



A validated Arabic version of the clinical learning evaluation questionnaire for medical students and interns

Ahmed Amir Samir, Kerollos Abdelsayed, Hebatalla Abdelmaksoud Abdelmonsef Ahmed, Ahmed Almaahdy Mohamed, Ahmed Reda Bahr, Naji Al-bawah, Ali Malik Tiryag, Alla' Khirfan, Mohamed Yacoub & Ramy Mohamed Ghazy

To cite this article: Ahmed Amir Samir, Kerollos Abdelsayed, Hebatalla Abdelmaksoud Abdelmonsef Ahmed, Ahmed Almaahdy Mohamed, Ahmed Reda Bahr, Naji Al-bawah, Ali Malik Tiryag, Alla' Khirfan, Mohamed Yacoub & Ramy Mohamed Ghazy (2026) A validated Arabic version of the clinical learning evaluation questionnaire for medical students and interns, Medical Education Online, 31:1, 2632522, DOI: [10.1080/10872981.2026.2632522](https://doi.org/10.1080/10872981.2026.2632522)

To link to this article: <https://doi.org/10.1080/10872981.2026.2632522>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 19 Feb 2026.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)

A validated Arabic version of the clinical learning evaluation questionnaire for medical students and interns

Ahmed Amir Samir^{a,1} , Kerollos Abdelsayed^{b,1} , Hebatalla Abdelmaksoud Abdelmonsef Ahmed^c , Ahmed Almahdly Mohamed^d , Ahmed Reda Bahr^e , Naji Al-bawah^f , Ali Malik Tiryag^g , Alla' Khirfan^h, Mohamed Yacoubⁱ and Ramy Mohamed Ghazy^{j,k} 

^aFaculty of Medicine, Al-Azhar University, Cairo, Egypt; ^bHeart Rhythm Science Center, Minneapolis Heart Institute Foundation, Minneapolis, Minnesota, USA; ^cPublic Health and Community Medicine, Faculty of Medicine, Kafr Elsheikh University, Kafr Elsheikh, Egypt; ^dCardiovascular Department, University of Minnesota Twin Cities, Minneapolis, USA; ^eFaculty of Medicine, Alexandria University, Alexandria, Egypt; ^fFaculty of Medicine, Sana'a University, Sana'a, Yemen; ^gFundamentals of Nursing, College of Nursing, University of Basrah, Basrah, Iraq; ^hFaculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan; ⁱDepartment of English, Florida International University, Miami, FL, USA; ^jFamily and Community Medicine Department, College of Medicine, King Khalid University, Abha, Saudi Arabia; ^kTropical Health Department, High Institute of Public Health, Alexandria University, Alexandria, Egypt

ABSTRACT

Background: The Clinical Learning Evaluation Questionnaire (CLEQ) is a comprehensive and reliable tool designed to systematically assess the quality and effectiveness of clinical learning environments, which are essential for undergraduate medical education. This study aims to translate, culturally adapt, and validate the CLEQ into Arabic (A-CLEQ) to enhance undergraduate clinical education in Arabic-speaking settings.

Methods: A cross-sectional study using an anonymous, online, self-administered questionnaire was conducted between July and September 2025 in four randomly selected Arab countries (Egypt, Jordan, Yemen, and Iraq). The CLEQ was translated into Arabic through a combination of forward and backward translation. The internal consistency was calculated using Cronbach's alpha. The content validity was determined through a review by experts. An exploratory factor analysis (EFA) was performed to discover the factor structure, followed by confirmatory factor analysis (CFA) to validate the model. Convergent and discriminant validity were checked using the average variance extracted and factor correlations.

Results: A total of 500 participants were included. Cronbach's alpha of the A-CLEQ was 0.92, indicating excellent reliability. The content validity index (CVI) for the items ranged from 0.83 to 1.0. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy for factor analysis was 0.90, and Bartlett's test of sphericity was significant ($P < .0001$). EFA supported a three-factor structure of the A-CLEQ, accounting for 60.6% of the variance (Cases, Motivation to learn, and Supervision & Organization factors). The correlations between A-CLEQ domains were < 0.70 , ascertaining discriminant validity. CFA model fit indices indicated good fit, with a Comparative Fit Index (CFI) of 0.94 and a Goodness-of-Fit Index (GFI) exceeding 0.90.

Conclusion: The 18-item A-CLEQ demonstrates strong psychometric properties, including good reliability and validity, making it a suitable tool for assessing the clinical learning environment among clinical-year medical students and interns in Arabic-speaking populations.

ARTICLE HISTORY

Received 24 December 2025
Revised 10 February 2026
Accepted 11 February 2026

KEYWORDS

Medical education; clinical learning; medical students; medical interns; cultural adaptation; questionnaire validity

Introduction

The early exposure to the clinical environment is increasingly considered a crucial determinant of a successful medical education curriculum [1]. Successful early clinical engagement provides medical students with essential clinical, professional, and ethical skills, preparing them to become competent healthcare providers. It also facilitates a smooth transition from undergraduate education to postgraduate

CONTACT Naji Al-bawah  najialbawah@gmail.com  Faculty of Medicine, Sana'a University, Sana'a, Yemen

¹These authors contributed equally to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/10872981.2026.2632522>.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

training and independent practice. Therefore, the quality and effectiveness of this early clinical exposure are crucial in shaping the overall competence of the medical education process [2].

The integration of clinical training during the medical education curriculum is a multifaceted process influenced by various factors. These include the depth of clinical exposure, duration of patient contact, availability of proper supervision and feedback, as well as the implementation of standardised assessments [3]. Additionally, there is growing interest in how medical students perceive the adequacy of their clinical training and exposure throughout their education [2].

In the Middle East and North Africa (MENA) region, there exists significant diversity in the methods of delivering medical education. Some medical schools adhere to the traditional British model, which consists of three years focused on basic sciences followed by three years of clinical training, along with an optional internship year. On the other hand, certain medical schools have adopted a more modern approach, where the curriculum spans five years and incorporates clinical experience starting from the first year. A few medical schools emulate the United States system, which involves four undergraduate years to obtain a bachelor's degree, followed by four years of medical school, with clinical exposure beginning in the first year [4]. Moreover, recent curricular reforms in some MENA countries, such as Egypt, have introduced new complexities. These include challenges in transitioning from the old system, concerns regarding student mental health under the new and still experimental model, and ongoing debates about its overall competence compared to the established curriculum [5,6].

Therefore, it is of increasing importance to have validated and reliable tools where educators and professors can evaluate the quality, quantity, and overall effectiveness of the clinical learning process, both based on objective and subjective parameters [2]. Fortunately, multiple assessment tools have been introduced, refined, and validated for the assessment of the general clinical training environment during medical education, including the Dundee Ready Educational Environment Measure (DREEM) [7], the Postgraduate Hospital Educational Environment Measure (PHEEM) [8], and the Clinical Learning Environment Inventory (CLEI) [9]. However, these tools had their shortcomings, especially in assessing the quality of the medical education process from the perspective of the students [2]. Moreover, two other tools were also recently provided: the Cleveland Clinical Teaching Effectiveness Inventory (CCTEI) [10], and the study by Pololi and Price, which introduced a tool to measure the perception of medical students of the learning environment [11]. Although those instruments addressed crucial determinants of the learning environment, such as the teacher-learner relationship, self-efficacy, and physician-patient relationship, they did not sufficiently address the perception of the clinical experiences and the potential organisational challenges throughout the process [2]. Therefore, the Clinical Learning Evaluation Questionnaire (CLEQ) was recently developed to assess the subjective perception of both undergraduate medical students and teachers of the clinical learning environment [2,12].

The original CLEQ was introduced in 2014 based on 37 items to assess six factors: cases, authenticity of clinical experience, supervision, organisation of the doctor-patient encounter, motivation to learn, and self-awareness. The questionnaire was developed in English and validated on a sample of 182 students from three medical colleges in Saudi Arabia [2]. However, a confirmatory factor analysis (CFA) using structural equation modelling in 2020 could not achieve model fit with the six-factor questionnaire when used on 185 other students from the same country. Instead, this study showed a good fit for another proposed four-factor reduced version of the CLEQ questionnaire that was based on 18 items [12]. The latter form of the questionnaire was then cross-culturally adapted and validated in China (Chinese version) and Iran (Persian version) and showed favourable validity and reliability [13,14]. Although the questionnaire was originally developed in English and used among Arabic-speaking students, it had not been previously translated or validated in Arabic. This may have introduced inherent bias due to possible misinterpretation or cultural differences in language use [15]. Moreover, it could magnify the non-response selection bias in the study by only involving individuals who are confident in their English language [16]. Therefore, this study aimed to cross-culturally adapt and validate an Arabic version of the CLEQ questionnaire, where the questionnaire can be later used for other Arabic-speaking countries in the MENA region.

Methodology

Study design, setting, and duration

This observational cross-sectional study was conducted between July and September 2025. The research was performed in four randomly selected Arabic-speaking countries: Egypt, Iraq, Yemen, and Jordan.

Sample size calculation

The CLEQ consists of 18 items and adheres to standard recommendations for exploratory factor analysis (EFA), which suggest recruiting approximately 10–20 respondents per item [17]. Accordingly, we targeted 200 participants for the EFA dataset. For CFA, best-practice guidelines recommend a minimum of 200–300 participants to ensure model stability and accurate estimation of fit indices [18]; therefore, we targeted 300 participants. In total, the planned analysable sample size was 500 participants, with 125 from each of the four participating countries. Within each country, participants were allocated in a 40:60 ratio for EFA and CFA, respectively, resulting in EFA datasets ($n = 200$; 50 per country) and CFA datasets ($n = 300$; 75 per country).

Sampling method

Convenience and snowball sampling methods were used to obtain the required sample size from selected Arab countries. Each country had a designated group of data collectors, referred to as the CLEQ group. Before collecting data, all data collectors received standardised online training on survey distribution strategies and data collection protocols.

Eligibility criteria

Undergraduate medical students currently enrolled in clinical years (typically years 3–6, depending on curriculum and our included countries system) and medical interns (post-final year/s students undergoing clinical internship training), who have at least six months' clinical rotation experience, were invited to voluntarily participate in the study after providing informed consent. Medical interns were part of this study because they embody a transitional stage in early medical training, where their learning is significantly shaped by the clinical environment and oversight. Including them facilitated a more thorough assessment of the clinical learning environment across various levels of clinical training. All the countries involved in the study adhere to undergraduate medical education systems similar to the Bachelor of Medicine, Bachelor of Surgery (MBBS) model, with differences primarily stemming from the structure of the curricula rather than the type of degree awarded. The equivalence among different years or phases of training was evaluated based on practical clinical exposure, which includes supervised rotations and direct patient interaction, rather than the titles of the curricula or formal year classifications.

Data collection and handling

Data was collected through an anonymous self-administered online link to a web-based survey, created with the Google Forms platform. The survey was disseminated by the study's collaborators via various social media platforms and online platforms accessible to medical students/interns, including university social media channels, student portals, online student forums, and discussion groups. Participation was voluntary, anonymous, and completed independently by students across the selected Arab countries. In the first page of the survey, participants received comprehensive information on the study's objectives, estimated time required for completion, informed consent, voluntary nature, anonymity, and confidentiality safeguards. To avoid duplicate responses by IP address, we enabled a restriction option for a single response in the Google Form settings.

Data collection tools and procedures

The survey consists of two sections. The first section encompasses sociodemographic data, including age, sex, country, residence, type of medical school, clinical year, and current medical education system (traditional; features 2–3 years pre-clinical basic sciences followed by 3 years clinical rotations and 1-year internship programme vs integrated; modular and problem-based learning approaches blending sciences and clinical exposure from first year). In addition, the survey includes questions regarding having a clear idea about clinical training, attending orientation/workshop on clinical training, hours per week in hospitals/clinics, number of patients directly interacted with per week, explaining objectives/expectations of clinical training, and use of simulation-based training and digital tools for clinical learning. The second section presents the Arabic translation of the 18-item English version of the CLEQ, which was refined by Nuha Alnaami et al. [12], to assess four factors: (1) clinical cases (4 items), (2) organisation of patient encounters (5 items), (3) supervision (4 items), and (4) motivation to learn (5 items) (**Data S1**). Each item was rated on a 5-point Likert scale (1 = Strongly disagree, 2 = Disagree, 3 = Undecided, 4 = Agree, and 5 = Strongly agree).

Questionnaire cultural adaptation and validation

Translation and adaptation

Cross-cultural adaptation and translation requirements were followed [19]. The cross-cultural adaptation of the CLEQ followed a multi-step process. Initially, two bilingual translators, who are native speakers of Arabic, independently translated the text from English to Arabic. These translations were then combined into one agreed-upon version. Following this, a single native English speaker, unaware of the original instrument, translated this consensus version back into English. This step helped pinpoint and rectify any conceptual inconsistencies or ambiguities that might have arisen during the initial translation (**Data S2**) [20]. Standard Arabic was used in this study because it is the official language of 22 Arab countries, and standard Arabic is extensively taught, understood, and spoken by local Arabs.

Content validity and expert evaluation

Content validity was assessed quantitatively. The process began with the development of a content validation form to guide the expert panel. A panel of six experts in medical education was assembled to evaluate the questionnaire. The experts assessed the items for their relevance to the specified constructs, clarity of wording, and suggested any necessary modifications. For each item, experts provided two separate ratings on a 4-point Likert scale: one for relevance (1 = not relevant, 4 = highly relevant) and one for clarity (1 = not clear, 4 = very clear).

Content validity indices were calculated based on the relevance ratings. The Item-Content Validity Index (I-CVI) was calculated for each item as the proportion of experts rating its relevance as 3 or 4. The Scale-Content Validity Index/Average (S-CVI/Ave), computed as the mean of all I-CVI scores, was used to evaluate the entire scale; a benchmark of ≥ 0.90 indicates excellent overall content validity. Finally, the Scale-Content Validity Index/Universal Agreement (S-CVI/UA), representing the proportion of items achieving a relevance rating of 3 or 4 from all experts, was reported as a more conservative indicator of consensus.

Pilot testing

A pilot study was conducted with 50 participants following translation and adaptation. We evaluated understanding, readability, language, cultural appropriateness, and the time required for participants to complete the questionnaire. Additionally, participants were invited to provide their comments on the questions while filling out the survey.

Psychometric analysis, data management, and statistical programmes used

Reliability and item analysis

Cronbach's alpha was used to assess the internal consistency of the scale. A Cronbach's alpha between 0.70 and 0.80 is considered satisfactory, while a value above 0.80 indicates excellent reliability [21].

Construct validity

To analyse the construct validity of the A-CLEQ, we conducted EFA and CFA in two phases. Before performing the EFA, we assessed the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and conducted Bartlett's test of sphericity. The KMO statistic ranges from 0 to 1, with values closer to 1 indicating greater suitability for factor analysis. A KMO value below 0.60 suggests poor adequacy [22]. Additionally, if the p -value from Bartlett's test is less than 0.05, it indicates that factorial analysis is appropriate [23]. Factors were extracted according to Kaiser's criterion (eigenvalues > 1) and confirmed through visual inspection of the scree plot. EFA was conducted using principal component analysis (PCA) with oblimin rotation, acknowledging potential correlations among factors. For interpreting item loading, 0.32 is considered poor, 0.45 fair, 0.55 good, 0.63 very good, and 0.71 excellent [24]. Regarding the uniqueness value, representing the proportion of a variable's variance not explained by the common factors in the model, a cutoff of >0.60 indicates a poor fit model [25]. Moreover, the average variance extracted (AVE) of 50.0% or greater means that, on average, more variance is explained by the construction than by measurement error [26]. Subsequently, CFA was conducted to evaluate the factor structure identified in the EFA. Standardised factor loadings of 0.50 or higher were considered acceptable indicators of practical significance [27]. Model fit for the CFA was evaluated using a standard set of fit indices and the following established benchmarks for good fit: Comparative Fit Index (CFI) > 0.90 , Tucker-Lewis Index (TLI) > 0.90 , Goodness of Fit Index (GFI) > 0.90 , Root Mean Square Error of Approximation (RMSEA) < 0.06 , and Standardised Root Mean Square Residual (SRMR) < 0.08 . The chi-square statistic was reported but was not used as a primary fit criterion due to its known sensitivity to sample size [28]. Discriminant validity was assessed using the Heterotrait-Monotrait (HTMT) ratio of correlations. Following the recommended threshold, an HTMT value below 0.85 was used to indicate sufficient discriminant validity between constructs [29]. Moreover, we utilised the factor correlation matrix of the domains. Discriminant validity was evaluated when the inter-factor correlation was below 0.70 [30]. Following the CFA, measurement invariance across countries was evaluated using multigroup CFA. Three levels of invariance were tested: configural (same factor structure), metric (equal factor loadings), and scalar (equal item intercepts). Chen [31] proposed cutoff values for changes in model fit indices (Δ) to assess measurement invariance. Invariance is supported when differences between nested models are sufficiently small, specifically: for metric invariance, $\Delta\text{CFI} \geq -0.01$, $\Delta\text{RMSEA} < 0.015$, and $\Delta\text{SRMR} < 0.030$; and for scalar and residual invariance, $\Delta\text{CFI} \geq -0.01$, $\Delta\text{RMSEA} < 0.015$, and $\Delta\text{SRMR} < 0.010$.

Statistical analysis

The responses were combined into an online spreadsheet where the data were coded. The mean \pm standard deviation (SD) was used to represent quantitative variables, and one-way analysis of variance (ANOVA) was used to compare between groups, followed by Bonferroni post hoc tests to identify pairwise differences. Numbers and percentages were used to describe qualitative variables. Pearson's correlation analysis was utilised to calculate the item-to-total correlation and the inter-domain correlation of the questionnaire. The p -value was established at less than 0.05. We used the Statistical Package for Social Science (SPSS) (version 25, Chicago, USA) to execute most of the analyses. Meanwhile, JASP (Version 0.17.3) was used to conduct the EFA and CFA.

Results

Cultural adaptation

During the translation process and cultural adaptation, some sentences and expressions were found to be unclear and were carefully clarified to improve comprehension and fluency. Additionally, in the original version of the questionnaire, the third option on the Likert scale was labelled 'Undecided.' However, this

option was changed to 'Neutral' during translation to better reflect the intended meaning and to improve clarity and relevance for the target population. In terms of cultural relevance and the linguistic use of the terms to be investigated, the experts also provided valuable comments and suggestions, such as replacing certain words with more suitable ones, adding clarification to specific words or concepts, and making certain expressions more precise. This modification was made to increase clarity and cultural relevance; however, each item's original purpose and structure were maintained throughout the process. Furthermore, a few changes were made during the cognitive interview to improve the language clarity. These adjustments preserved the Arabic version's validity and accuracy for respondents who spoke Arabic while preserving its essential content. The modification based on expert evaluation is depicted in **Data S3**, and the final version of the questionnaire is attached to **Data S4**.

Sociodemographic characteristics

A total of 500 medical students/interns were included, with a mean age of 23.0 ± 3.0 years. More than half of the participants (54.60) were females, with most students (88.60%) enrolled in public medical schools. Of the participants, 61.60% were enrolled in traditional medical education programmes, and 39.40% participated in any orientation sessions or workshops related to clinical training. The full socio-demographic and clinical training characteristics of the sample are depicted in **Table 1**.

Descriptive statistics of the A-CLEQ items

Table 2 shows the Item-level descriptive statistics and reliability analysis of the A-CLEQ. The mean scores of the individual items ranged from 2.70 ± 1.24 (Item 14) to 3.88 ± 0.92 (Item 7). The total scale score had a mean of 60.17 ± 11.87 (**Figure 1**). The overall Cronbach's α coefficient for the A-CLEQ was 0.92, with

Table 1. Socio-demographic and clinical training characteristics of the study participants.

Studied variables		N (500)	%
Age (years) mean \pm SD			23.00 \pm 3.00
Gender	Female	273	54.60
	Male	227	45.40
Type of medical school	Public	443	88.60
	Private	57	11.40
Residence	Rural	95	19.00
	Urban	405	81.00
Academic year	3rd year	58	11.60
	4th year	116	23.20
	5th year	128	25.60
	6th year	96	19.20
	Internship year	102	20.40
Current medical education system	Traditional system	308	61.60
	Integrated system	192	38.40
Having a clear idea about clinical training	Yes	262	52.40
	No	238	47.60
Attending orientation/workshop on clinical training	Yes	197	39.40
	No	303	60.60
Hours per week in hospitals/clinics	<20 hours	356	71.20
	20–30 hours	111	22.20
	>30 hours	33	6.60
Patients' direct interaction per week	<10	343	68.60
	10–20	120	24.00
	>20	37	7.40
Objectives/expectations of clinical training explained	Always	66	13.20
	Often	177	35.40
	Sometimes	173	34.60
	Rarely	62	12.40
	Never	22	4.40
Use of simulation-based training	Always	44	8.80
	Often	115	23.00
	Sometimes	151	30.20
	Rarely	108	21.60
	Never	82	16.40
Use of digital tools for clinical learning	Yes	305	61.00
	No	195	39.00

Table 2. Item-level descriptive statistics and reliability analysis of the A-CLEQ.

A-CLEQ	Mean	SD	Cronbach's α if item deleted
Item 1	3.00	1.03	0.91
Item 2	3.39	1.01	0.91
Item 3	3.02	1.06	0.91
Item 4	2.98	1.14	0.92
Item 5	3.86	0.98	0.91
Item 6	3.42	1.02	0.91
Item 7	3.88	0.92	0.91
Item 8	3.57	0.97	0.91
Item 9	3.86	0.84	0.91
Item 10	3.45	0.99	0.91
Item 11	3.37	0.99	0.91
Item 12	3.51	0.92	0.91
Item 13	3.62	0.94	0.91
Item 14	2.70	1.24	0.92
Item 15	2.94	1.09	0.91
Item 16	3.20	1.02	0.91
Item 17	2.90	1.15	0.91
Item 18	3.48	1.00	0.91
Total A-CLEQ Scale	60.17	11.87	Cronbach's $\alpha = 0.92$

A-CLEQ: Arabic Clinical Learning Evaluation Questionnaire.

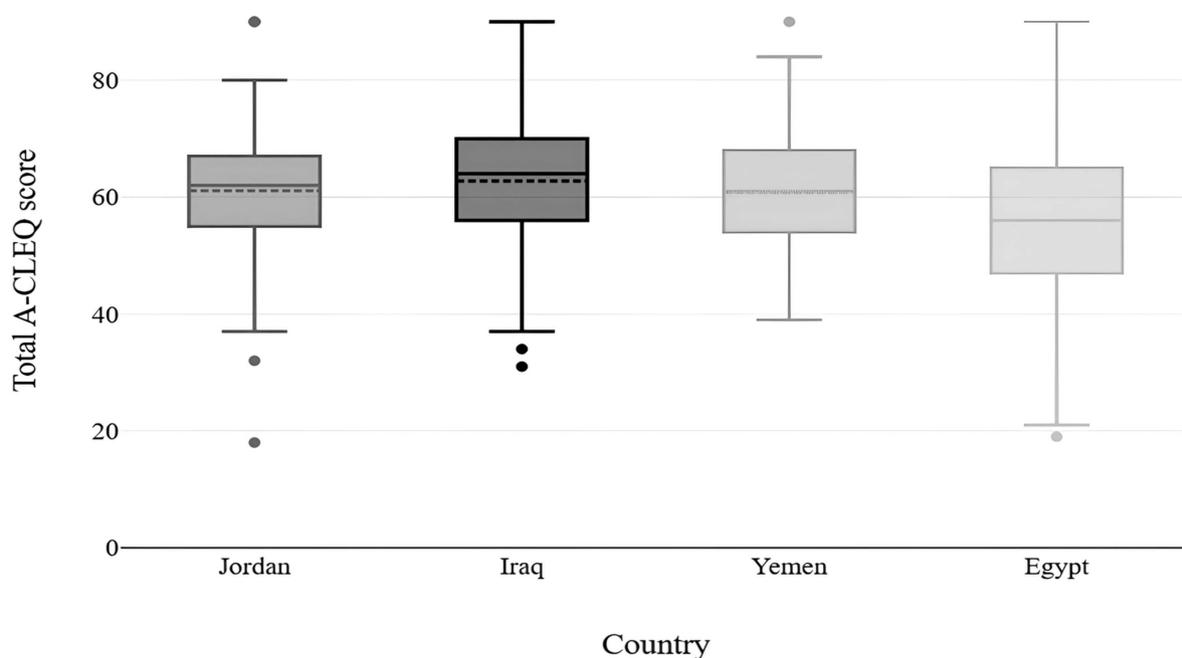


Figure 1. Boxplot of distribution of A-CLEQ scores among medical students across the studied countries.

Cronbach's α remaining stable across the scale if an item was deleted, ranging between 0.91 and 0.92, suggesting that no single item substantially reduced the internal consistency of the questionnaire. The percentages of the responses to each item of the A-CLEQ were illustrated in Figure S1.

Content validity

The I-CVI values for the items were between 0.83 and 1.0, where items 1, 2, and 4 received a score of 0.83, and the rest of them were rated as 1.0. Additionally, the S-CVI/UA and S-CVI/Ave were calculated and were 0.83 and 0.97, respectively.

EFA and reliability analysis of the A-CLEQ

The initial EFA of the A-CLEQ indicated a three-factor structure, based on eigenvalues greater than 1, which accounted for 60.60% of the total variance. However, given that the original questionnaire was developed with a four-factor model, we further examined this model in our data (Table S1). In the four-factor model, two items demonstrated cross-loadings that diverged from the original structure: Item 14 ('The number of students in the clinical sessions is appropriate') loaded more strongly on the Supervision factor rather than Organisation, and Item 18 ('I have the opportunity to prepare before the clinical encounter') loaded on Motivation to Learn instead of Organisation. Furthermore, several items displayed factor loadings below the conventional threshold of 0.32, indicating limited contribution to the latent construction. Consequently, we considered the three-factor solution (Cases, motivation to learn, and Supervision & Organisation factors) to be more parsimonious and theoretically coherent, with Supervision and Organisation integrated into a single domain. The three-factor EFA of the A-CLEQ had a KMO value of 0.90, and the significance level of Bartlett's sphericity test was <0.001. The three domains represented excellent reliability, with all above 0.70. The eigenvalue for factor 1 was 8.56, which consisted of items reflecting the 'Supervision & Organisation' domain; factor 2 was 1.94, composed of items corresponding to 'Cases'; factor 3 was 1.54, representing 'Motivation to learn'. Uniqueness for all items was below the recommended threshold of 0.60. Most items loaded strongly on their respective domains, supporting construct validity; moreover, the item-domain correlations ranged from 0.53 to 0.79, suggesting adequate convergent validity (Table 3).

CFA of the A-CLEQ items

Regarding the model fit indices for CFA of the A-CLEQ, they demonstrated a good model fit with CFI = 0.94, TLI = 0.93, and GFI = 0.98 above the threshold of 0.90. Moreover, the RMSEA of 0.06 and the SRMR of 0.04 were well below the threshold of 0.08. The HTMT values for all factor pairs were below the threshold of 0.85, ranging from 0.59 to 0.79, suggesting adequate discriminant validity. The factor loading ranged from 0.54 to 0.89. Each loading was statistically significant ($P < 0.001$) (Table S2). Moreover, the standardised factor loadings ranged from 0.54 to 0.83 onto three-dimensional factors (Figure 2).

Measurement invariance of the A-CLEQ Items by country

Table S3 presents the results of the multigroup CFA to assess the measurement invariance of the A-CLEQ across countries. The configural model (M1) demonstrated an acceptable fit ($\chi^2/df = 1.65$, CFI = 0.88, RMSEA = 0.09, SRMR = 0.07), supporting the baseline model structure. In metric invariance (M2), the fit

Table 3. Results of the exploratory factor analysis of the A-CLEQ ($N = 200$).

A-CLEQ	Cases	Motivation to learn	Supervision & organisation	Uniqueness	Item-domain correlation
Item 1	0.70			0.35	0.66
Item 2	0.77			0.29	0.71
Item 3	0.90			0.16	0.79
Item 4	0.77			0.43	0.63
Item 5		0.64		0.55	0.60
Item 6		0.53		0.55	0.61
Item 7		0.66		0.38	0.53
Item 8		0.64		0.41	0.63
Item 9		0.71		0.48	0.55
Item 10			0.51	0.28	0.70
Item 11			0.82	0.26	0.68
Item 12			0.80	0.29	0.71
Item 13			0.63	0.40	0.63
Item 14			0.71	0.53	0.55
Item 15			0.73	0.28	0.60
Item 16			0.48	0.39	0.68
Item 17			0.61	0.49	0.65
Item 18			0.35	0.55	0.54
Cronbach's α	0.85	0.79	0.89		

A-CLEQ: Arabic Clinical Learning Evaluation Questionnaire.

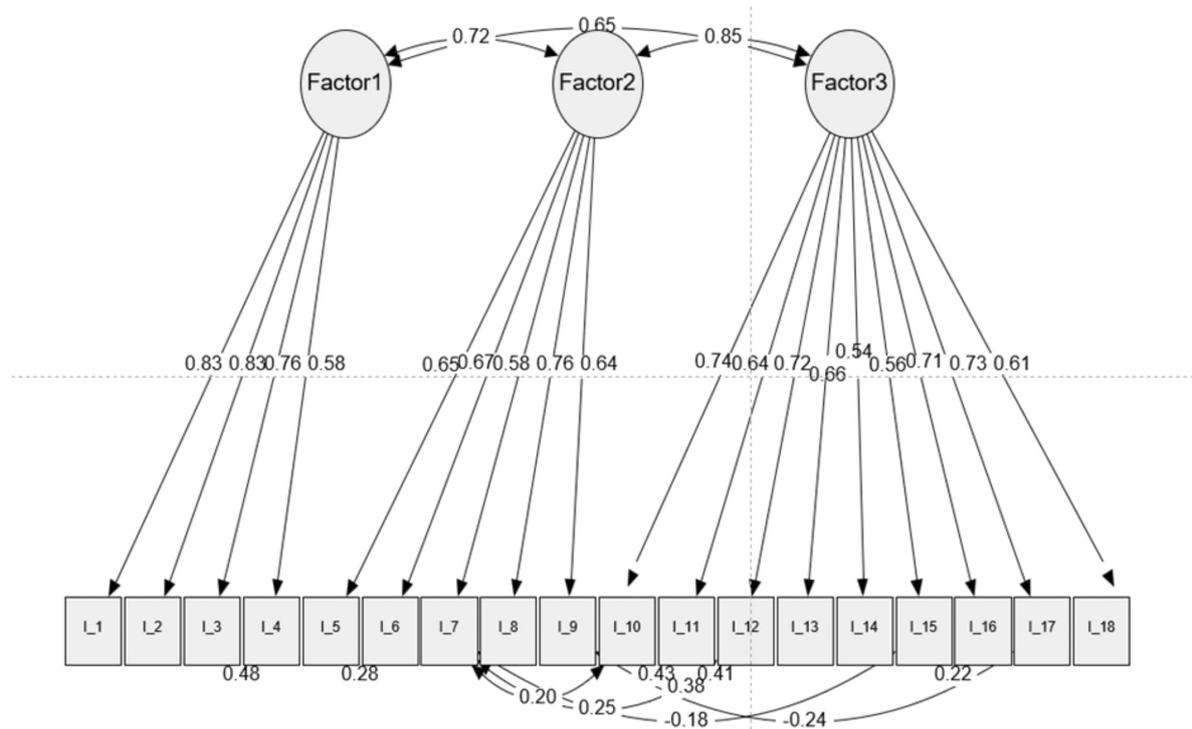


Figure 2. Confirmatory Factor Analysis for the Arabic Clinical Learning Evaluation Questionnaire (A-CLEQ) (Factor 1: Cases; Factor 2: Motivation to learn; Factor 3: Supervision & Organization).

remained comparable within commonly accepted thresholds ($\Delta\text{CFI} = -0.01$; $\Delta\text{RMSEA} = -0.00$; $\Delta\text{SRMR} = 0.02$). Moreover, in scalar invariance (M3), the model showed an acceptable comparable level.

Inter-correlations between A-CLEQ domains

The correlation analysis demonstrated strong and statistically significant associations among the A-CLEQ subscales and the total scale (all $P < 0.001$). The strongest correlation was observed between *Supervision & Organisation* and the *Total Scale* ($r = 0.93$). The correlations between A-CLEQ domains were positive and significant, with a range of ($r = 0.51:0.65$) below the cutoff value of 0.70, ascertaining discriminant validity (Table S4).

Cross-country comparisons of the A-CLEQ items

Table S5 presents a comparison of the A-CLEQ domains and total scores across the countries studied. Significant differences were observed in all domains and the total scale score, with the lowest observed scores in Egypt compared to the other three countries.

Discussion

This study examined the applicability of the four-factor model of the CLEQ. The original 18-item CLEQ was translated into Arabic and administered to medical students and interns across four Arab countries in the MENA region to evaluate its cross-cultural adaptability, construct validity, and reliability. The findings revealed that a revised three-factor model, termed the A-CLEQ, demonstrated good reliability, internal consistency, and strong psychometric properties.

Most items ($n = 15$) had a perfect I-CVI of 1.0 for relevance, meaning that every expert assessed it as either quite or extremely relevant. These values indicate that the questionnaire had good overall content validity [32]. Similar findings were reported in previous Persian questionnaires, which ranged from 0.80 to 0.90, demonstrating that the items were representative of the construct being tested and measured [14].

Additionally, the relevance revealed a perfect score among the panel for both S-CVI/Ave and S-CVI/UA, indicating that the questionnaire had good overall content validity.

Reliability testing confirmed the robustness of the A-CLEQ. Compared with the original six-factor 37-item CLEQ ($\alpha = 0.60\text{--}0.86$) and the four-factor 18-item version ($\alpha = 0.72\text{--}0.87$), the A-CLEQ demonstrated higher indices [2,12]. The overall Cronbach's α coefficient for the A-CLEQ was 0.92, which aligned with the findings of the Chinese validation and was higher than the Persian validation, 0.87, indicating that the CLEQ's core items maintain strong internal consistency across diverse cultural and educational contexts [2,12].

The outcomes of the EFA of the A-CLEQ show a three-factor structure. These findings suggest that the distinction between supervision and organisation may not be empirically robust in our sample, as both domains reflect related and integrated aspects of the clinical learning environment. Consequently, we considered the three-factor solution to be more parsimonious and theoretically coherent, with supervision and organisation integrated into a single domain. The merger of the 'Supervision' and 'Organisation' domains is not a methodological weakness but a reflection in line with the cultural and educational context in which clinical training is delivered. From a theoretical perspective, sociocultural learning theory views learning as a socially mediated process that occurs through participation in shared practices rather than isolated instruction. Situated learning further emphasises that clinical learning develops through engagement in authentic workplace activities shaped by social, organisational, and hierarchical contexts [33–35]. Accordingly, supervision in clinical settings is inherently embedded within organisational arrangements, such as team structure, workload, scheduling, and access to supervisors, that jointly shape learners' opportunities for participation and feedback.

In the current study, the merging of supervision and organisation reflects how these elements are experienced as closely interconnected during clinical training, where scheduling, workload, and supervisor availability shape daily learning. Interpretation of these findings should consider the educational context of the included cohorts, which differ in clinical year, educational system, clinical exposure, and participation in orientation activities. The inclusion of both clinical-phase students and medical interns represents a continuum of clinical training within similar organisational settings. In many Arabic-speaking medical schools, these features are closely intertwined, leading students to experience supervision and organisation as inseparable dimensions of their clinical training. The emergence of a unified Supervision & Organisation factor in the A-CLEQ, therefore, represents a theoretically coherent and contextually meaningful depiction of clinical learning within hierarchical health care systems. Notably, a similar merge was also recommended in the validation of the Chinese version of the CLEQ to mitigate method bias [13].

As psychometric structures vary across languages and contexts, this revised three-domain model better aligns with learners' perceptions while preserving all 18 items. However, it may not fully capture variation arising from regional dialects and localised language use within the same language. Despite this structural modification, the A-CLEQ demonstrated excellent reliability, strong construct validity, and robust measurement invariance, enhancing its cultural relevance and utility as a validated tool across diverse medical education settings in the MENA Region. The findings highlight that validating culturally adapted assessment tools is essential, and future longitudinal research is needed to explore CLEQ's ability to predict clinical competence.

Regarding CFA of the A-CLEQ, the model demonstrated a good model fit, all of which meet commonly accepted thresholds for model fit [36]. Similar findings were reported in the four-factor CLEQ by Alnaami et al. [12], with CFI = 0.951, GFI = 0.903, and RMSEA = 0.052. While the RMSEA in Alnaami et al.'s study [37] was slightly better (0.052 vs. 0.06 in our study), the overall fit of our three-factor model is acceptable, particularly as the GFI exceeds 0.90, and further improvements might be achieved with larger sample sizes. Also, the Persian validation showed acceptable values (CFI = 0.95, GFI = 0.91); however, the RMSEA was relatively high at 0.092, indicating a slightly weaker fit [14]. The Chinese version also demonstrated good indices, with a CFI of 0.956 and an RMSEA of 0.057 [13]. Overall, the findings indicate that the A-CLEQ demonstrates strong construct validity and robust psychometric properties, supporting its use among medical students and interns in Arab countries.

A significant difference in CLEQ score was observed in all domains and the total scale score, with the lowest observed scores in Egypt compared to the other three countries, suggesting potential challenges in

clinical supervision, organisational structure, or student engagement that may warrant targeted intervention. This variability in scores across participants' countries likely reflects national contextual differences in clinical training environments, institutional structures, and educational practices, proving that the quality of the clinical learning environment is influenced by broader socioeconomic, institutional, and curricular factors. These observations are consistent with prior research using analogous tools like the DREEM, which evaluates comparable domains in medical education. For instance, a multinational study across Arab countries reported significantly lower DREEM scores in Egypt compared to Saudi Arabia, Iraq, Yemen, and Syria, attributing Egypt's challenges to socioeconomic factors like budget constraints, political instability, teacher shortages, and limited resources [38]. To facilitate evidence-based reforms, it is essential to validate tools such as the A-CLEQ within specific national contexts and among comparable cultural groups. Future multicenter studies should employ mixed methods to assess clinical learning environments throughout the Arab world, thereby informing targeted interventions. Notably, while disparities in scores between countries may indicate genuine differences, the CLEQ alone cannot reveal the intricate socioeconomic factors contributing to these variations. These results underscore the need for next-generation surveys that combine traditional educational metrics with explicit evaluations of how socioeconomic factors impact clinical learning.

Strengths and limitations

To our knowledge, this is the first to validate and translate a tool assessing the clinical learning environment for clinical years medical students and interns in Arabic. Conducted with 500 participants and across four Arab countries, the study provides strong generalisability. Robust psychometric analyses and expert translation confirmed the tool's validity, reliability, and cross-cultural applicability. However, this study has some limitations. First, utilising non-probability sampling may limit wider representativeness within each country and increase the risk of selection bias. The cross-sectional design does not assess temporal stability or predictive validity. Self-reported data introduced response bias, and the test-retest reliability was not assessed. Additionally, curricula and clinical training structures differ from country to country. This variation may affect students' perceptions. Also, differences in demographics among participants might have influenced their responses. Moreover, the use of an odd-numbered Likert scale may have introduced response bias via a neutral middle option. Furthermore, the non-probability sampling strategy, which relied on online social media dissemination, likely created a selection bias toward students who are digitally connected and academically engaged—often from urban or resource-rich institutions. Consequently, our findings may modestly overrepresent the experiences of students with greater access to academic and digital resources.

Conclusion

The A-CLEQ demonstrates strong psychometric properties, including good reliability and construct validity, establishing it as a culturally appropriate and effective tool for evaluating clinical learning environments in Arab medical schools. Its application can help institutions identify specific areas for improvement, such as clinical supervision and resource allocation, and provide actionable data to guide curricular development and enhance the overall quality of clinical education.

Acknowledgements

We appreciate the study participants for their commitment and patience. We sincerely thank the medical education experts who assisted in validating the questionnaire, especially Dr. Hany Salem Atwa, Dr. Shima Elarbaby, Dr. Yasser El-Wazir, [Professors of Medical Education, Faculty of Medicine, Suez Canal University, Ismailia, Egypt], Dr. Ebtehal Dongol [Faculty of Medicine, South Valley University, Qena, Egypt], and Dr. Mahmoud Basiouny [Faculty of Medicine, Alexandria University, Alexandria, Egypt].

We would like to thank ALHaqwi and Kuntze, who designed a total of 37 CLEQ items, and thank Nuha Alnaami et al., for the latest version of the CLEQ.

We are thankful to the members of the CLEQ- Collaborative group for their outstanding performance during data collection.

Author contributions

AAS and **KA**: Conceptualisation of the research, methodology, writing the original draft, reviewing, and editing the manuscript. **HAAA**: Formal analysis and writing original draft. **AAM**: Writing original draft and editing. **ARB, NA, AMT, AK**: Data collection and writing original draft. **MY**: Scale translation, review & editing. **RMG**: Conceptualisation, methodology, study supervision, and critical revision of the manuscript for important intellectual content. **CLEQ-Collaborative group**: Data collection. All the authors gave their consent.

Disclosure statement

The authors declare no competing financial or non-financial interests.

Funding

No funds were received by the authors for this manuscript.

ORCID

Ahmed Amir Samir  0009-0003-3264-031X

Kerollos Abdelsayed  0009-0001-2499-5570

Hebatalla Abdelmaksoud Abdelmonsef Ahmed  0000-0002-1882-6623

Ahmed Almahdy Mohamed  0009-0004-3320-7662

Ahmed Reda Bahr  0009-0001-8719-4738

Naji Al-bawah  0009-0001-3519-6985

Ali Malik Tiryag  0000-0002-5240-8652

Ramy Mohamed Ghazy  0000-0001-7611-706X

Data availability statement

Data is available at the reasonable request of the corresponding author.

Ethics approval and consent to participate

The study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the College of Medicine, Al-Iraqia University, Iraq (Approval No. FMSA248). Participation was entirely voluntary, and informed consent was obtained electronically from all participants after a clear explanation of the study's objectives, ensuring anonymity and confidentiality. The informed consent form was provided in Arabic, in alignment with the Arabic version of the CLEQ questionnaire and the linguistic preference of the target population.

Clinical trial number

Not applicable.

Consent for publication

Not applicable.

CLEQ-collaborative group

Samara Ali, Ahmed Abdulmughni, Saif Alaribi, Suhaib Nabil Othman, Mariam Mohamed Elsaheed, Fatma Ali Al-Humais, Kareem Abdo, Asmaa Gamal Emesh, Ahmed Al-Suraihi, Omer Thamer Ghariri, Tuqa Mohammad Etoom, Abdallah Adel Nofal.

References

- [1] Tayade MC, Latti RG. Effectiveness of early clinical exposure in medical education: Settings and scientific theories - Review. *J Educ Health Promot.* 2021;10:117. doi: [10.4103/jehp.jehp_988_20](https://doi.org/10.4103/jehp.jehp_988_20)
- [2] AlHaqwi AI, Kuntze J, Van Der Molen HT. Development of the clinical learning evaluation questionnaire for undergraduate clinical education: factor structure, validity, and reliability study. *BMC Med Educ.* 2014;14:44. doi: [10.1186/1472-6920-14-44](https://doi.org/10.1186/1472-6920-14-44)

- [3] Miri K, Hoseinabadi TS, Yaghobi A, et al. Cost management analysis of objective structured clinical examination (OSCE): guide to the universities of medical sciences. *BMC Med Educ.* 2024;24:1241. doi: [10.1186/s12909-024-06264-7](https://doi.org/10.1186/s12909-024-06264-7)
- [4] Cancarevic I, Plichtová L, Tohid H. Similarities and the differences in the education systems around the world. In: Tohid H, Maibach H, editors. *International Medical Graduates in the United States*. Cham: Springer International Publishing; 2021. p. 13–44. doi: [10.1007/978-3-030-62249-7_2](https://doi.org/10.1007/978-3-030-62249-7_2)
- [5] Samir AA, Elamir AH, Basyouni Helal M, et al. Sociodemographic, lifestyle, and psychological factors as controllable predictors of academic self-efficacy after reforming a medical education system; the Egyptian nationwide experience. *BMC Med Educ.* 2025;25:211. doi: [10.1186/s12909-025-06805-8](https://doi.org/10.1186/s12909-025-06805-8)
- [6] Badrawi N, Hosny S, Ragab L, et al. Radical reform of the undergraduate medical education program in a developing country: the Egyptian experience. *BMC Med Educ.* 2023;23:143. doi: [10.1186/s12909-023-04098-3](https://doi.org/10.1186/s12909-023-04098-3)
- [7] Roff S, McAleer S, Harden RM, et al. Development and validation of the dundee ready education environment measure (DREEM). *Med Teach.* 1997;19:295–299. doi: [10.3109/01421599709034208](https://doi.org/10.3109/01421599709034208)
- [8] Vieira JE. The postgraduate hospital educational environment measure (PHEEM) questionnaire identifies quality of instruction as a key factor predicting academic achievement. *Clinics.* 2008;63:741–746. doi: [10.1590/S1807-59322008000600006](https://doi.org/10.1590/S1807-59322008000600006)
- [9] Chan DSK. Validation of the clinical learning environment inventory. *West J Nurs Res.* 2003;25:519–532. doi: [10.1177/0193945903253161](https://doi.org/10.1177/0193945903253161)
- [10] Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med.* 2000;75:161–166. doi: [10.1097/00001888-200002000-00015](https://doi.org/10.1097/00001888-200002000-00015)
- [11] Pololi L, Price J. Validation and use of an instrument to measure the learning environment as perceived by medical students. *Teach Learn Med.* 2000;12:201–207. doi: [10.1207/S15328015TLM1204_7](https://doi.org/10.1207/S15328015TLM1204_7)
- [12] Alnaami N, Al Haqwi A, Masuadi E. Clinical learning evaluation questionnaire: a confirmatory factor analysis. *AMEP.* 2020;11:953–961. doi: [10.2147/AMEP.S243614](https://doi.org/10.2147/AMEP.S243614)
- [13] Yang L, Sun J, Wang R, et al. Cross-cultural adaptation and validation of the clinical learning evaluation questionnaire with Chinese clinical interns. *BMC Med Educ.* 2025;25:495. doi: [10.1186/s12909-025-07088-9](https://doi.org/10.1186/s12909-025-07088-9)
- [14] Ostovarfar J, Soufi SK, Moosavi M, et al. Clinical learning evaluation questionnaire: a reliable and valid tool for the evaluation of clinical education by educators and students. *BMC Med Educ.* 2023;23:618. doi: [10.1186/s12909-023-04601-w](https://doi.org/10.1186/s12909-023-04601-w)
- [15] Valdez D, Montenegro MS, Crawford BL, et al. Translation frameworks and questionnaire design approaches as a component of health research and practice: a discussion and taxonomy of popular translation frameworks and questionnaire design approaches. *Soc Sci Med.* 2021;278:113931. doi: [10.1016/j.socscimed.2021.113931](https://doi.org/10.1016/j.socscimed.2021.113931)
- [16] Davern M. Nonresponse rates are a problematic indicator of nonresponse bias in survey research. *Health Serv Res.* 2013;48:905–912. doi: [10.1111/1475-6773.12070](https://doi.org/10.1111/1475-6773.12070)
- [17] Price LR. *Psychometric Methods: Theory into Practice*. New York: The Guilford Press; 2017.
- [18] Kline RB. *Principles and Practice of Structural Equation Modeling*. Guilford Publications; 2023.
- [19] Hilton A, Skrutkowski M. Translating instruments into other languages: development and testing processes. *Cancer Nurs.* 2002;25:1–7. doi: [10.1097/00002820-200202000-00001](https://doi.org/10.1097/00002820-200202000-00001)
- [20] Gjersing L, Caplehorn JR, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC Med Res Methodol.* 2010;10:13. doi: [10.1186/1471-2288-10-13](https://doi.org/10.1186/1471-2288-10-13)
- [21] Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334. doi: [10.1007/BF02310555](https://doi.org/10.1007/BF02310555)
- [22] Alexander. Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy. In: *Statistics How To*. 2016. Accessed 14 Sept 2025. <https://www.statisticshowto.com/kaiser-meyer-olkin/>
- [23] Psychometric testing of the German versions of three implementation outcome measures. 2021. doi: [10.1007/s43477-021-00019-y](https://doi.org/10.1007/s43477-021-00019-y)
- [24] Cleare S, Gumley A, Cleare CJ, et al. An investigation of the factor structure of the self-compassion scale. *Mindfulness.* 2018;9:618–628. doi: [10.1007/s12671-017-0803-1](https://doi.org/10.1007/s12671-017-0803-1)
- [25] Costello AB, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract Assess Res Eval.* 2019;10(7).
- [26] Cheung GW, Cooper-Thomas HD, Lau RS, et al. Reporting reliability, convergent and discriminant validity with structural equation modeling: a review and best-practice recommendations. *Asia Pac J Manag.* 2024;41:745–783. doi: [10.1007/s10490-023-09871-y](https://doi.org/10.1007/s10490-023-09871-y)
- [27] Marsh HW, Guo J, Dicke T, et al. Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and Set-ESEM: optimal balance between goodness of fit and parsimony. *Multivar Behav Res.* 2020;55:102–119. doi: [10.1080/00273171.2019.1602503](https://doi.org/10.1080/00273171.2019.1602503)
- [28] Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J Acad Mark Sci.* 2015;43:115–135. doi: [10.1007/s11747-014-0403-8](https://doi.org/10.1007/s11747-014-0403-8)
- [29] Voorhees CM, Brady MK, Calantone R, et al. Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies. *J Acad Mark Sci.* 2016;44:119–134. doi: [10.1007/s11747-015-0455-4](https://doi.org/10.1007/s11747-015-0455-4)
- [30] Lee H, Smith WZ. Fit indices for measurement invariance tests in the thurstonian IRT model. *Appl Psychol Meas.* 2020;44:282–295. doi: [10.1177/0146621619893785](https://doi.org/10.1177/0146621619893785)

- [31] Chen FF. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*. 2007;14(3):464–504. doi: [10.1080/10705510701301834](https://doi.org/10.1080/10705510701301834)
- [32] Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. *Res Soc Admin Pharm*. 2019;15:214–221. doi: [10.1016/j.sapharm.2018.03.066](https://doi.org/10.1016/j.sapharm.2018.03.066)
- [33] Lave J, Wenger E. *Situated Learning: Legitimate Peripheral Participation*. 1st ed. Cambridge University Press; 1991.
- [34] Billett S. Learning through work: workplace affordances and individual engagement. *J Workplace Learn*. 2001;13:209–214. doi: [10.1108/EUM0000000005548](https://doi.org/10.1108/EUM0000000005548)
- [35] Teunissen PW, Scheele F, Scherpbier AJJA, et al. How residents learn: qualitative evidence for the pivotal role of clinical activities. *Med Educ*. 2007;41:763–770. doi: [10.1111/j.1365-2923.2007.02778.x](https://doi.org/10.1111/j.1365-2923.2007.02778.x)
- [36] Shemwell JT, Chase CC, Schwartz DL. Seeking the general explanation: a test of inductive activities for learning and transfer. *J Res Sci Teach*. 2015;52:58–83. doi: [10.1002/tea.21185](https://doi.org/10.1002/tea.21185)
- [37] Bujang MA, Omar ED, Foo DHP, et al. Sample size determination for conducting a pilot study to assess reliability of a questionnaire. *Restor Dent Endod*. 2024;49(1):e3. doi: [10.5395/rde.2024.49.e3](https://doi.org/10.5395/rde.2024.49.e3)
- [38] Jumaa ZY, Hermis AH, Dahshan A, et al. The impact of educational environment on academic thriving among medical students: insights from a multinational cross-sectional survey. *BMC Med Educ*. 2025;25:1449. doi: [10.1186/s12909-025-08053-2](https://doi.org/10.1186/s12909-025-08053-2)