Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: https://www.science-gate.com/IJAAS.html

Development and validation of a scale for evaluating STEM faculty teaching effectiveness in higher education

Robert Jay N. Angco *, Ma. Lubella B. Angco

Department of Mathematics and Statistics, Cebu Technological University, Cebu, Philippines

ARTICLE INFO

Article history:
Received 2 September 2025
Received in revised form
10 January 2026
Accepted 19 January 2026

Keywords:
Teaching effectiveness
STEM education
Scale development
Outcomes-based education
Higher education

ABSTRACT

This study aimed to develop and validate a scale for assessing STEM faculty teaching effectiveness and improving educational outcomes in STEM disciplines. The study is grounded in the principles of Outcomes-Based Education (OBE) and Outcomes-Based Teaching and Learning (OBTL) and aligns intended learning outcomes, teaching and learning activities, and assessment tasks. A sequential exploratory mixed-methods design, with qualitative methods followed by quantitative methods, was employed. Administrators and faculty members participated in the development and validation of the STEM Faculty Teaching Effectiveness Scale for higher education. The scale showed excellent inter-rater agreement, as indicated by Cohen's Kappa. Exploratory factor analysis identified three main factors: Communicating Intended Learning Outcomes, Facilitating Teaching and Learning Activities, and Implementing Assessment Tasks. Confirmatory factor analysis using the Maximum Likelihood method was then conducted. The results showed good model fit for Model 2 (CFI = 0.9260; TLI = 0.9170; RMSEA = 0.0610; SRMR = 0.0471; χ^2/df = 1.8070), supporting the three-factor structure. Reliability analysis indicated high internal consistency, with a Cronbach's alpha value of 0.953. The final instrument consists of 28 items and is recommended for use by higher education institutions to evaluate STEM faculty teaching effectiveness.

1. Introduction

Evaluating Science, Technology, Engineering, and Mathematics (STEM) faculty teaching effectiveness is critical in higher education, as it directly impacts several aspects, including faculty performance, student learning outcomes, research quality, and community participation. Such rigorous evaluation identifies effective educators and maintains teaching excellence standards.

Evaluating faculty teaching effectiveness in STEM, particularly within Outcomes-Based Education (OBE) and Outcomes-Based Teaching and Learning (OBTL) frameworks, requires acknowledging its complex and multidimensional nature. Effectiveness is determined not only by content mastery but also by how well faculty communicate learning outcomes, facilitate learning activities, and assess student performance in alignment with desired outcomes. Despite recognition of these complexities, quantifying these dimensions for practical use in personnel decisions remains challenging.

In STEM education, OBE adoption holds particular significance as it necessitates a constructivist approach that engages students in active learning experiences (Allen et al., 2016). OBE shifts the focus from lecturers as content deliverers to facilitators, enabling students to drive learning through explicit objectives and collaboration. This approach integrates theoretical knowledge with practical skills, preparing students for workforce demands. OBE-based teaching evaluation, therefore, analyzes material delivery, active learning facilitation, and outcome attainment, aligning education with professional requirements (Asim et al., 2021).

In the Philippines, most higher education institutions have shifted toward OBE in response to demands for standardization and quality improvement (Mufanti et al., 2024). The Commission on Higher Education (CHED) mandates these standards through CMO No. 46, series 2012, requiring curriculum innovation aligned with current trends. OBTL emphasizes explicit learning

* Corresponding Author.
Email Address: robertjay.angco@ctu.edu.ph (R. J. N. Angco)
https://doi.org/10.21833/ijaas.2026.01.024
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-5321-2371

outcomes and uses constructive alignment to match teaching with specified requirements.

Although existing scales have been developed to evaluate STEM faculty teaching effectiveness (Landrum et al., 2017; Karpudewan et al., 2022; Yang et al., 2023), these instruments often fail to capture the unique aspects of STEM instruction within an OBE/OBTL framework. For instance, Yang et al. (2023) explored self-efficacy among integrated STEM activity implementers, yet their scale may not fully represent the diverse experiences of STEM teachers.

These limitations underscore the need for a tailored evaluation tool that captures the specific alignment of learning outcomes, teaching activities, and assessments characteristic of effective STEM teaching. The current study addresses this gap by developing and validating a scale grounded in OBE and OBTL principles, specifically designed to evaluate STEM faculty teaching effectiveness in higher education. This scale provides actionable insights for faculty development, enhancing the practical utility of teaching evaluations

2. Research methodology

This section presents a detailed description of the research methodology employed in this study. It covers the research design, environment, respondents, instruments, the data-gathering process, data analysis, and ethical considerations that govern the conduct of the study.

This study utilized the sequential exploratory (QUAL→quan) mixed method research design to develop and validate an instrument to evaluate STEM faculty teaching effectiveness in higher education institutions. In this study, the development of the scale followed the practical guidelines of Barry et al. (2011). The scale underwent four development and validation stages: (1) defining constructs, (2) developing the scale design and structure, (3) generating sample items, and (4) pretesting the scale. These steps characterize the sequential exploratory mixed-methods design, in which qualitative data were first collected and analyzed, and emerging themes were used to develop a quantitative scale.

In the initial qualitative phase, purposive sampling was employed to select participants, focusing on administrators and faculty members who played a key role in implementing OBTL. The aim was to ensure that the selected participants had sufficient experience and insight into the OBTL framework, which is critical for gathering relevant and meaningful data for scale development.

In the subsequent quantitative phase, random sampling was used to gather a representative sample of STEM faculty across various higher education institutions. This method ensured that the sample accurately reflected the diversity of STEM disciplines, institution types, and faculty experience levels. The scale was administered to this sample to evaluate its reliability and validity through statistical analyses, including confirmatory factor analysis and a path diagram. The research design included a path diagram to visually represent the relationships between the constructs derived from the qualitative phase and their corresponding items in the quantitative scale. The path diagram clearly illustrates how the constructs interconnect, providing a visual framework that supports the developed scale's overall structure and coherence.

During the preliminary phase of this study, data saturation was achieved with 10 respondents, including the vice-president for academic affairs, college deans, academic chairpersons, and faculty members. These interviews provided insights into their conceptions regarding the OBTL. After generating a pool of potential items for the scale, feedback from a group of experts in the field was used to assess the content validity of the items. For the quantitative part, the scale was administered to a larger and more diverse sample of administrators and faculty members to ensure robustness in assessing its reliability and validity (Table 1).

The qualitative sample included participants from diverse positions and institution types, ensuring multiple perspectives on OBTL implementation in Philippine STEM higher education. State universities represented 60% of the sample, with private institutions comprising 40%. All participants had at least five years of experience in STEM education and direct involvement in OBTL implementation.

Table 1: Demographic characteristics of qualitative phase participants (n = 10)

Participant	Position	Institution type	STEM discipline	Years of experience
P1	Vice president for academic affairs	State university	Engineering	15+
P2	College dean	Private university	Science	12
P3	College dean	State university	Mathematics	10
P4	College dean	State university	Technology	14
P5	Department chairperson	Private university	Engineering	8
P6	Department chairperson	State university	Science	9
P7	Department chairperson	Private university	Mathematics	7
P8	Faculty member	State university	Engineering	6
P9	Faculty member	Private university	Technology	5
P10	Faculty member	State university	Science	8

Specific identifiers removed to protect participant confidentiality

The item generation began with a thorough literature review conducted before the interviews. This review encompassed related works, recent studies, and widely recognized theoretical frameworks in the academic community, highlighting the critical aspects of OBTL.

Item generation was grounded in comprehensive qualitative research. Semi-structured interviews were conducted with ten participants purposively selected based on their involvement in STEM education and OBTL implementation. Participants included one vice-president for academic affairs, three college deans, three academic chairpersons, and three faculty members from both state and private universities across the Philippines. Each interview lasted 45–60 minutes and was conducted either face-to-face or online via Zoom/Google Meet, based on participant convenience. All interviews were audio-recorded with consent and transcribed verbatim. Interview transcripts were analyzed using thematic analysis following Braun and Clarke (2006). The analysis involved systematic coding of data to identify patterns and themes relevant to OBTL implementation. This process resulted in three primary themes that aligned with OBTL principles: (1) Communicating Intended Learning Outcomes, (2) Facilitating Teaching-Learning Activities, and (3) Implementing Assessment Tasks. Integrating insights from the literature review with the qualitative data collected through interviews was a critical step in the item generation process, ensuring that the item pool reflected both the theoretical and practical dimensions of OBTL. Table 2 illustrates how insights from participant interviews were systematically transformed into specific scale items, demonstrating the empirical grounding of the instrument in STEM educators' lived experiences with OBTL implementation.

Table 2: Examples of the theme-to-item transformation process

Construct	Representative Interview Quote	Participant	Derived Scale Item
Communicating intended learning outcomes (ILO)	In terms of communicating our intended learning outcomes to our students, we have to define clear outcomes; that's the first thing we have to do. These outcomes must be aligned with national and international standards in STEM education.	Informant C (Dean)	ILO1: The faculty clearly communicates the intended learning outcomes at the beginning of the course. ILO2: The faculty aligns course objectives with national standards in STEM education. ILO3: The faculty aligns course objectives with international standards in STEM education.
Facilitating teaching–learning activities (TLA)	It is not only one-on-one learning, but teamwork is encouraged because this allows us to observe and assess students' knowledge and learning in groups. My classes include project-based learning, laboratory and hands-on activities, and competency demonstrations, which are necessary to deliver meaningful learning.	Informant A (Vice president for academic affairs); Informant F (Faculty)	TLA1: The faculty promotes collaborative learning by engaging students in group projects. TLA2: The faculty designs teaching activities that emphasize practical and real-world applications of STEM concepts. TLA3: The faculty facilitates hands-on and laboratory-based activities to enhance student learning. TLA4: The faculty helps students apply theoretical knowledge to practical situations.
Implementing assessment tasks (AT)	We use both formative and summative assessments for our students. They are evaluated using rubrics, and aligning assessments with intended learning outcomes is essential.	Informant C (Dean); Informant J (Faculty)	AT1: The faculty regularly assesses students' understanding through formative assessments. AT2: The faculty uses summative assessments to evaluate overall student learning. AT3: The faculty uses clearly defined rubrics to assess student performance. AT4: The faculty aligns assessment tasks with the course learning objectives.

In the validation phase, the researchers sought assistance from identified experts to evaluate the content validity of the items that assess STEM faculty teaching effectiveness in the implementation of OBTL. Initial constructs of the instrument for assessing STEM faculty effectiveness emerged from administrator and faculty conceptions based on the interviews and OBTL literature review. Additionally, content experts confirmed that the items could effectively assess STEM faculty teaching effectiveness in OBTL.

The development of content validity followed a systematic two-stage approach, comprising two separate established methods. The first part involved a qualitative review of the instrument by five experts who provided feedback based on their professional judgment on the rationale of each item under the three constructs: (1) Communicating Intended Learning Outcomes, (2) Facilitating Teaching-Learning Activities, and (3) Implementing Assessment Tasks. Each item was then reviewed by the experts on its level of content alignment to the specified construct.

The first version of the scale consisted of 40 items. Of the 40 items, eight were considered inadequate. The inadequate items were item number 4 for the first construct and items 14, 17, 18, 20, 22, and 25 for the second construct. For the third construct, one item was considered inadequate. These eight items were removed, resulting in a second version of the instrument. After content validation, experts suggested additional revisions beyond the initial eight items, addressing redundancy, clarity, and construct alignment. The experts identified additional redundant, unclear, or misaligned items with the constructs they were meant to measure. Finally, experts deemed most items significant for evaluating STEM faculty teaching effectiveness, resulting in the final 28-item instrument validated through subsequent exploratory and confirmatory factor analyses.

3. Data analysis

Kappa Statistic (κ) was used to determine the consensus index of inter-rater agreement beyond chance levels, complementing the Content Validity Index (CVI). For construct validity, exploratory factor analysis was performed using principal component analysis (PCA) for factor extraction and Varimax for factor rotation. This study also employed the Average Variance Extracted (AVE) approach. To confirm construct distinctiveness, AVE was calculated for each construct. Reliability was tested after establishing validity evidence for the scores obtained from the instrument.

4. Results

Exploratory factor analysis (EFA) is a statistical test used to re-express many observed variables in terms of fewer factors when variables are presumed to share common underlying factors. Data suitability for structure detection is indicated by the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's Test of Sphericity (BTS). The KMO value should exceed 0.6 (preferably close to 1.0), and the BTS should be significant ($p < .05$) (Hair et al., 2010). Furthermore, the closer the KMO value is to 1.0, the more closely correlated the variables. The results revealed KMO = 0.933, χ^2 = 3864.821, $p < .001$. The KMO statistic indicates excellent sampling adequacy, while the computed χ^2 is significant.

Table 3 presents the 28 items with their communalities. Communalities refer to the proportion of the variance of each variable that the factors can explain. The values under the Extraction column are all greater than 0.2. This means these items were retained, as their communalities satisfied the threshold value. Table 4 reveals three factors based on eigenvalues greater than 1. These factors explained 62.053% of the variance in the original variables. The remaining items should be able to explain at least 50% of the total variance in the final check of the factor solution. Factor solutions were examined using Varimax rotation.

The pattern matrix from the principal component analysis with Varimax rotation revealed the factor structure of the OBTL instrument. Table 5 presents the factor loadings for all 28 items across the three extracted components, with loadings above 0.30 considered substantive. The rotation converged in 5 iterations, indicating a stable factor solution.

Table 3: Communalities of the 28 items

Item	Extraction
ILO1	0.653
ILO2	0.724
ILO3	0.653
ILO4	0.692
TLA1	0.339
TLA2	0.533
TLA3	0.509
TLA4	0.586
TLA5	0.649
TLA6	0.554
TLA7	0.602
TLA8	0.512
TLA9	0.542
TLA10	0.539
TLA11	0.526
TLA12	0.665
TLA13	0.536
TLA14	0.530
TLA15	0.492
AT1	0.575
AT2	0.576
AT3	0.566
AT4	0.592
AT5	0.623
AT6	0.500
AT7	0.684
AT8	0.548
AT9	0.625

Extraction method: Principal component analysis

Table 4: Total variance explained in the factor analysis

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	12.411	44.325	44.325	12.411	44.325	44.325	7.480	26.713	26.713
2	2.420	8.642	52.967	2.420	8.642	52.967	5.405	19.304	46.017
3	1.294	4.620	57.587	1.294	4.620	57.587	3.240	11.571	57.587
4	1.250	4.466	62.053						
5	.901	3.219	65.272						
6	.835	2.984	68.255						
7	.759	2.712	70.967						
8	.755	2.697	73.664						
9	.659	2.354	76.018						
10	.591	2.111	78.129						
11	.561	2.004	80.133						
12	.525	1.874	82.008						
13	.506	1.808	83.816						
14	.469	1.677	85.493						
15	.467	1.668	87.161						
16	.405	1.447	88.607						
17	.397	1.417	90.024						
18	.366	1.306	91.330						
19	.321	1.147	92.477						
20	.303	1.083	93.561						
21	.296	1.056	94.617						
22	.287	1.025	95.642						
23	.249	.889	96.530						
24	.242	.864	97.394						
25	.207	.740	98.135						
26	.197	.704	98.839						
27	.174	.623	99.462						
28	.151	.538	100.000						

The results from the pattern matrix analysis provide detailed insights into the structure of the Outcome-Based Teaching and Learning (OBTL) framework, comprising three constructs: Intended

Learning Outcomes (ILO), Teaching and Learning Activities (TLA), and Assessment Tasks (AT). This analysis, conducted using Principal Component Analysis and Varimax rotation, demonstrates a clear and distinct factor structure for each dimension.

Table 5: Pattern matrix: Items and factor loadings for the 3 dimensions of OBTL

Items	Component		
	1	2	3
ILO1	.187	.303	.726
ILO2	.323	.275	.737
ILO3	.331	.198	.710
ILO4	.247	.393	.691
TLA1	.420	.214	.342
TLA2	.677	.161	.221
TLA3	.621	.292	.197
TLA4	.717	.254	.085
TLA5	.748	.259	.150
TLA6	.638	.212	.319
TLA7	.711	.234	.206
TLA8	.677	.142	.181
TLA9	.630	.266	.272
TLA10	.636	.362	.063
TLA11	.661	.232	.190
TLA12	.766	.225	.164
TLA13	.689	.144	.200
TLA14	.664	.161	.250
TLA15	.676	.181	.042
AT1	.253	.704	.123
AT2	.273	.706	.040
AT3	.275	.628	.309
AT4	.212	.736	.068
AT5	.260	.707	.234
AT6	.205	.620	.271
AT7	.228	.734	.304
AT8	.162	.627	.358
AT9	.242	.708	.255

Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser normalization

The Communicating Intended Learning Outcomes factor is defined by four items associated with Intended Learning Outcomes, with loadings from 0.691 to 0.737. The high loadings of ILO2 (0.737) and ILO1 (0.726) indicate that these items are strong indicators of the learning goals and objectives that students are expected to achieve. These loadings suggest that the items within this component effectively capture the intended learning outcomes, providing a precise measure of the knowledge, skills, and competencies that students should acquire. The slightly lower, yet still strong, loadings of ILO3 (0.710) and ILO4 (0.691) further support the reliability of this component. These items encompass various learning domains, including cognitive, affective, and psychomotor outcomes, ensuring a comprehensive evaluation of student learning.

The second component encompasses nine items related to Assessment Tasks, all showing high loadings between 0.620 and 0.736. The strongest indicators, AT4 (0.736) and AT7 (0.734), reflect key elements of assessment practices, including formative and summative assessments that gauge student learning and performance. The high loadings across all items (e.g., AT9, AT5, AT2) indicate that this component comprehensively captures the essence of assessment tasks. The consistency of these loadings suggests that the items are well-aligned with the construct of assessment tasks, ensuring that the component reliably measures the various facets of how student learning is evaluated. This might include traditional tests, performance-based assessments, and other evaluative methods contributing to a holistic understanding of student achievement.

This last component is defined by 15 items related to TLA, with factor loadings ranging from 0.420 to 0.766. The high loadings of TLA12 (0.766), TLA5 (0.748), and TLA4 (0.717) indicate that these items are particularly strong indicators of teaching and learning activities. These items likely reflect key aspects of the instructional methods and learning experiences provided in the educational setting. The consistent loadings above 0.6 for most items (e.g., TLA7, TLA13, TLA8) suggest a robust internal consistency within this component. Even the lower loading of TLA1 (0.420), although lower than the others, still signifies a meaningful contribution to the dimension. The range of loadings within this component highlights the diversity of teaching and learning activities included in your model, from highly structured methods to more flexible, student-centered approaches.

Figs. 1 and 2 present the path diagrams illustrating the relationships among three latent constructs. The model includes factor loadings for the observed variables, path coefficients between the latent constructs, and correlations among the latent constructs.

The initial path diagram (Model 1) includes latent constructs Communicating ILO, Facilitating TLA, and Implementing AT. This model does not include covariances among the error terms of the observed variables. Model 2 is the refined path diagram in which additional covariances among the error terms of observed variables were included to improve model fit. Initially, fit indices for Model 1 were examined. Common fit indices include Chi-square, RMSEA, CFI, and TLI. Acceptable thresholds are often Chi-square (non-significant, though difficult with large samples), RMSEA (< 0.06), CFI (> 0.95), and TLI (> 0.95). Model 1 review revealed that some fit indices did not meet acceptable thresholds.

In Model 2, covariances among the error terms of observed variables were introduced. Including these covariances improved model fit, evidenced by better-fit indices (e.g., lower Chi-square, lower RMSEA, higher CFI, and TLI). With these additions, Model 2 shows improved fit indices, such as RMSEA = 0.05, CFI = 0.95, and TLI = 0.93, indicating a better-fitting model. For Model 2, the factor loadings for ILO ranged from 0.82 to 0.89, indicating a strong relationship between the observed variables (ILO1 to ILO4) and the latent construct "Communicating Intended Learning Outcomes." Factor loadings above 0.70 are considered high and indicate a good measure of the construct (Hair et al., 2010). This indicates that ILO1 to ILO4 were good indicators of the ILO construct. On the other hand, the factor loadings for TLA range from 0.61 to 0.82, with most loadings above 0.70, indicating a strong relationship between the observed variables (TLA1 to TLA15) and the latent construct "Facilitating Teaching-Learning Activities." Hair et al. (2010) also suggested that loadings above 0.60 are acceptable in social sciences, making these values generally acceptable. This suggests that these items are strong indicators of the TLA construct. Finally, the factor loadings for AT range from 0.68 to 0.82, indicating a strong relationship between the observed variables (AT1 to AT9) and the latent construct "Implementing Assessment Tasks." This means that the AT1 to AT9 are good indicators of the AT construct. Again, factor loadings above 0.70 are considered strong, and values above 0.60 are acceptable, according to Hair et al. (2010).

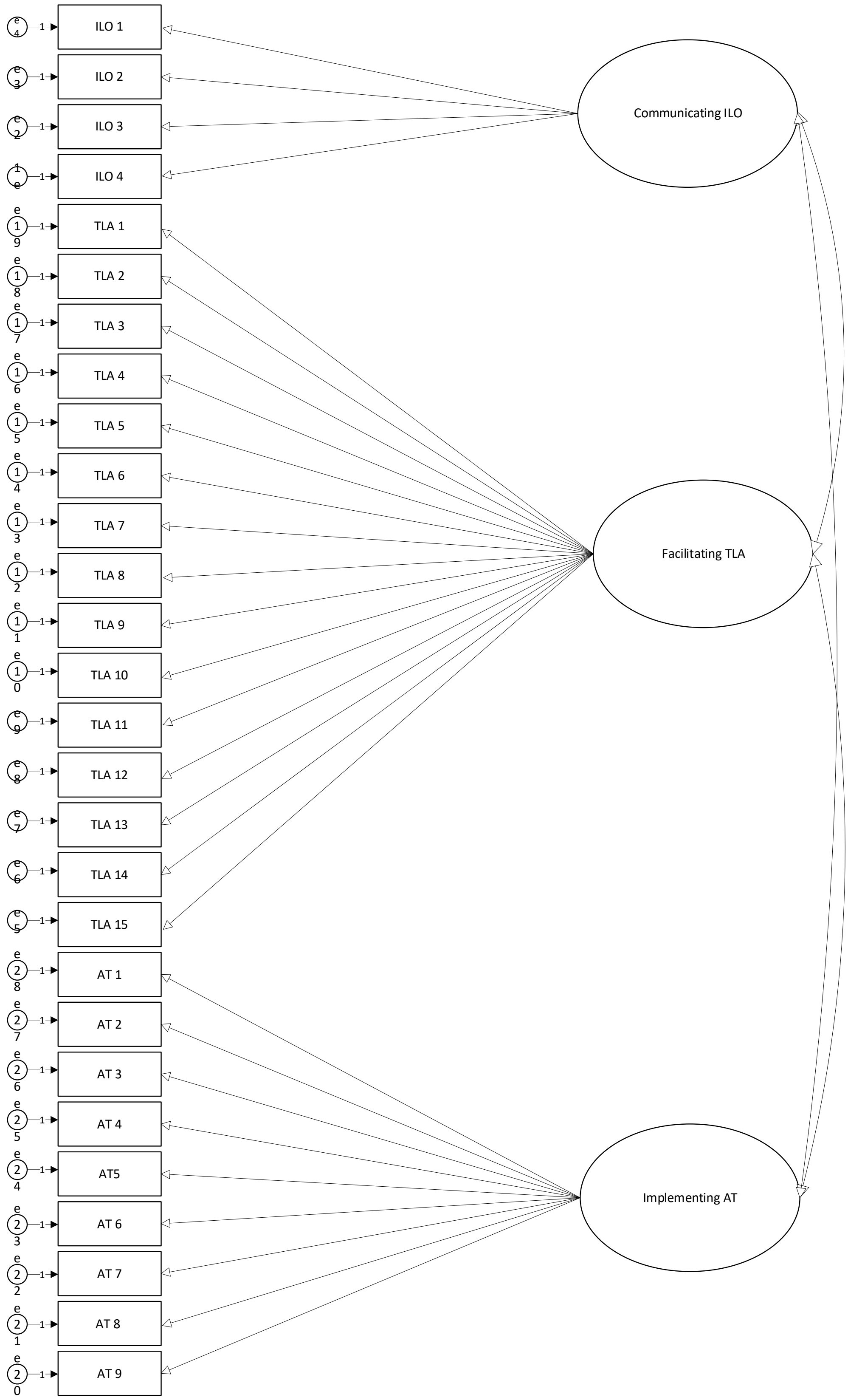

Fig. 1: Path diagram for model 1

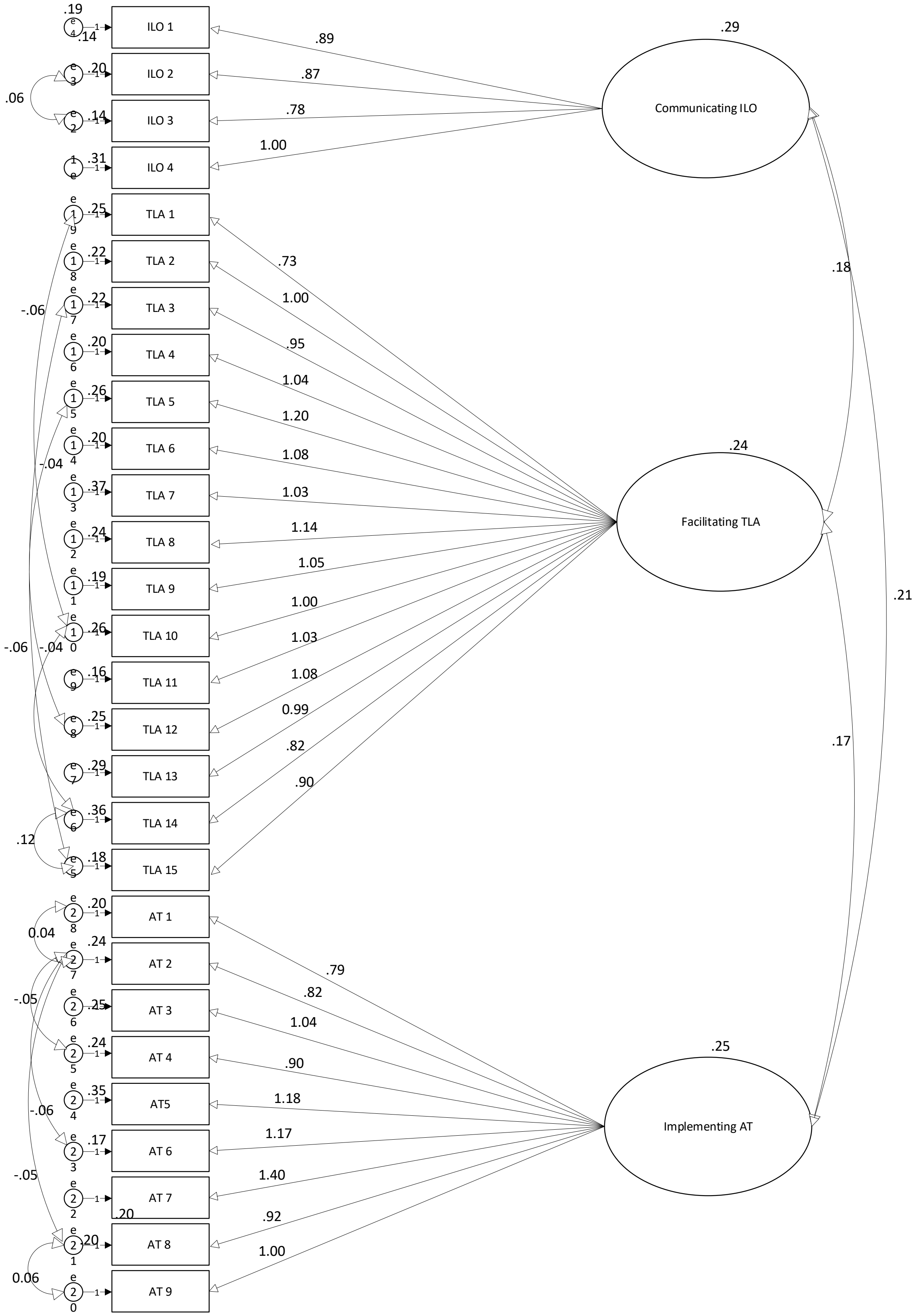

Fig. 2: Path diagram for model 2

Table 6 shows the model fit indices that determined how well each proposed model represented the data. Model 1 showed a Comparative Fit Index (CFI) of 0.881. According to Hair et al. (2010), a CFI greater than 0.90 is acceptable, indicating that Model 1 was only 0.019 below the acceptable threshold. The Tucker-Lewis Index (TLI) for Model 1 is 0.871. Hair et al. (2010) suggested that a TLI greater than 0.90 indicates an acceptable fit, indicating that Model 1 approached this criterion with acceptable fit. The Root Mean Square Error of Approximation (RMSEA) for Model 1 is 0.076. The RMSEA of 0.076, below the 0.08 threshold, indicated reasonable fit, so Model 1 is within this range, signifying an acceptable fit. The Standardized Root Mean Square Residual (SRMR) of

Model 1 is 0.0516. An SRMR of 0.08 or less indicates a good fit, indicating that Model 1 fits well within this threshold and is acceptable. Finally, the χ^2/df ratio for Model 1 was 2.225. Hair et al. (2010) considered a ratio below 3.00 to indicate a good fit, which means Model 1 meets this criterion, suggesting a good fit overall. In contrast, Model 2 demonstrates superior fit indices across all metrics. The CFI for Model 2 is 0.926, surpassing the 0.90 threshold and thus indicating a very good fit. Model 2 has a TLI of 0.917, above the 0.85 threshold suggested by Hair et al. (2010), indicating that Model 2 is a good fit. The RMSEA for Model 2 is 0.061, below the 0.08 threshold, suggesting a good fit. Model 2's SRMR is 0.0471, which, according to Hair et al. (2010), should be 0.08 or less for a good fit, signifying that Model 2 meets this criterion. Additionally, the Chi-square/df ratio for Model 2 is 1.807. Hair et al. (2010) considered a ratio below 3.00 to indicate a good fit, so Model 2 satisfies this condition. Both models exhibit satisfactory fit indices when compared to the proposed threshold values. Nevertheless, Model 2 consistently outperformed Model 1. Specifically, Model 2's CFI, TLI, RMSEA, SRMR, and Chi-square/df ratio all indicate a better fit than Model 1. Therefore, based on these indices, Model 2 is the preferred model as it better meets the standards for good model fit. Based on the overall model fit indices, there is clear evidence supporting the scale's construct validity for evaluating STEM faculty teaching effectiveness. Hence, the three-factor model identified through EFA was confirmed through CFA.

Table 6: Model data fit indices results

Model fit indices	Proposed threshold value	Model 1	Model 2
CFI	> 0.90	0.8810	0.9260
TLI	> 0.90	0.8710	0.9170
RMSEA	< 0.08	0.0760	0.0610
SRMR	≤ 0.08	0.0516	0.0471
χ^2/df	< 3.00	2.2250	1.8070

Table 7 shows the convergent and internal consistency results of the scale. For Communicating Intended Learning Outcomes, the standardized factor loadings for these items ranged from 0.686 to 0.823, indicating that the items adequately reflected the characteristics of each construct, as all loadings are above the commonly accepted threshold of 0.50 (Gefen et al., 2000). The Average Variance Extracted (AVE) is 0.577838, which exceeds the threshold of 0.50, indicating that the items in the scale reflect the characteristics of each research variable in the model (Srinivasan et al., 2002). This indicates good convergent validity. The Composite Reliability (CR) is 0.84499, well above the acceptable threshold of 0.70, indicating a high level of internal consistency. Thus, the scale demonstrated acceptable convergent validity.

The high internal consistency and convergent validity across constructs suggested that the scale reliably assessed STEM faculty teaching effectiveness. The following are the Cronbach's Alpha results for each of the constructs: Communicating Intended Learning Outcomes (α = 0.856), Facilitating Teaching-Learning Activities (α = .935),Implementing Assessment Tasks (α = 0.906), and the overall reliability (α = 0.953).

Table 7: Convergent and internal consistency results of the scale

Constructs	Item	Standardized factor loading	Average variance extracted	Composite reliability
Communicating intended learning outcomes	ILO1	0.7450	0.577838	0.84499
	ILO2	0.7800		
	ILO3	0.6860		
	ILO4	0.8230		
Facilitating teaching-learning activities	TLA1	0.5360	0.49890	0.936795
	TLA2	0.7020		
	TLA3	0.7030		
	TLA4	0.7340		
	TLA5	0.7970		
	TLA6	0.7200		
	TLA7	0.7480		
	TLA8	0.6760		
	TLA9	0.7200		
	TLA10	0.7130		
	TLA11	0.7040		
	TLA12	0.7980		
	TLA13	0.6930		
	TLA14	0.6820		
	TLA15	0.6290		
Implementing assessment tasks	AT1	0.6850	0.523592	0.907892
	AT2	0.6760		
	AT3	0.7340		
	AT4	0.6660		
	AT5	0.7740		
	AT6	0.7030		
	AT7	0.7980		
	AT8	0.7150		
	AT9	0.7500		

Table 8 shows the discriminant validity results of the scale. The results suggest that the three constructs (ILO, TLA, and AT) demonstrate good discriminant validity. This is indicated by the fact that the √AVE for each construct is greater than the correlations between that construct and the others. Specifically, the √AVE values for ILO (0.760156), TLA (0.706322), and AT (0.723596) are higher than the correlations between ILO and TLA (0.622), ILO and AT (0.652), and TLA and AT (0.633). This result confirmed that each construct was distinct from the others and effectively measured its unique dimension, as intended by the measurement model.

Table 8: Discriminant validity results

Variable	ILO	TLA	AT
ILO	1		
TLA	0.622	1	
AT	0.652	0.633	1
AVE	0.577838	0.49889	0.523592
$\sqrt{AVE}$	0.760156	0.706322	0.723596

5. Discussion

This study successfully developed and validated a comprehensive scale for evaluating STEM faculty teaching effectiveness within OBE and OBTL frameworks. The resulting 28-item instrument demonstrates strong psychometric properties and offers a theoretically grounded tool for assessing STEM teaching in higher education.

5.1. Interpretation of the three-factor structure

The emergence of three distinct factors—Communicating Intended Learning Outcomes (ILO), Facilitating Teaching-Learning Activities (TLA), and Implementing Assessment Tasks (AT)—aligns with the constructive alignment principle central to OBTL (Biggs, 1996). This empirical finding validates the theoretical framework that effective teaching requires coherent integration of learning outcomes, teaching methods, and assessment strategies.

The separation of TLA from AT is particularly significant, indicating that facilitating learning activities and implementing assessments represent distinct competencies. This has important implications for faculty development: rather than treating teaching as a monolithic skill, professional development programs should address each dimension separately while acknowledging their interconnections.

The factor loadings provide insights into each dimension's structure. The ILO factor (loadings 0.691-0.737) comprises four items with strong internal consistency, suggesting that communicating outcomes can be assessed efficiently. The TLA factor (15 items, loadings 0.420-0.766) reflects the multifaceted nature of STEM pedagogy, encompassing diverse approaches including laboratory work, collaborative projects, differentiated instruction, and experiential learning. The AT factor (9 items, loadings 0.620-0.736) captures both traditional and contemporary assessment practices aligned with OBE principles.

5.2. Comparison with existing scales and contribution to literature

The three-factor model both converges with and diverges from existing instruments in meaningful ways. Unlike unidimensional scales, our model captures the complexity of STEM teaching within the OBE framework, aligning with calls for more nuanced evaluation systems.

Karpudewan et al.'s (2022) instrument for K-12 Malaysian teachers included teaching strategies and assessment factors but lacked explicit focus on communicating learning outcomes—a critical OBE component.

The STEM Faculty Instructional Barriers and Identity Survey (FIBIS) by Sturtevant and Wheeler (2019) focused on self-efficacy and barriers rather than providing a comprehensive evaluation framework. Yang et al.'s (2023) work on integrated STEM activities examined self-efficacy and commitment but did not capture the full range of teaching competencies. Our scale addresses these gaps by providing a performance-based assessment tool explicitly aligned with OBE/OBTL principles and incorporating STEM-specific practices.

5.3. Factor correlations and discriminant validity

The moderate to strong correlations between factors (r = 0.622 to 0.652) indicate that while the three dimensions are related aspects of teaching effectiveness, they remain distinct constructs. The highest correlation between ILO and AT (r = 0.652) suggests that faculty who clearly communicate learning outcomes tend to align their assessments accordingly, a fundamental principle of constructive alignment.

Discriminant validity was established despite these correlations, with √AVE for each construct exceeding its correlations with other constructs. This confirms that each dimension captures unique variance and represents a distinct aspect of teaching effectiveness, supporting the need for holistic faculty development that addresses all three dimensions while recognizing their interconnections.

5.4. Model fit and structural validity

The confirmatory factor analysis results provide strong evidence for structural validity. Model 2 demonstrated excellent fit: CFI = 0.926, TLI = 0.917, RMSEA = 0.061, SRMR = 0.047, and $\chi^2/df = 1.807$. All values met or exceeded recommended thresholds (Hair et al., 2010), confirming that the three-factor model adequately represented the data.

The improvement from Model 1 to Model 2 through theoretically justified error covariances likely reflects methodological factors (similar item wording) or substantive overlap between related practices. Factor loadings in Model 2 ranged from 0.61 to 0.89, with most exceeding 0.70, demonstrating strong relationships between observed variables and their respective latent constructs.

5.5. Reliability and internal consistency

The scale demonstrated excellent internal consistency, with Cronbach's alpha coefficients of 0.856 for Communicating Intended Learning Outcomes, 0.935 for Facilitating Teaching–Learning Activities, 0.906 for Implementing Assessment Tasks, and 0.953 for the overall instrument. These values substantially exceeded the recommended threshold of 0.70 and approached the 0.90 level, indicating excellent reliability. The high reliability of the Facilitating Teaching–Learning Activities dimension, despite comprising 15 items, suggests that diverse teaching practices cohere into a unified dimension of instructional facilitation.

Composite reliability further supported the internal consistency of the scale, with values of 0.845 for Communicating Intended Learning Outcomes, 0.937 for Facilitating Teaching–Learning Activities, and 0.908 for Implementing Assessment Tasks. In addition, Average Variance Extracted values were 0.578 for Communicating Intended Learning Outcomes, 0.499 for Facilitating Teaching–Learning Activities, and 0.524 for Implementing Assessment Tasks, providing additional evidence of reliability and convergent validity. Although the AVE value for Facilitating Teaching–Learning Activities was marginally below the conventional 0.50 threshold, it remains acceptable given the high composite reliability and the breadth of instructional practices encompassed by this construct (Fornell and Larcker, 1981). The final STEM faculty teaching effectiveness scale was designed for use by administrators and faculty members to evaluate teaching effectiveness within outcomes-based education and outcomes-based teaching and learning frameworks. The instrument operationalizes teaching effectiveness across three dimensions: communicating intended learning outcomes, facilitating teaching–learning activities, and implementing assessment tasks.

The Communicating Intended Learning Outcomes dimension consists of items that assess the extent to which faculty clearly articulate course learning outcomes at the beginning of instruction and align course objectives with national and international standards in STEM education, as well as with discipline-specific outcomes prescribed by outcomes-based education frameworks.

The Facilitating Teaching–Learning Activities dimension captures a wide range of instructional practices characteristic of effective STEM teaching. These include the use of scientific journals and scholarly resources, the design of learning activities that emphasize practical and real-world applications of STEM knowledge, collaborative and active learning strategies, differentiated instruction, integration of recent advances in STEM disciplines, use of appropriate laboratory equipment, and the implementation of experiential, hands-on, and self-directed learning activities that prepare students for professional practice.

The Implementing Assessment Tasks dimension evaluates assessment practices aligned with outcomes-based principles, including the use of formative and summative assessments, timely and constructive feedback, rubric-based evaluation, diverse STEM-specific assessment methods, alignment between assessments and course objectives, data-driven evaluation of teaching interventions, and the use of challenging yet attainable assessment tasks to support student learning.

Responses to the scale are recorded using a five-point Likert format ranging from 1 (Not demonstrated) to 5 (Very well demonstrated), allowing for systematic evaluation of observable teaching practices across the three dimensions of outcomes-based STEM instruction.

6. Conclusion and recommendations

This study successfully developed and validated a comprehensive, theoretically grounded scale for evaluating STEM faculty teaching effectiveness within OBE and OBTL frameworks. The 28-item instrument with three factors—Communicating Intended Learning Outcomes, Facilitating Teaching-Learning Activities, and Implementing Assessment Tasks—demonstrated excellent psychometric properties, including strong reliability (α = 0.953), convergent and discriminant validity, and good model fit. The scale addresses critical gaps in existing instruments by explicitly incorporating OBE/OBTL principles and capturing STEM-specific pedagogical practices.

The three-factor structure aligns with constructive alignment principles and provides a comprehensive framework for assessing, developing, and researching STEM teaching effectiveness. The scale offers practical utility for summative evaluation, formative assessment, faculty development, and institutional quality assurance. While limitations related to sample characteristics, cross-sectional design, and criterion validity warrant attention in future research, the current evidence strongly supports the scale's use in evaluating STEM faculty teaching effectiveness in higher education.

As higher education institutions worldwide continue to adopt OBE frameworks and seek to enhance STEM education quality, this validated instrument provides a valuable tool for systematic, fair, and comprehensive assessment of teaching effectiveness. By focusing evaluation on specific, observable practices aligned with educational outcomes, the scale supports evidence-based improvement in STEM teaching and learning. Future research validating the scale in diverse contexts, examining relationships to student outcomes, and tracking changes over time will further establish its value for promoting excellence in STEM education.

List of abbreviations

AT	Assessment tasks
AVE	Average variance extracted
BTS	Bartlett's test of sphericity
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CHED	Commission on Higher Education
CMO	Commission Memorandum Order
CR	Composite reliability
CVI	Content validity index
Df	Degrees of freedom
EFA	Exploratory factor analysis
ILO	Intended learning outcomes
KMO	Kaiser–Meyer–Olkin measure of sampling adequacy
OBE	Outcomes-based education
OBTL	Outcomes-based teaching and learning
PCA	Principal component analysis
QUAL	Qualitative
Quan	Quantitative
RMSEA	Root mean square error of approximation
SRMR	Standardized root mean square residual
STEM	Science, technology, engineering, and mathematics
TLA	Teaching and learning activities
TLA1–TLA15	Items measuring facilitating teaching–learning activities
TLI	Tucker–Lewis index
α	Cronbach's alpha
κ	Cohen's kappa
χ^2	Chi-square statistic

Acknowledgment

This research was funded by the Office of the Vice President for Research and Development of Cebu Technological University.

Compliance with ethical standards

Ethical considerations

This study was conducted in accordance with established ethical principles of research. Ethical approval was obtained from the Research Ethics Committee (REC; CNU-REC Code 1037/2024-05 Angco). Participation was voluntary, with informed consent obtained from all participants. Participants were informed of the study's purpose and their right to withdraw at any time. Confidentiality and anonymity were ensured through the removal of personal identifiers and secure data storage.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Allen M, Webb AW, and Matthews CE (2016). Adaptive teaching in STEM: Characteristics for effectiveness. Theory into Practice, 55(3): 217-224. https://doi.org/10.1080/00405841.2016.1173994

Asim HM, Vaz A, Ahmed A, and Sadiq S (2021). A review on outcome based education and factors that impact student learning outcomes in tertiary education system. International Education Studies, 14(2): 1-11. https://doi.org/10.5539/ies.v14n2p1

Barry AE, Chaney EH, Stellefson ML, and Chaney JD (2011). So you want to develop a survey: Practical recommendations for scale development. American Journal of Health Studies, 26: 97-105.

Biggs J (1996). Enhancing teaching through constructive alignment. Higher Education, 32: 347-364. https://doi.org/10.1007/BF00138871

Braun V and Clarke V (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2): 77-101. https://doi.org/10.1191/1478088706qp063oa

Fornell C and Larcker DF (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1): 39-50. https://doi.org/10.1177/002224378101800104

Gefen D, Straub D, and Boudreau MC (2000). Structural equation modeling and regression: Guidelines for research practice. Communications of the Association for Information Systems, 4: 7. https://doi.org/10.17705/1CAIS.00407

Hair JF Jr, Black WC, Babin BJ, and Anderson RE (2010). Multivariate data analysis. 7th Edition, Pearson Prentice Hall, Upper Saddle River, USA.

Karpudewan M, Krishnan P, Ali MN, and Yoon Fah L (2022). Designing instrument to measure STEM teaching practices of Malaysian teachers. PLOS ONE, 17(5): e0268509. https://doi.org/10.1371/journal.pone.0268509 **PMid:35594272 PMCid:PMC9122257**

Landrum RE, Viskupic K, Shadle SE, and Bullock D (2017). Assessing the STEM landscape: The current instructional climate survey and the evidence-based instructional practices adoption scale. International Journal of STEM Education, 4: 25. https://doi.org/10.1186/s40594-017-0092-1 **PMid:30631681 PMCid:PMC6310375**

Mufanti R, Carter D, and England N (2024). Outcomes-based education in Indonesian higher education: Reporting on the understanding, challenges, and support available to teachers. Social Sciences and Humanities Open, 9: 100873. https://doi.org/10.1016/j.ssaho.2024.100873

Srinivasan R, Lilien GL, and Rangaswamy A (2002). Technological opportunism and radical technology adoption: An application to e-business. Journal of Marketing, 66(3): 47-60. https://doi.org/10.1509/jmkg.66.3.47.18508

Sturtevant H and Wheeler L (2019). The STEM faculty instructional barriers and identity survey (FIBIS): Development and exploratory results. International Journal of STEM Education, 6: 35. https://doi.org/10.1186/s40594-019-0185-0

Yang KL, Wu HK, Yeh YF, Lin KY, Wu JY, and Hsu YS (2023). Implementers, designers, and disseminators of integrated STEM activities: Self-efficacy and commitment. Research in Science and Technological Education, 41(4): 1433-1451. https://doi.org/10.1080/02635143.2021.2008343