon period" (Lorber et al., 2015) and would be less likely to give overly positive scores.

To distribute the survey, it was posted online using student networking sites such as university communication systems and student social media, which represents a convenience sampling method. All students were encouraged to follow a link to the survey and share it with their fellow students. Participation was completely voluntary, and participants signed informed consent before completing the survey.

Before the analysis began, responses were screened for missing values and straight lining. Cases with either of these problems were removed from the final dataset. Missing values on individual items did not exist in the final analytic samples.

Along with collecting survey responses, participants' age and self-reported gender were also recorded. In addition to being described in the analysis of the respondent sample, age and gender were used in the measurement invariance test.

Plan of Analysis

For the expert ratings described above, content validity was quantified using the Item-Level Content Validity Index (I-CVI), defined as the proportion of experts rating an item as "essential". Items meeting the .70 retention threshold were carried forward, and the Scale-Level Content Validity Index (S-CVI) was calculated as the mean of retained I-CVI values. Expert comments on wording were also considered when identifying ambiguity or interpretation problems (Zamanzadeh et al., 2015).

Pilot-stage exploratory factor analysis

Following the content validity assessment, the retained items were administered to a pilot sample of undergraduate students to examine their psychometric functioning. The data were analysed using Principal Axis Factoring in IBM SPSS 27 with oblique rotation (direct oblimin), as the underlying constructs were expected to be correlated (Field, 2013; Hair et al., 2019). Sampling adequacy was assessed using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity.

The number of factors to retain was determined using the Kaiser criterion (eigenvalues greater than 1), inspection of the scree plot, the proportion of explained variance, and theoretical interpretability of the resulting factor solution.

Communalities below .40 were treated as an indication that an item required closer examination rather than as an automatic deletion criterion. Item refinement considered communalities, factor loadings and cross-loading patterns together with theoretical interpretability and the intended construct coverage.

Pilot-stage reliability was assessed separately for the two resulting factors. Cronbach's alpha was used for the five-item Instructional Practices factor, whereas split-half reliability was used for the two-item Student Study Effort factor.

Independent validation and confirmatory factor analysis

The refined seven-item set was subsequently administered to the independent validation sample. CFA was conducted using AMOS 24 with maximum likelihood estimation to test the fit of the two-factor measurement model identified during the pilot-stage EFA (Hayduk & Littvay, 2012). Model fit was evaluated using the comparative fit index (CFI), Tucker–Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). CFI and TLI values ≥ .95, RMSEA values ≤ .06, and SRMR values ≤ .08 were used as commonly recommended benchmarks indicative of good model fit (Brown, 2015; Hair et al., 2019; Kline, 2016; West et al., 2012). These thresholds were specified a priori as interpretive guidelines rather than rigid decision rules.

Measurement invariance across gender was then examined to determine whether the instrument operated equivalently across gender groups. The analysis followed a stepwise procedure including configural, metric, and scalar invariance models. Changes in model fit were evaluated primarily using ΔCFI and ΔRMSEA, which are considered robust indicators of measurement invariance (Chen, 2007; Putnick & Bornstein, 2016). A change in CFI (ΔCFI) of less than .01 was used as the primary criterion for invariance.

Finally, reliability was evaluated separately for the two factors. Internal consistency of the five-item Instructional Practices factor was assessed using Cronbach's alpha, with values above .70 specified *a priori* as indicating satisfactory reliability (Taber, 2018). For the two-item Student Study Effort factor, reliability was additionally evaluated using the inter-item correlation and split-half reliability.

Results

Content Validity

Content validity was evaluated by a panel of eight SMEs, including university professors and graduate researchers with expertise in instructional practices and student engagement. Experts assessed the relevance and clarity of the initial pool of 21 items using a content validation form, and Item-Level Content Validity Index (I-CVI) and Scale-Level Content Validity Index (S-CVI) values were computed.

Of the 21 initial items, 10 did not meet the recommended threshold (I-CVI < .70) and were therefore excluded from further analysis. The remaining 11 items demonstrated strong content validity, with I-CVI values ranging from .88 to 1.00. The S-CVI for the retained items was .91, indicating adequate content validity of the revised item set.

Accordingly, the 11 retained items were carried forward to the pilot study.

Pilot Testing

Following content validation, the 11-item questionnaire was examined in the pilot sample (N = 2,233) using principal axis factoring with direct oblimin rotation. Sampling adequacy was high (KMO = .900), and Bartlett's test of sphericity was statistically significant, $\chi^2(55) = 12{,}216.174$, $p < .001$. The Kaiser criterion identified two factors with eigenvalues greater than 1 (5.328 and 1.447, respectively), while the third eigenvalue was .978. Inspection of the scree plot was also consistent with retention of a two-factor solution. Together with the theoretical interpretability of the solution, these results supported a two-factor structure. The two extracted factors accounted for 53.81% of the variance.

Item refinement subsequently considered communalities, factor loadings, cross-loading patterns, and theoretical interpretability. Two items had extraction communalities below .40, while the remaining refinement decisions reflected the combined psychometric and conceptual evaluation of the items. Four items were ultimately removed, resulting in the final seven-item version.

The EFA was then repeated using the seven retained items. The resulting solution again supported two factors corresponding to Instructional Practices and Student Study Effort. Sampling adequacy was satisfactory (KMO = .825), and Bartlett's test of sphericity was statistically significant, $\chi^2(21) = 7{,}188.224$, $p < .001$. The two-factor solution explained 60.78% of the variance. All retained items loaded clearly on their respective factors, with loadings ranging from .66 to .85 for Instructional Practices and loadings of .77 for both Student Study Effort items. The resulting factor solution is presented in Table 2.

Pilot-stage reliability was satisfactory for Instructional Practices (Cronbach's α = .884) and Student Study Effort (split-half reliability = .748). No reverse-keyed items were included in the original 21-item pool, the 11-item pilot version, or the final seven-item version.

Table 2

Exploratory factor analysis results for the TAGGED scale (pilot sample)

Item	Instructional Practices	Student Study Effort
The teacher organises the delivery of the curriculum well (Q1)	.85	
The teacher stimulates interest in the topic of the course (Q2)	.83	
The teacher delivers course-related concepts effectively (Q3)	.82	
The teacher encourages students to participate and ask questions (Q4)	.72	
The content being taught was well organised (Q5)	.66	
I study frequently (Q6)		.77
Hours of study per week (Q7)		.77

Note. N = 2,233. Extraction method: principal axis factoring. Rotation method: direct oblimin with Kaiser normalization. Rotation converged in three iterations. Total variance explained = 60.78%. Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy = .825. Bartlett's test of sphericity was significant, $\chi^2(21) = 7{,}188.224$, $p < .001$. Only factor loadings ≥ .40 are displayed.

Validation Sample

Of the 15,252 responses collected during the validation phase, 119 were excluded because of missing responses or invariant response patterns across the target questionnaire items, resulting in a final analytic sample of 15,133. No item-level missing values remained in the final analytic dataset.

The final 7-item version of the TAGGED scale was administered to the validation sample (N = 15,133) to evaluate the psychometric properties of the instrument. The analyses focused on examining the factor structure, descriptive properties of the items, reliability, and measurement invariance across student subgroups.

Specifically, CFA was conducted to test the fit of the proposed measurement model, followed by assessments of reliability and construct validity. Measurement invariance was examined to evaluate the stability of the factor structure across student groups.

Descriptive Statistics

The descriptive statistics reported below refer to the independent validation/CFA sample.

Descriptive statistics were computed for all seven items of the TAGGED scale, including the mean, standard deviation, minimum, maximum, skewness, and kurtosis. Item means ranged from 2.31 to 3.66, indicating moderate to relatively high endorsement of items related to instructional practices and student study effort. Standard deviations indicated adequate variability across response categories.

Inspection of the distributional properties showed acceptable levels of skewness and kurtosis, with values within commonly recommended thresholds for factor analytic procedures (|skewness| < 2). Kurtosis values were also within acceptable limits (|kurtosis| < 2), supporting the assumption of approximate univariate normality.

No evidence of substantial floor or ceiling effects was observed, and all items demonstrated sufficient variance for subsequent latent structure analyses.

Descriptive statistics for the TAGGED items are presented in Table 3.

Table 3

Descriptive statistics for TAGGED scale items (validation sample)

Item	*M*	*SD*	Min	Max	Skewness	Kurtosis
Q1. The teacher organises the delivery of the curriculum well	3.66	1.02	1	5	-0.41	-0.72
Q2. The teacher stimulates interest in the topic of the course	3.44	1.08	1	5	-0.33	-0.61
Q3. The teacher delivers course-related concepts effectively	3.51	1.03	1	5	-0.36	-0.69
Q4. The teacher encourages students to participate and ask questions	3.58	1.05	1	5	-0.39	-0.73
Q5. The content being taught was well organised	3.47	1.01	1	5	-0.31	-0.62
Q6. I study frequently	2.75	1.12	1	5	0.18	-0.48
Q7. Hours of study per week	2.31	1.19	1	5	0.32	-0.40

Note. *N* = 15,133. *M* = mean; *SD* = standard deviation. Skewness and kurtosis values are within acceptable limits (|2|).

Confirmatory Factor Analysis (CFA)

A CFA was conducted using AMOS 24 on the validation sample (N = 15,133) to test the fit of the hypothesized two-factor measurement model underlying the TAGGED scale. The model was estimated using the maximum likelihood (ML) method.

Model fit was evaluated using several goodness-of-fit indices, including the comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The results indicated an excellent fit of the model to the data: CFI = .998, TLI = .996, RMSEA = .028, and SRMR = .007. All fit indices met

commonly recommended criteria for good model fit (Brown, 2015; Hair et al., 2019; Kline, 2016; West et al., 2012), indicating a very good fit of the proposed two-factor model.

The excellent model fit should be interpreted in light of the model's parsimony, as models with a small number of indicators often produce highly favourable fit indices.

All standardised factor loadings were statistically significant. Loadings ranged from .67 to .89 for the Instructional Practices factor and from .61 to .98 for the Student Study Effort factor. These results indicate strong relationships between the observed items and their respective latent constructs.

Further evaluation of the measurement model supported the reliability and validity of the constructs. Internal consistency was high for Instructional Practices (α = .91). Because the Student Study Effort factor consisted of two items, reliability was assessed using split-half reliability, which yielded an acceptable value (.74). The inter-item correlation for this factor was also within acceptable limits.

Given the limited number of indicators for the Student Study Effort factor, these results should be interpreted with caution, although they are consistent with recommended thresholds for two-item constructs.

Convergent validity was supported by average variance extracted (AVE) values exceeding the recommended threshold of .50, while discriminant validity was indicated as AVE values exceeded the corresponding maximum shared variance (MSV) for both constructs (Hair et al., 2014).

The standardized factor loadings and validity indicators for the measurement model are presented in Table 4.

Measurement Invariance

Measurement invariance across gender was examined using multigroup confirmatory factor analysis. A stepwise procedure was applied, including configural, metric, and scalar invariance models. Nested models were compared by examining changes in model fit between constrained and unconstrained models.

The chi-square difference test indicated no significant deterioration in model fit across gender ($\Delta\chi^2$, p = .993). In addition, changes in comparative fit indices were negligible (ΔCFI < .01), further supporting invariance across groups.

These findings support configural, metric, and scalar invariance of the TAGGED scale across gender, indicating that the factor structure, factor loadings, and item intercepts are comparable for male and female students.

This suggests that the instrument operates equivalently across gender groups, allowing for meaningful comparisons of latent constructs.

Table 4

Measurement Model for the TAGGED Scale (Confirmatory Factor Analysis)

Construct / Item	Loadings (Male)	Loadings (Female)	Loadings (Overall)	Internal Consistency (Male)	Internal Consistency (Female)	Internal Consistency (Overall)	AVE (Male)	AVE (Female)	AVE (Overall)	MSV (Male)	MSV (Female)	MSV (Overall)
Instructional Practices (IP)				.91	.91	.91	.67	.67	.67	.07	.05	.06
Organized delivery of curriculum (Q1)	.82	.84	.83									
Stimulation of interest in the course (Q2)	.90	.89	.89									
Effective delivery of concepts (Q3)	.89	.89	.89									
Encouragement of student participation (Q4)	.79	.78	.78									
Good organization of content (Q5)	.67	.66	.67									
Student Study Effort (SSE)				.74	.74	.74	.67	.65	.66	.07	.05	.06
Frequency of study (Q6)	.99	.97	.98									
Weekly study hours (Q7)	.60	.61	.61									

Note. CFA fit statistics for the overall model: CMIN/DF = 12.46, $p < .001$, CFI = .998, TLI = .996, RMSEA = .028, PCLOSE = 1.000, SRMR = .007. Reliability was estimated using Cronbach's α for Instructional Practices and split-half reliability for Student Study Effort. AVE = Average Variance Extracted. MSV = Maximum Shared Variance. N = 15,133 (male = 6,436; female = 8,697). All factor loadings were statistically significant ($p < .001$).

Discussion

The present study aimed to develop and validate the TAGGED scale as a concise instrument for assessing student-reported instructional practices associated with teaching effectiveness alongside the separate construct of student study effort within Greek HE. The findings provide evidence for the psychometric adequacy of the instrument, as both exploratory and confirmatory factor analyses converged on a stable two-factor structure representing Instructional Practices and Student Study Effort. This structure distinguishes students' perceptions of observable instructional practices from their own behavioural investment in learning, while retaining both as conceptually distinct constructs within the broader teaching-learning context (Fredricks et al., 2004; Kahu & Nelson, 2018).

Previous research has consistently emphasised that effective teaching in HE involves not only the delivery of course content but also the creation of learning environments that encourage active student participation and engagement (Bardach & Klassen, 2020; Sofyan et al., 2021).

The present findings provide empirical support for this literature in the context of Greek HE: organization, clarity, stimulation of interest, and encouragement of involvement form a stable factor of Instructional Practices that was preserved through both the exploratory and confirmatory phases of scale development. These findings provide empirical support for treating these student-observable practices as a focused course-level domain within broader conceptualisations of teaching effectiveness, rather than as an exhaustive representation of teacher effectiveness.

The Instructional Practices factor captures key aspects of instructional practice experienced by students, including the organisation of course delivery, clarity of explanations, and the ability of instructors to stimulate interest and encourage participation. Their convergence within a single factor is consistent with previous research identifying these characteristics as central to effective, interactive, and student-centred teaching in HE (Darling-Hammond, 2000; Klassen & Tze, 2014). The findings also complement evidence that effective instruction extends beyond knowledge transmission to practices that support cognitive engagement and deeper learning (Kim et al., 2019). However, the construct captured by the TAGGED scale reflects a focused subset of observable instructional practices rather than the full multidimensional scope of teaching effectiveness discussed in the literature.

The separate Student Study Effort construct identified in the analysis reflects the behavioural investment students make in their academic work. Although conceptually distinct from teaching performance, student effort may be associated with the instructional environment and with students' experiences of teaching. Research on student engagement has similarly highlighted associations between students' behavioural engagement and the instructional and learning environments they experience (Fredricks et al., 2004; Kahu, 2013; Zhoc et al., 2019). The present findings reinforce this conceptual distinction, as the indicators of study effort formed a factor distinct from Instructional Practices. This result aligns with engagement theories that distinguish between students' engagement behavior and the instructional context of that engagement (Fredricks et al., 2004; Kahu, 2013; Kahu & Nelson, 2018). At the same time, the operationalisation of study effort in the present study captures primarily the behavioural dimension of engagement, rather than its cognitive or emotional components.

The separation of Student Study Effort from Instructional Practices is substantively important. The two-factor solution represents two distinct constructs, not two dimensions of a single teacher-effectiveness construct; students' study effort must not be interpreted as teacher effectiveness. Rather, study effort is an adjacent behavioural-engagement construct included to examine how students' own investment in learning coexists with their perceptions of instruction. Accordingly, institutional interpretation should report the two constructs separately and should not combine Student Study Effort with Instructional Practices into a global teacher-effectiveness score.

An important finding of the study concerns the stability of the measurement model across student groups. The invariance of the factor structure across gender suggests that male and female students interpret the items of the scale in a comparable manner and that the instrument measures the same underlying constructs across groups. Measurement invariance is a critical property for evaluation instruments used in HE, as it allows meaningful comparisons between

student populations and reduces the risk of systematic bias in teaching evaluations (Putnick & Bornstein, 2016). These findings support the use of the TAGGED scale for comparative analyses across male and female students within HE contexts.

A methodological consideration relates to the Student Study Effort factor, which is represented by two indicators. While psychometric guidelines typically recommend the use of multi-item constructs, previous research suggests that two-item measures may be appropriate when the construct being measured is narrowly defined and the indicators demonstrate adequate association (Bergkvist & Rossiter, 2007; Yong & Pearce, 2013). Their use nevertheless requires particular attention to reliability and construct coverage because fewer indicators provide a more restricted representation of the underlying construct (Zeller & Carmines, 1980). In the present study, the two indicators of study effort demonstrated a moderate association with each other while maintaining relatively low correlations with the remaining items of the scale, providing preliminary support for their interpretation as a specific behavioural dimension of student engagement distinct from perceptions of instructional practices.

Overall, the findings indicate that the TAGGED scale provides a concise and psychometrically sound instrument for assessing students' perceptions of instructional practices together with the separately measured construct of study effort. The results obtained have clear implications for the recent demand for simpler evaluation methods that are grounded in solid evidence and applicable in practice (Quansah et al., 2024; Weenink et al., 2024). The fact that the structure of the measure retains five items focused on instructional practices suggests that a small number of observable behaviors may be sufficient to form a meaningful factor representing a focused component of the broader construct of teaching effectiveness. The content of this factor aligns with specific instructional features previously identified in the HE literature (Tickell & Klassen, 2024; Zhang & Wu, 2025).

The emergence of Instructional Practices and Student Study Effort as distinct, potentially related constructs is theoretically consistent with contemporary engagement frameworks, which conceptualise learning as the result of reciprocal interactions between teaching environments, institutional conditions, and student participation (Kahu & Nelson, 2018; Mercer-Mapstone et al., 2017; Zhoc et al., 2019). The empirical differentiation of the two TAGGED constructs is consistent with the theoretical distinction in this literature between characteristics of the instructional environment and students' own behavioural engagement (Kahu & Nelson, 2018; Mercer-Mapstone et al., 2017; Zhoc et al., 2019). The present findings also complement recent evidence linking instructional quality with students' motivation, engagement, and perceptions of learning effectiveness across diverse HE settings (Miró-Colmenárez et al., 2025; Zhang & Wu, 2025). From this perspective, the scale contributes to ongoing discussions regarding how teaching quality can be operationalised in ways that remain theoretically meaningful while also being feasible for large-scale implementation.

Importantly, although the instrument was developed within the Greek HE system, the underlying constructs are not specific to the Greek context. Organisation of learning activities, clarity of instruction, stimulation of interest, encouragement of participation, and student study effort are aspects of instructional practice or student engagement discussed across diverse HE systems internationally (Mercer-Mapstone et al., 2017; Tickell & Klassen, 2024; Tight, 2024). The correspondence between the constructs represented in TAGGED and constructs identified in international research provides preliminary conceptual support for their potential relevance

beyond the Greek setting. This is pertinent to recent literature emphasising the value of interpretable indicators of teaching quality and student engagement across institutional and national contexts, particularly where student feedback contributes to quality assurance and enhancement processes (Ciuchi et al., 2024; Quansah et al., 2024; Weenink et al., 2024). However, conceptual correspondence with international literature does not establish cross-cultural measurement equivalence. Accordingly, the present findings should be interpreted as evidence of conceptual correspondence rather than evidence of cross-cultural equivalence.

The brevity of the instrument may be particularly relevant in contemporary HE environments where institutions increasingly depend on student evaluation systems while simultaneously facing declining response rates, concerns regarding evaluation fatigue, and ongoing debates regarding the validity and practical use of student feedback data (Perrett, 2021; Quansah et al., 2024; Sidwell et al., 2025). Recent scholarship has argued that sustainable evaluation systems require instruments that balance methodological rigour with practical usability in institutional settings (Ciuchi et al., 2024; Cook et al., 2024). The present findings contribute to this literature by showing that a seven-item instrument can retain a clearly interpretable factor structure and satisfactory psychometric performance in a large validation sample. TAGGED therefore illustrates the potential feasibility of combining brevity with psychometric adequacy, although its practical performance, including response rates, respondent burden, and sensitivity to change, should be evaluated directly in future institutional applications.

Practical Implications

The findings of this study have several implications for policy and practice within Greek HE. Earlier studies highlight concerns about the psychometric validity, interpretability, and effectiveness of student evaluations of teaching, not only in Greek HE (Papadogiannis et al., 2024; Quansah et al., 2024). To address this problem, TAGGED provides a practical tool that uses an effective yet simple instrument, psychometrically evaluated and based on student-observable instructional practices, while also taking into account Student Study Effort as an additional element of student engagement. In this way, the tool can support better and more effective evaluation practices in terms of institutional quality assurance.

This practical use is consistent with research emphasising that student-evaluation systems are most useful when they generate interpretable feedback for quality enhancement rather than functioning as stand-alone judgements of teaching quality (Cook et al., 2024; Quansah et al., 2024). In line with this literature, TAGGED should be viewed as part of a broader evaluation framework rather than as a standalone method for measuring teacher effectiveness. It should be interpreted alongside other complementary sources of evidence, such as peer evaluations, teaching portfolios, and other indicators of teaching efforts.

From an institutional perspective, the integration of the TAGGED scale into internal evaluation systems could support evidence-based decision-making related to teaching development initiatives. Because the Instructional Practices factor provides focused feedback on organisation, clarity, stimulation of interest, and encouragement of participation, repeated use of the instrument could help institutions identify specific areas for teaching development rather than relying solely on broad or global judgements of teaching quality. In addition, the brevity of the instrument may reduce respondent burden, an important practical consideration in evaluation systems in which survey fatigue and participation are recognised concerns (Quansah et al., 2024).

At the individual level, the scale may also serve as a useful tool for faculty reflection and professional development. By providing structured feedback on key aspects of instructional practice, the instrument can help teachers identify strengths in their teaching and areas where pedagogical approaches may be refined. This formative application is consistent with previous research advocating the use of student feedback to support teaching enhancement rather than as an isolated evaluative judgement (Cook et al., 2024; Quansah et al., 2024). TAGGED may therefore contribute to reflective and developmental evaluation processes while preserving the distinction between student-observable instructional practices and the broader, multidimensional construct of teacher effectiveness.

Theoretical implications

The present study contributes to the HE teaching and learning literature by demonstrating that key dimensions of instructional quality can be captured through a parsimonious set of student-reported indicators while maintaining conceptual coherence and empirical robustness. Although teaching effectiveness is widely recognised as a multidimensional construct encompassing pedagogical, relational, organisational, and contextual components (Klassen et al., 2023; Tight, 2024), the present findings suggest that students' evaluations of teaching are strongly anchored in a limited number of observable instructional practices, particularly instructional organisation, clarity of communication, intellectual stimulation, and opportunities for active participation. These findings support recent calls for evaluation frameworks that balance conceptual comprehensiveness with measurement efficiency and practical applicability within quality enhancement systems (Quansah et al., 2024; Weenink et al., 2024).

The identification of student study effort as a distinct yet related latent construct further contributes to contemporary understandings of the teaching–learning relationship. Current theoretical models increasingly conceptualise student engagement as a dynamic process emerging from interactions between learners and their educational environments rather than as a simple outcome of instructional quality (Kahu & Nelson, 2018; Mercer-Mapstone et al., 2017). Consistent with this perspective, the findings indicate that instructional practices and behavioural engagement represent interconnected but conceptually distinct constructs within the educational experience. The emergence of student study effort as a separate factor suggests that students differentiate between the quality of instructional provision and their own behavioural investment in learning, supporting their treatment as separate constructs that may be related, rather than interpreting study effort as a subdimension or indicator of teacher effectiveness.

More broadly, the study contributes to ongoing debates regarding the conceptualisation and operationalisation of teaching quality in HE. While many evaluation systems attempt to capture an extensive range of teaching dimensions, such approaches raise an important tension between conceptual breadth, respondent burden, and practical implementation, a concern also identified in research on questionnaire burden and contemporary student-evaluation systems (Cook et al., 2024; Rolstad et al., 2011; Sahlqvist et al., 2011). The present findings suggest that concise measures centred on observable instructional practices can retain conceptual coherence and satisfactory psychometric performance while reducing the number of indicators required for assessment. Consequently, the TAGGED framework offers a potential bridge between theoretical models of teaching effectiveness and the practical demands of contemporary quality assurance and quality enhancement processes in HE. However, the present results support parsimony only

within the explicitly delimited domain of student-observable instructional practices and should not be interpreted as demonstrating that five instructional items exhaust the broader multidimensional construct of teacher effectiveness.

Limitations and Future Research

Several limitations should be considered when interpreting the findings of the present study. First, TAGGED was developed and validated with undergraduate students aged 18–24 years, which limits the generalisability of its measurement properties to postgraduate students, mature students, or other academic-level groups. Although the large validation sample provides robust evidence within the population studied, it does not establish that the factor structure, reliability, or measurement properties of TAGGED are equivalent across populations with different educational or demographic characteristics. Future research should therefore test factorial validity and measurement invariance across academic levels, age groups, and different forms of higher-education provision.

Second, data were collected through a web-based survey distributed via university communication channels and student networks. Although online surveys facilitate efficient access to large samples, they may introduce self-selection bias because participation depends on students' access to digital platforms and willingness to engage with the survey (Van Mol, 2017). Consequently, students who are more engaged with institutional communication channels or more motivated to provide feedback may have been overrepresented in the sample. Future studies could strengthen the representativeness of findings using complementary recruitment strategies, including institutionally coordinated sampling procedures and in-class data collection.

Third, the present study relied exclusively on self-reported data. Student evaluations provide valuable insights into learners' experiences of teaching; however, they represent only one perspective on instructional quality and may be influenced by individual expectations, response tendencies, or contextual factors unrelated to teaching practices. Previous research has highlighted potential sources of bias in student evaluations, including leniency effects and systematic differences associated with instructor characteristics (Boring & Ottoboni, 2016; Spooren et al., 2013). Consequently, future research should examine the relationship between TAGGED scores and alternative indicators of teaching quality, including peer evaluations, classroom observations, teaching portfolios, and student learning outcomes, to establish a more comprehensive understanding of the instrument's validity.

Fourth, the Student Study Effort construct assessed by TAGGED is represented by two indicators. Although psychometric guidelines suggest that two-item measures may be appropriate when assessing narrowly defined constructs and when indicators demonstrate adequate association (Bergkvist & Rossiter, 2007), the limited number of indicators may constrain the breadth of the construct and its conceptual coverage. Future research could further refine this construct by incorporating additional indicators reflecting broader behavioural, cognitive, and self-regulatory aspects of engagement.

Fifth, a further limitation concerns construct coverage of the five-item Instructional Practices factor. The final retained items assess organisation, delivery/clarity, stimulation of interest, and participation, but they do not directly assess other plausible components of teacher effectiveness such as disciplinary content knowledge or the design and fairness of assessment. The strong

psychometric performance of a brief factor should therefore not be interpreted as evidence that these omitted domains are unimportant. Future extensions of TAGGED should test whether adding theoretically justified domains improves content coverage without undermining feasibility, and should compare the incremental validity of broader versus shorter versions.

Beyond these methodological considerations, an important direction for future research concerns the broader applicability of the TAGGED framework. Although the instrument was developed within the Greek HE context, the separate constructs of perceived instructional practices and student study effort reflect constructs that are widely recognised within international HE research. Future studies should therefore investigate the cross-cultural validity and measurement invariance of the scale across different countries, disciplinary fields, institutional settings, and languages. Such research would help determine the extent to which TAGGED can support comparative studies of teaching quality and student engagement and contribute to the development of more coherent and internationally comparable approaches to teaching evaluation in HE.

Conclusion

This study developed and validated the TAGGED scale, a concise student-reported instrument designed to assess perceived instructional practices and, separately, student study effort in higher education. The findings supported a stable two-factor structure, satisfactory reliability, construct validity, and measurement invariance across gender, providing evidence that the instrument offers a psychometrically robust yet practical approach to teaching evaluation.

By focusing on observable instructional practices and behavioural engagement, TAGGED responds to growing calls for evaluation frameworks that balance conceptual relevance, methodological rigour, and feasibility for large-scale implementation. The instrument offers a concise alternative to lengthier evaluation systems and may support institutional quality assurance, teaching enhancement initiatives, and evidence-informed decision-making in HE.

Accordingly, TAGGED is best interpreted as a focused course-level measure comprising a student-reported Instructional Practices factor and a separate Student Study Effort factor, rather than as an exhaustive global measure of teacher effectiveness.

Although developed within the Greek HE context, the distinct constructs assessed by TAGGED reflect instructional and engagement-related processes that are widely recognised across international HE systems. As such, the scale has the potential to contribute to comparative research on teaching quality and student engagement beyond national boundaries. Future research should examine its cross-cultural validity and measurement invariance across different educational systems, disciplinary contexts, and languages.

Acknowledgements

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all

individual participants included in the study. The data that support the findings of this study are available from the corresponding author upon reasonable request. The authors declare that they did not use artificial intelligence (AI) tools or AI-assisted technologies in the conception, writing, data analysis, interpretation of findings, or preparation of this manuscript.

The authors list the following CRediT contributions: **Mastrokoukou**: Conceptualization, Methodology, Investigation, Writing - original draft. **Seremetaki**: Investigation, Data curation, Formal analysis, Writing - review & editing. **Karali**: Investigation, Data curation, Resources, Writing - review & editing. **Settanni**: Methodology, Formal analysis, Validation, Writing - review & editing. **Ijaz**: Formal analysis, Validation, Data curation, Writing - review & editing. **Longobardi**: Conceptualization, Supervision, Project administration, Writing - review & editing.

References

Alzafari, K., & Ursin, J. (2019). Implementation of quality assurance standards in European higher education: Does context matter? *Quality in Higher Education, 25*(1), 58–75. https://doi.org/10.1080/13538322.2019.1578069

Ammigan, R., & Jones, E. (2018). Improving the student experience: Learning from a comparative study of international student satisfaction. *Journal of Studies in International Education, 22*(4), 283–301. https://doi.org/10.1177/1028315318773137

Bardach, L., & Klassen, R. M. (2020). Smart teachers, successful students? A systematic review of the literature on teachers' cognitive abilities and teacher effectiveness. *Educational Research Review, 30*, Article 100312. https://doi.org/10.1016/j.edurev.2020.100312

Baeten, M., Dochy, F., Struyven, K., Parmentier, E., & Vanderbruggen, A. (2016). Student-centred learning environments: An investigation into student teachers' instructional preferences and approaches to learning. *Learning Environments Research, 19*(1), 43–62. https://doi.org/10.1007/s10984-015-9190-5

Beausaert, S. A. J., Segers, M. S. R., & Wiltink, D. P. A. (2013). The influence of teachers' teaching approaches on students' learning approaches: The student perspective. *Educational Research, 55*(1), 1–15. https://doi.org/10.1080/00131881.2013.767022

Becket, N., & Brookes, M. (2006). Evaluating quality management in university departments. *Quality Assurance in Education, 14*(2), 123–142. https://doi.org/10.1108/09684880610662015

Behera, S. K., & Mukherjee, S. (2019). Assessment of teacher effectiveness of university teachers in Purulia District of West Bengal. *International Journal of Research in Teacher Education, 10*(2), 1–19. https://ijrte.inased.org/makale/926

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*(2), 175–184. https://doi.org/10.1509/jmkr.44.2.175

Blaskova, M., Blasko, R., Figurska, I., & Sokol, A. (2015). Motivation and development of the university teachers' motivational competence. *Procedia - Social and Behavioral Sciences, 182*, 116–126. https://doi.org/10.1016/j.sbspro.2015.04.746

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Cinches, M. F. C., Russell, R. L. V., Chavez, J. C., & Ortiz, R. O. (2017). Student engagement: Defining teacher effectiveness and teacher engagement. *Journal of Institutional Research South East Asia, 15*(1).

Ciuchi, O. M., Șerbănescu, L. E., Dobre, C. M., Georgescu, B. G., Țigănoaia, B. D., & Țucă, P. L. (2024). The impact of student evaluation of teaching staff on enhancing the quality of teaching in higher education in Romania. *Sustainability*, *16*(23), 10196. https://doi.org/10.3390/su162310196

Coertjens, L., Brahm, T., Trautwein, C., & Lindblom-Ylänne, S. (2017). Students' transition into higher education from an international perspective. *Higher Education, 73*(3), 357–369. https://doi.org/10.1007/s10734-016-0092-y

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* (2nd ed.). Pegem Akademi.

Cook, S., Watson, D., & Webb, R. (2024). Performance evaluation in teaching: Dissecting student evaluations in higher education. *Studies in Educational Evaluation*, *81*, 101342. https://doi.org/10.1016/j.stueduc.2024.101342

Drosos, D., Gkika, E. C., Kargas, A., Komisopoulos, F., & Ntanos, S. (2023). A multi-criteria analysis method for the evaluation of university students' satisfaction. In *The International Conference on Strategic Innovative Marketing and Tourism* (pp. 595–603). Springer. https://doi.org/10.1007/978-3-031-51038-0_64

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment, 29*(2), 136–151. https://doi.org/10.1177/1073191120957102

European Association for Quality Assurance in Higher Education. (2015). *Standards and guidelines for quality assurance in the European Higher Education Area (ESG)*.

Fernandes, E. L., & Flores, M. A. (2022). Assessment in higher education: Voices of programme directors. *Assessment & Evaluation in Higher Education, 47*(1), 45–60. https://doi.org/10.1080/02602938.2021.1888869

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.

Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science, 49*(2), 78–88. https://doi.org/10.1037/cbs0000069

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gupta, S., Grover, S., Menon, V., Indu, P. V., Vidhukumar, K., & Chacko, D. (2025). Item generation and establishing face and content validity of a rating scale: A primer. *Indian Journal of Psychiatry*, *67*(8), 816–822. https://doi.org/10.4103/indianjpsychiatry_750_25

Freire, P. (2007). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). Continuum. (Original work published 1970)

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson.

Hellenic Republic. (2005). *Law 3374/2005: Quality assurance in higher education, credit transfer and accumulation system, diploma supplement and other provisions*. *Government Gazette of the Hellenic Republic*, Series A, No. 189, 2 August 2005.

Hellenic Republic. (2011). *Law 4009/2011: Structure, operation, quality assurance of studies and internationalisation of higher education institutions*. *Government Gazette of the Hellenic Republic*, Series A, No. 195, 6 September 2011.

Hellenic Statistical Authority. (2025). *Higher education statistics: Academic year 2023–2024*. Hellenic Statistical Authority. https://www.statistics.gr

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education, 38*(5), 758–773. https://doi.org/10.1080/03075079.2011.598505

Kim, L. E., Jörg, V., & Klassen, R. M. (2019). A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. *Educational Psychology Review, 31*, 163–195. https://doi.org/10.1007/s10648-018-9458-2

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, *12*, 59-76. https://doi.org/10.1016/j.edurev.2014.06.001

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Krause, K. L., & Coates, H. (2008). Students' engagement in first-year university. *Assessment & Evaluation in Higher Education, 33*(5), 493–505. https://doi.org/10.1080/02602930701698892

Little, T. D. (2024). *Longitudinal structural equation modeling*. Guilford Publications.

López, L. F. G., & Dávila, M. G. V. (2019). La evaluación del desempeño docente en la educación superior. *Propósitos y Representaciones*, *7*(2), 479-515. https://doi.org/10.20511/pyr2019.v7n2.255

Mastrokoukou, S., & Crawford-Lee, M. (2023). Guest editorial: Profound digital pedagogies: Global perspectives. *Higher Education, Skills and Work-Based Learning, 13*(4), 645–648. https://doi.org/10.1108/HESWBL-08-2023-291

Mastrokoukou, S., Kaliris, A., Donche, V., Chauliac, M., Karagiannopoulou, E., Christodoulides, P., & Longobardi, C. (2022). Rediscovering teaching in university: A scoping review of teacher effectiveness in higher education. *Frontiers in Education, 7*, 861458. https://doi.org/10.3389/feduc.2022.861458

Mercer-Mapstone, L., Dvorakova, S. L., Matthews, K. E., Abbot, S., Cheng, B., Felten, P., ... & Swaim, K. (2017). A systematic literature review of students as partners in higher education. *International Journal for Students as Partners*, *1*(1). https://doi.org/10.15173/ijsap.v1i1.3119

Miró-Colmenárez, P. J., Durán-Alonso, S., Díaz-Cruces, E., & Zamora-Ledezma, C. (2025). Enhancing university teaching through student-led review articles as a pathway to early research engagement. *Education Sciences*, *15*(2), 249. https://doi.org/10.3390/educsci15020249

Organisation for Economic Co-operation and Development (OECD). (2011). *Education policy advice for Greece* (Strong Performers and Successful Reformers in Education). OECD Publishing. https://doi.org/10.1787/9789264119581-en

Papadogiannis, I., Vassilakis, C., Wallace, M., & Katsis, A. (2024). Challenges and Trends in Student Evaluation of Teaching: Analysis of SET Data from the University of Peloponnese. *Information*, *15*(9), 576. https://doi.org/10.3390/info15090576

Quansah, F., Cobbinah, A., Asamoah-Gyimah, K., & Hagan, J. E., Jr. (2024). Validity of student evaluation of teaching in higher education: A systematic review. *Frontiers in Education, 9*, Article 1329734. https://doi.org/10.3389/feduc.2024.1329734

Şahin Kürşad, M., İlgün Dibek, M., & Erdoğan, T. (2026). A multilevel meta-analysis of student evaluations of teaching and academic achievement. *Studies in Educational Evaluation, 90*, 101644. https://doi.org/10.1016/j.stueduc.2026.101644

Sidwell, D., Lee, D., Zimmerman, P. A., Bentley, S., & Barton, M. (2025). Teaching faculty experiences with student evaluation of instruction: A mixed-methods study. *Teaching and Learning in Nursing*, *20*(1), e276-e284. https://doi.org/10.1016/j.teln.2024.11.009

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598–642. https://doi.org/10.3102/0034654313496870

Strayhorn, T. L. (2018). *College students' sense of belonging: A key to educational success for all students*. Routledge.

Tickell, R. M., & Klassen, R. M. (2024). Developing the Teacher mentoring self-efficacy scale (TMSES) using the Delphi method and exploratory factor analysis. *Teaching and Teacher Education*, *139*, 104452. https://doi.org/10.1016/j.tate.2023.104452

Tight, M. (2024). The crisis literature in higher education. *Higher Education Quarterly*, *78*(4), e12504. https://doi.org/10.1111/hequ.12504

Trowler, V. (2010). Student engagement literature review. *The Higher Education Academy*, *11*(1), 1-15. https://doi.org/10.1080/13562517.2015.1016417

Van Mol, C. (2017). Improving web survey efficiency: the impact of an extra reminder and reminder content on web survey response. *International Journal of Social Research Methodology*, *20*(4), 317-327. https://doi.org/10.1080/13645579.2016.1185255

Weenink, K., Aarts, N., & Jacobs, S. (2024). Purposes and tensions in organising knowledge: trajectories of student evaluations in two research universities. *Assessment & Evaluation in Higher Education*, *49*(6), 810-823. https://doi.org/10.1080/02602938.2024.2314168

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247. https://doi.org/10.1016/j.ijresmar.2010.02.004

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). Guilford Press.

Zhang, J., & Wu, Y. (2025). Impact of university teachers' digital teaching skills on teaching quality in higher education. *Cogent Education*, *12*(1), 2436706. https://doi.org/10.1080/2331186x.2024.2436706

Zhoc, K. C., Webster, B. J., King, R. B., Li, J. C., & Chung, T. S. (2019). Higher education student engagement scale (HESES): Development and psychometric evidence. *Research in Higher Education, 60*(2), 219–244. https://doi.org/10.1007/s11162-018-9510-6

Appendix 1: The Greek version of the TAGGED questionnaire

Βαθμολογική Κλίμακα	1= Διαφωνώ απόλυτα	2=Διαφωνώ	3= Ούτεδιαφωνώ/ Ούτε συμφωνώ	4= Συμφωνώ	5= Συμφωνώ απόλυτα

	1	2	3	4	5
1. Οι στόχοι του μαθήματος ήταν σαφείς;	□	□	□	□	□
2. Η ύλη που καλύφθηκε ανταποκρινόταν στους στόχους του μαθήματος;	□	□	□	□	□
3. Η ύλη που διδάχθηκε ήταν καλά οργανωμένη;	□	□	□	□	□
4. Το εκπαιδευτικό υλικό που χρησιμοποιήθηκε βοήθησε στην καλύτερη κατανόηση του θέματος;	□	□	□	□	□
5. Τα εκπαιδευτικά βοηθήματα (σύγγραμμα, σημειώσεις, πρόσθετη βιβλιογραφία) χορηγήθηκαν εγκαίρως;	□	□	□	□	□
6. Πόσο ικανοποιητικό βρίσκετε το(α) κύριο βιβλίο(α) ή τις σημειώσεις;	□	□	□	□	□
7. Πόσο εύκολα διαθέσιμη είναι η βιβλιογραφία στην Πανεπιστημιακή Βιβλιοθήκη;	□	□	□	□	□
8. Πόσο απαραίτητα κρίνετε τα προαπαιτούμενα του μαθήματος, εάν υπάρχουν;	□	□	□	□	□
9. Σε ποιό βαθμό χρησιμοποιήθηκαν γνώσεις από άλλα μαθήματα;	□	□	□	□	□
10. Πώς κρίνετε το επίπεδο δυσκολίας του μαθήματος για το έτος του;	□	□	□	□	□
11. Πώς κρίνετε τον αριθμό Διδακτικών Μονάδων σε σχέση με τον φόρτο εργασίας;	□	□	□	□	□
12. Ο/η Καθηγητής/-τρια οργανώνει καλά την παρουσίαση της ύλης στα μαθήματα;	□	□	□	□	□
13. Ο/η Καθηγητής/-τρια επιτυγχάνει να διεγείρει το ενδιαφέρον για το αντικείμενο του μαθήματος;	□	□	□	□	□
14. Ο/η Καθηγητής/-τρια αναλύει και παρουσιάζει τις έννοιες με τρόπο απλό και ενδιαφέροντα, χρησιμοποιώντας παραδείγματα;	□	□	□	□	□
15. Ο/η Καθηγητής/-τρια ενθαρρύνει τους/τις φοιτητές/-τριες να διατυπώνουν απορίες και ερωτήσεις, για να αναπτύξουν την κρίση τους;	□	□	□	□	□
16. Ο/η Καθηγητής/-τρια ήταν συνεπής στις υποχρεώσεις του/της (παρουσία στα μαθήματα, έγκαιρη διόρθωση εργασιών ή εργαστηριακών αναφορών, ώρες συνεργασίας με τους/τις φοιτητές/-τριες);	□	□	□	□	□
17. Ο/η Καθηγητής/-τρια αξιοποιεί τις νέες τεχνολογίες, π.χ.: e-class, για να διευκολύνει την επικοινωνία για εκπαιδευτικούς λόγους;	□	□	□	□	□
18. Ο/η Καθηγητής/-τρια είναι γενικά προσιτός/-ή στους/στις φοιτητές/-τριες;	□	□	□	□	□
19. Παρακολουθώ τακτικά τις διαλέξεις.	□	□	□	□	□
20. Μελετώ συστηματικά την ύλη.	□	□	□	□	□

21. Αφιερώνω εβδομαδιαία για μελέτη του συγκεκριμένου μαθήματος: 1= <2 Ώρες, 2=2-4 Ώρες, 3=4-6 Ώρες, 4=6-8 Ώρες, 5= >8 Ώρες	□	□	□	□	□

Appendix 2: The English version of the "Teacher Assessment in Greek Higher Education Development (TAGGED)" questionnaire

Rating scale	**1= I strongly disagree**	**2=I disagree**	**3= Neither disagree nor agree**	**4= I agree**	**5=I strongly agree**

	1	2	3	4	5
1. The aims of the course are clear.	□	□	□	□	□
2. The proposed curriculum meets the aims of the course.	□	□	□	□	□
3. The content being taught was well organized.	□	□	□	□	□
4. The material delivered helped in understanding better the subject.	□	□	□	□	□
5. The course material (main coursebook, course/lecture notes, tutorial material. additional bibliography) was delivered on time.	□	□	□	□	□
6. The main coursebook(s) is/are satisfactory.	□	□	□	□	□
7. The bibliography is available in the University library.	□	□	□	□	□
8. How necessary do you consider the prerequisites of the course, if any?	□	□	□	□	□
9. To what extent was knowledge from other courses used?	□	□	□	□	□
10. The course's level of difficulty is appropriate for an undergraduate program.	□	□	□	□	□
11. How do you evaluate the number of ECTS in relation to the workload (i.e., studying hours, etc.)?	□	□	□	□	□
12. The teacher organizes the delivery of the curriculum well.	□	□	□	□	□
13. The teacher stimulates interest in the topic of the Course.	□	□	□	□	□
14. The teacher delivers course-related concepts in an effective way.	□	□	□	□	□
15. The teacher encourages students to participate and express their questions.	□	□	□	□	□

16. The teacher was consistent in their obligations (attendance at classes, timely correction of assignments or laboratory reports, hours of collaboration with students).	□	□	□	□	□
17. The teacher makes use of new technologies e.g., e-class to facilitate communication for educational purposes.	□	□	□	□	□
18. The teacher is accessible to students.	□	□	□	□	□
19. I attend frequently the course's total lectures and tutorials.	□	□	□	□	□
20. I study frequently.	□	□	□	□	□
21. Hours of study per week: 1= <2 hours, 2=2-4 hours , 3=4-6 hours , 4=6-8 hours , 5= >8 hours	□	□	□	□	□