MERAL Publications and Solutions Pvt. Ltd.

Original Article

Development, implementation, and preliminary psychometric evaluation of a musculoskeletal OSCE in postgraduate physiotherapy education

Claudio Chamorro Lange[1], Hernán Andrés de la Barra Ortiz[1]*

[1]Exercise and Rehabilitation Sciences Institute, School of Physical Therapy, Faculty of Rehabilitation Sciences, Universidad Andrés Bello, Santiago de Chile 7591538, Chile.

Correspondence: Hernán Andrés de la Barra Ortiz, Exercise and Rehabilitation Sciences Institute, School of Physical Therapy, Faculty of Rehabilitation Sciences, Universidad Andrés Bello, Santiago de Chile 7591538, Chile. hdelabarra@unab.cl

Received: 11 July 2026; **Revised:** 14 August 2026; **Accepted:** 19 August 2026

ABSTRACT

The Objective Structured Clinical Examination (OSCE) is widely used in health professions education to assess clinical performance under standardized conditions. However, evidence regarding its psychometric performance in postgraduate musculoskeletal physiotherapy education remains limited. This pilot study aimed to implement a musculoskeletal OSCE and examine its preliminary content validity, score distribution, internal consistency, and internal structure. A cross-sectional methodological pilot study included 20 qualified physiotherapists enrolled in a Master's Programme in Musculoskeletal Rehabilitation. The OSCE comprised three stations assessing physiotherapy assessment, therapeutic exercise, and orthopaedic manual therapy through 10 criteria scored on a 4-point analytic rating scale. Content validity was examined using the Content Validity Index and Fleiss' kappa. Internal consistency was assessed using Cronbach's alpha, and exploratory factor analysis (EFA) with principal-factor extraction and Promax rotation was used to examine preliminary internal structure. All criteria showed an I-CVI of 1.00 for relevance and representativeness, whereas clarity and item–station coherence showed greater variability. Internal consistency was satisfactory for Physiotherapy Assessment (α=0.807) and Orthopaedic Manual Therapy (α=0.789), while Therapeutic Exercise showed inconsistent relationships among its criteria (α=−0.268). Several items showed ceiling effects. EFA suggested a three-factor pattern broadly aligned with station membership, although the marginal KMO value (approximately 0.50) and small sample size require cautious interpretation. These findings informed development of a revised 12-item candidate OSCE. Prospective evaluation in a larger independent cohort is required to confirm its psychometric performance.

Keywords: Objective structured clinical examination, Clinical competence, Education, Graduate, Physical therapy specialty, Psychometrics

Introduction

Musculoskeletal disorders are a major cause of disability worldwide and place a substantial burden on individuals and healthcare systems [1]. Their management requires comprehensive clinical assessment and evidence-based interventions, including education, therapeutic exercise, and, when appropriate, manual therapy [2]. Physiotherapists therefore require more than technical proficiency; they must be able to interpret examination findings, select and adapt interventions according to the patient's presentation, and justify clinical decisions based on current evidence [3]. These abilities are especially relevant in postgraduate musculoskeletal education, where learners are already qualified physiotherapists and training is expected to develop advanced clinical reasoning and performance [3, 4]. Accordingly, postgraduate assessment

Access this article online	
Website: www.japer.in	**E-ISSN:** 2249-3379

How to cite this article: Lange CC, de la Barra Ortiz HA. Development, implementation, and preliminary psychometric evaluation of a musculoskeletal OSCE in postgraduate physiotherapy education. J Adv Pharm Educ Res. 2026;16(1):1-11. https://doi.org/10.51847/H0vnNPkbC9

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

should move beyond isolated knowledge or technical skills and capture how clinicians integrate examination findings, clinical reasoning, and intervention decisions in practice [4].

The Objective Structured Clinical Examination (OSCE) is widely used in health professions education to assess clinical performance in a structured and reproducible manner [5-7]. Through a series of predefined stations, learners are required to demonstrate specific competencies under standardized conditions. Depending on its design, an OSCE can assess practical skills, clinical reasoning, decision-making, communication, and the integration of knowledge with clinical performance [5, 6]. In physiotherapy education, this format is particularly useful because learners can be asked not only to perform an assessment or intervention, but also to select, interpret, and justify their clinical approach within a realistic scenario [8-10]. However, the usefulness of an OSCE depends not only on the clinical relevance of its stations but also on the quality of its scoring criteria and on how well these criteria represent the competencies being assessed.

Evidence on the use of OSCEs in musculoskeletal physiotherapy remains limited [11]. Available studies have largely focused on student populations and show considerable variation in station number and duration, scoring methods, competencies assessed, and psychometric evaluation [11, 12]. Although some studies have reported favorable findings for internal consistency and inter-rater reliability, comparatively little is known about the internal structure of musculoskeletal OSCEs. Evidence in postgraduate physiotherapy education is particularly scarce, although OSCE-based approaches have been explored in qualified therapists undergoing further clinical training [13]. This is relevant because postgraduate programmes must assess competencies beyond isolated technical performance, including the integration of clinical reasoning, practical skills, and evidence-based decision-making. Assessment strategies capable of making these competencies observable and systematically assessable are therefore important in postgraduate clinical education.

In this context, OSCEs may provide a useful framework for assessing advanced musculoskeletal competencies across different areas of clinical practice. Nevertheless, implementation alone does not establish the quality of an assessment instrument. Contemporary approaches to educational assessment emphasize that interpretation of scores should be supported by multiple sources of evidence, including content and internal structure [14]. Previous work in physiotherapy has also shown that examining internal consistency and factorial structure can identify weaknesses in OSCE scoring criteria and inform subsequent refinement. Preliminary assessment of content validity, score distribution, internal consistency, and internal structure may therefore help identify areas requiring modification and guide further psychometric evaluation. Accordingly, the aim of this study was to implement a three-station musculoskeletal OSCE in the Master's Programme in Musculoskeletal Rehabilitation and to examine its preliminary content validity, score distribution, internal consistency, and internal structure.

Materials and Methods

Design

A non-experimental, cross-sectional methodological pilot study was conducted to implement and preliminarily evaluate a three-station musculoskeletal OSCE designed to assess clinical performance in postgraduate physiotherapy students.

Participants and context

The OSCE was implemented on July 26, 2026, among students enrolled in the Master's Programme in Musculoskeletal Rehabilitation at Andrés Bello University, Chile. The assessment was conducted as part of the course Musculoskeletal Assessment of the Lower Extremity Based on Scientific Evidence and Clinical Reasoning, which is specifically designed for qualified physiotherapists undertaking advanced postgraduate training in musculoskeletal rehabilitation.

Twenty-four students were enrolled in the course. Twenty completed the OSCE and were included in the present analysis, whereas four were unable to attend because of health-related or work-related circumstances. The final sample consisted of 20 postgraduate students, all qualified physiotherapists, including 16 men (80%) and 4 women (20%), with a mean age of 30.8 years (SD 4.5). The course aims to develop advanced competencies in musculoskeletal assessment, clinical reasoning, therapeutic exercise, and orthopaedic manual therapy within an evidence-based framework. Student achievement in the course was assessed through two components: the OSCE, accounting for 70% of the final course grade, and a theoretical examination, accounting for the remaining 30%.

Selection criteria

Eligible participants were postgraduate physiotherapists formally enrolled in the course in which the OSCE was administered. Students were included in the analysis if they completed all three OSCE stations and had complete scoring data available for analysis. Students who did not attend the assessment or did not complete the OSCE were excluded from the psychometric analysis.

Instruments: OSCE

A three-station musculoskeletal OSCE was developed by the course instructors in accordance with the learning outcomes established for the postgraduate course and was designed to assess the integration of clinical reasoning and practical performance in lower-extremity musculoskeletal assessment and intervention. The instrument was aligned with three course learning outcomes addressing: (i) critical analysis of musculoskeletal physiotherapy

assessment; (ii) design of evidence-based physiotherapy intervention plans according to clinical assessment findings; and (iii) justification of therapeutic strategies based on scientific evidence and the patient's clinical presentation. The detailed learning-outcome mapping is presented in **Table 1**.

The OSCE comprised three stations: S1, Physiotherapy Assessment; S2, Therapeutic Exercise; and S3, Orthopaedic Manual Therapy. The complete instrument included 10 scored criteria: four in S1 and three each in S2 and S3. Each criterion was rated using a 4-point analytic rating scale ranging from 0 (not achieved) to 3 (totally achieved). Maximum scores were 12 points for S1 and 9 points each for S2 and S3, yielding a maximum overall score of 30 points. A 60% achievement criterion was applied independently to each station, corresponding to minimum passing scores of 8/12 for S1 and 6/9 for S2 and S3. Overall OSCE passing required meeting the passing criterion in all three stations.

Three clinical scenarios were used: an intermediate stage following Achilles tendon repair, arthrogenic quadriceps muscle inhibition after anterior cruciate ligament reconstruction, and chronic hip osteoarthritis. Six students were assessed using the Achilles tendon repair scenario, six using the anterior cruciate ligament reconstruction scenario, and eight using the hip osteoarthritis scenario.

The OSCE was administered as a three-station circuit. Each station lasted six minutes, including one minute to read the instructions and five minutes for task performance, resulting in a total examination time of 18 minutes per student. Three course instructors served as examiners, with one examiner assigned to each station. Examiners scored student performance using the predefined station-specific analytic criteria. This assessment represented the initial pilot administration of the complete OSCE.

Content validity was evaluated by three experts with experience in musculoskeletal physiotherapy and postgraduate education. Each expert evaluated the 10 scoring criteria in terms of relevance to the corresponding learning outcome, clarity of wording, content representativeness, and item–station coherence using a four-point rating scale.

Table 1. Structure and Blueprint of the Three-Station OSCE

Station	Station name	Aim	LO*	Modality	Criteria**	Maximum score (points)	Passing score (60% criterion)***	Clinical reasoning items	Practical skills items
S1	Physiotherapy Assessment	To assess the student's ability to select, perform, interpret, and clinically justify a physiotherapy assessment according to the health condition, rehabilitation phase, and information provided in the clinical case.	LO1, LO3	Practical, case-based OSCE station	Item 1. Physiotherapy assessment proposal: selects an appropriate assessment considering the health condition, rehabilitation phase, and clinical case. Item 2. Performance of the physiotherapy assessment: correctly performs the assessment considering procedural instructions and technical requirements. Item 3. Analysis and integration of findings: interprets the assessment findings accurately and integrates them coherently. Item 4. Clinical reasoning for the assessment: clearly and logically justifies the selected assessment procedures based on evidence.	12	8	1, 3, 4	2

S2	Therapeutic Exercise	To assess the student's ability to select, perform, and clinically justify a therapeutic exercise programme according to the health condition, rehabilitation phase, and clinical presentation.	LO2, LO3	Practical, case-based OSCE station	Item 1. Selection of the therapeutic exercise programme: selects a complete and appropriate programme considering the health condition, rehabilitation phase, and clinical manifestations. Item 2. Performance of therapeutic exercise: performs the proposed exercises considering verbal instructions, technical execution, interpretation of findings, and logical progression. Item 3. Clinical reasoning for therapeutic exercise: justifies the proposed exercises considering the biomechanics of the involved structure, stage of rehabilitation, and assessment findings.	9	6	1, 3	2
S3	Orthopaedic Manual Therapy	To assess the student's ability to select, perform, dose, and clinically justify an orthopaedic manual therapy intervention according to the characteristics and evolution of the clinical case.	LO2, LO3	Practical, case-based OSCE station	Item 1. Technique selection: selects an appropriate manual therapy intervention considering the health condition, rehabilitation phase, and clinical manifestations. Item 2. Performance of the technique: performs the technique considering therapist and patient positioning, hand placement, and direction of the force or movement vector. Item 3. Clinical reasoning for manual therapy: justifies the selected technique, identifies the therapeutic objective, and supports the dosage or strategy used to assess the response to treatment.	9	6	1, 3	2

*LO, learning outcome; OSCE, Objective Structured Clinical Examination. LO1: Critical analysis in the physiotherapy assessment of individuals with musculoskeletal conditions, integrating neurophysiological, biomechanical, biopsychosocial, evidence-based, and clinical reasoning perspectives. LO2: Design of evidence-based physiotherapy intervention plans within a biopsychosocial framework according to clinical assessment findings. LO3: Justification of therapeutic strategy selection and implementation based on available scientific evidence and the patient's health condition and clinical manifestations.

**All criteria were scored using a 4-point analytic rating scale ranging from 0 (not achieved) to 3 (totally achieved).

***A 60% achievement criterion was used for each station; the minimum passing scores were 8/12 for S1 and 6/9 for S2 and S3.

Statistical analysis

Content validity was quantified using the item-level Content Validity Index (I-CVI) and the scale-level Content Validity Index based on the average method (S-CVI/Ave) [15]. For I-CVI calculations, expert ratings of 3 or 4 on the four-point scale were considered indicative of adequate content validity. The I-CVI was calculated as the proportion of experts assigning a rating of 3 or 4 to each criterion, whereas the S-CVI/Ave was obtained by averaging I-CVI values across the 10 criteria. Inter-rater agreement among the three experts was additionally examined using Fleiss' kappa based on the original four-category ratings (1–4) [16].

Item-level scores were summarized using mean and standard deviation [mean (SD)], median and first and third quartiles [median (Q1–Q3)], observed range, and frequencies and percentages for each score category (0–3). Floor and ceiling effects were defined as the percentages of participants obtaining the minimum (0) and maximum (3) possible item scores, respectively. Station and overall OSCE scores were summarized using mean (SD), median (Q1–Q3), and observed range. Passing rates were calculated according to the predefined achievement criteria. Normality of station and overall OSCE scores was assessed using the Shapiro–Wilk test [17]; normality testing was

not applied to individual items because they were scored on a four-category ordinal scale.

Internal consistency was examined separately for each station and for the overall OSCE using Cronbach's alpha (α) [18]. Raw and standardized alpha coefficients, corrected item–rest correlations, and average inter-item correlations were calculated. Because academic performance and pass/fail decisions were determined at the station level, station-specific estimates were considered the primary internal-consistency analyses.

The preliminary internal structure of the OSCE was examined using exploratory factor analysis (EFA) with principal-factor extraction followed by oblique Promax rotation [19]. Sampling adequacy was assessed using the Kaiser–Meyer–Olkin (KMO) measure, and factor retention was guided by visual inspection of the scree plot, the eigenvalue pattern, and theoretical interpretability. Factor loadings with absolute values ≥ 0.40 were considered salient, and item uniqueness was examined to identify criteria less well represented by the extracted factors. Given the limited sample size, the factor analysis was considered exploratory rather than confirmatory. Statistical significance was set at $p < 0.05$ where applicable. All analyses were performed using Stata version 15.1 (StataCorp LLC, College Station, TX, USA).

Results and Discussion

Of the 24 postgraduate physiotherapy students enrolled in the course, 20 completed the OSCE and were included in the analysis, corresponding to a participation rate of 83.3%. Four students did not attend the assessment because of health-related or work-related circumstances. The final sample included 16 men (80%) and 4 women (20%), with a mean age of 30.8 years (SD 4.5).

Content validity and expert agreement

Table 2 presents the content-validity indices and inter-rater agreement for the 10 OSCE criteria. All items achieved an I-CVI of 1.00 for relevance and content representativeness. Item–station coherence was also high overall (S-CVI/Ave = 0.90), although S2-item3 showed an I-CVI of 0.00 for this dimension. Clarity showed greater variability across items, with an S-CVI/Ave of 0.77; lower clarity indices were observed for S2-item1 and S3-item3 (I-CVI = 0.33) and particularly for S2-item3 (I-CVI = 0.00). Fleiss' κ was 0.794 for clarity, 0.713 for representativeness, and 0.883 for item–station coherence. Fleiss' κ could not be estimated for relevance because there was no variability in expert ratings.

Table 2. Expert-Based Content Validity and Inter-Rater Agreement of the Musculoskeletal OSCE

Item	Relevance I-CVI*	Clarity I-CVI*	Representativeness I-CVI*	Item–station coherence I-CVI*
S1-item1	1.00	1.00	1.00	1.00
S1-item2	1.00	1.00	1.00	1.00
S1-item3	1.00	1.00	1.00	1.00
S1-item4	1.00	1.00	1.00	1.00
S2-item1	1.00	0.33	1.00	1.00
S2-item2	1.00	1.00	1.00	1.00
S2-item3	1.00	0.00	1.00	0.00
S3-item1	1.00	1.00	1.00	1.00
S3-item2	1.00	1.00	1.00	1.00
S3-item3	1.00	0.33	1.00	1.00
S-CVI/Ave	1.00	0.77	1.00	0.90
Fleiss' κ**	NE	0.794	0.713	0.883

Abbreviations: I-CVI, item-level Content Validity Index; NE, not estimable; S-CVI/Ave, scale-level Content Validity Index based on the average I-CVI; κ, Fleiss' kappa. Ratings of 3 or 4 on the 4-point expert rating scale were considered indicative of adequate content validity for I-CVI calculations.

**Fleiss' kappa was calculated using the original four-category expert ratings (1–4).

Internal consistency

Internal-consistency estimates for each station and for the overall OSCE are shown in **Table 3**. S1 (Physiotherapy Assessment) showed a raw Cronbach's α of 0.807 and a standardized α of 0.815, with an average inter-item correlation of 0.525 and corrected item–rest correlations ranging from 0.472 to 0.751. Similarly, S3 (Orthopaedic Manual Therapy) showed α = 0.789 and standardized α = 0.792, with an average inter-item correlation of 0.560 and item–rest correlations ranging from 0.481 to 0.728.

In contrast, S2 (Therapeutic Exercise) showed negative raw and standardized alpha coefficients (α = −0.268 and standardized α = −0.102, respectively), together with a negative average inter-item correlation (−0.032) and item–rest correlations ranging from −0.339 to 0.231. These findings indicate inconsistent relationships among the three criteria within this station. For the complete 10-item OSCE, Cronbach's α was 0.659 and standardized α was 0.661, with an average inter-item correlation of 0.163.

Table 3. Internal Consistency of the Three OSCE Stations and Overall OSCE Score (n = 20)

Station/scale	No. of items	Cronbach's α	Standardized Cronbach's α	Average inter-item correlation	Item-rest correlation, range*
S1 – Physiotherapy Assessment	4	0.807	0.815	0.525	0.472–0.751
S2 – Therapeutic Exercise	3	−0.268**	−0.102**	−0.032	−0.339–0.231
S3 – Orthopaedic Manual Therapy	3	0.789	0.792	0.560	0.481–0.728
Overall OSCE	10	0.659	0.661	0.163	0.047–0.598

Abbreviations: α, Cronbach's alpha; OSCE, Objective Structured Clinical Examination. Standardized α was calculated from the inter-item correlation matrix.

*Item-rest correlations represent corrected correlations between each item and the remaining items within the corresponding station or overall OSCE.

**Negative α values for S2 resulted from negative average inter-item covariance/correlation and indicate inconsistent relationships among the station items rather than meaningful negative reliability.

Score distribution and student performance

Item-, station-, and overall score distributions are summarized in **Table 4**. Mean station scores were 10.00 (SD 2.05) for S1, 8.05 (SD 0.69) for S2, and 6.90 (SD 1.92) for S3. The corresponding median scores were 10.5 (Q1–Q3: 8.5–12.0), 8.0 (8.0–8.0), and 7.5 (7.0–8.0), respectively. The overall OSCE mean score was 24.95 (SD 3.12) out of 30, with a median of 26.0 (Q1–Q3: 22.0–27.5).

The predefined station-level passing criterion was achieved by 16 students (80%) in S1, all 20 students (100%) in S2, and 17 students (85%) in S3. Thirteen students (65%) met the passing criterion in all three stations and were therefore classified as having passed the overall OSCE.

Several items showed marked concentration at the upper end of the scoring scale. The largest ceiling effects were observed for S1-item1 (90%), S2-item1 (85%), S2-item3 (80%), and S3-item1 (75%). Conversely, S3-item2 showed the lowest mean item score [1.80 (SD 0.70)] and a ceiling effect of only 10%.

Shapiro–Wilk testing showed no statistically significant departure from normality for S1 ($p = 0.204$) or the overall OSCE score ($p = 0.295$). S2 showed a borderline result ($p = 0.050$), whereas S3 scores showed a clear departure from normality ($p < 0.001$).

Table 4. Item-, Station-, and Overall Score Distribution and Performance of the Musculoskeletal OSCE (n = 20)

Item/station score	Mean (SD)	Median (Q1–Q3)	Min–Max	Score 0, n (%)	Score 1, n (%)	Score 2, n (%)	Score 3, n (%)	Floor effect (%)	Ceiling effect (%)	Passing criterion	Passed, n (%)
S1-item1	2.90 (0.31)	3.0 (3.0–3.0)	2–3	0 (0%)	0 (0%)	2 (10%)	18 (90%)	0	90	—	—
S1-item2	2.40 (0.68)	2.5 (2.0–3.0)	1–3	0 (0%)	2 (10%)	8 (40%)	10 (50%)	0	50	—	—
S1-item3	2.40 (0.75)	3.0 (2.0–3.0)	1–3	0 (0%)	3 (15%)	6 (30%)	11 (55%)	0	55	—	—
S1-item4	2.30 (0.73)	2.0 (2.0–3.0)	1–3	0 (0%)	3 (15%)	8 (40%)	9 (45%)	0	45	—	—
S1 total	10.00 (2.05)	10.5 (8.5–12.0)	6–12	—	—	—	—	—	—	≥8/12	16 (80%)
S2-item1	2.85 (0.37)	3.0 (3.0–3.0)	2–3	0 (0%)	0 (0%)	3 (15%)	17 (85%)	0	85	—	—
S2-item2	2.40 (0.50)	2.0 (2.0–3.0)	2–3	0 (0%)	0 (0%)	12 (60%)	8 (40%)	0	40	—	—
S2-item3	2.80 (0.41)	3.0 (3.0–3.0)	2–3	0 (0%)	0 (0%)	4 (20%)	16 (80%)	0	80	—	—
S2 total	8.05 (0.69)	8.0 (8.0–8.0)	6–9	—	—	—	—	—	—	≥6/9	20 (100%)
S3-item1	2.55 (0.89)	3.0 (2.5–3.0)	0–3	1 (5%)	2 (10%)	2 (10%)	15 (75%)	5	75	—	—
S3-item2	1.80 (0.70)	2.0 (1.5–2.0)	0–3	1 (5%)	4 (20%)	13 (65%)	2 (10%)	5	10	—	—
S3-item3	2.55 (0.69)	3.0 (2.0–3.0)	1–3	0 (0%)	2 (10%)	5 (25%)	13 (65%)	0	65	—	—
S3 total	6.90 (1.92)	7.5 (7.0–8.0)	1–9	—	—	—	—	—	—	≥6/9	17 (85%)
Overall OSCE score	24.95 (3.12)	26.0 (22.0–27.5)	19–29	—	—	—	—	—	—	Pass all three stations	13 (65%)

Abbreviations: OSCE, Objective Structured Clinical Examination; Q1, first quartile; Q3, third quartile; SD, standard deviation. S1, Physiotherapy Assessment; S2, Therapeutic Exercise; S3, Orthopaedic Manual Therapy.

Floor and ceiling effects represent the percentage of participants obtaining the minimum (0) and maximum (3) possible item scores, respectively.

A 60% achievement criterion was applied to each station; because only whole-number station scores were possible, the minimum passing scores were 8/12 for S1 and 6/9 for S2 and S3. Overall OSCE passing was defined as meeting the minimum passing score in all three stations.

Preliminary internal structure

The EFA is presented in **Table 5**. Visual inspection of the scree plot **(Figure 1)** suggested retention of three factors, with the first three eigenvalues exceeding 1 (2.73, 2.04, and 1.11, respectively), followed by a progressive flattening of the eigenvalue curve. Following principal-factor extraction and oblique Promax rotation, the resulting pattern was broadly consistent with a station-related organization of clinical performance. Factor 1 was predominantly characterized by S1 criteria, with strong loadings for S1-item2 (0.779), S1-item3 (0.811), and S1-item4 (0.871). Factor 2 included strong loadings

for S2-item1 (0.652) and S2-item3 (0.818), although S1-item1 also loaded primarily on this factor (0.674). S2-item2 showed a cross-loading pattern, with loadings of 0.432 on Factor 1 and −0.419 on Factor 2. Factor 3 was clearly characterized by all three S3 criteria, with loadings ranging from 0.579 to 0.803. Overall, the factor-loading pattern appeared to be more closely aligned with station membership than with the predefined learning-outcome grouping. However, the overall KMO value was approximately 0.50, indicating marginal sampling adequacy. Given the small sample size (n = 20), these factorial findings should be interpreted as preliminary and hypothesis-generating.

Table 5. Exploratory Factor Structure of the Musculoskeletal OSCE

Item	Station	LO	Factor 1	Factor 2	Factor 3	Uniqueness
S1-item1	S1	LO1	0.327	0.674	0.111	0.378
S1-item2	S1	LO1	**0.779**	0.084	0.188	0.287
S1-item3	S1	LO1	**0.811**	0.036	−0.244	0.340
S1-item4	S1	LO3	**0.871**	0.067	−0.038	0.230
S2-item1	S2	LO2	−0.200	**0.652**	0.061	0.588
S2-item2	S2	LO2	**0.432**	**−0.419**	0.208	0.582
S2-item3	S2	LO3	0.130	**0.818**	−0.105	0.246
S3-item1	S3	LO2	−0.168	−0.153	**0.678**	0.483
S3-item2	S3	LO2	0.106	0.155	**0.579**	0.637
S3-item3	S3	LO3	−0.000	0.002	**0.803**	0.356

Abbreviations: LO, learning outcome; OSCE, Objective Structured Clinical Examination. Exploratory factor analysis was conducted using principal-factor extraction with oblique Promax rotation. Factor loadings with absolute values ≥ 0.40 are shown in bold. The overall Kaiser–Meyer–Olkin measure was 0.50.

Based on the combined findings from content-validity assessment, internal-consistency analysis, item-level score performance, and exploratory factor analysis, selected scoring criteria were identified for refinement. The overall station structure was retained, while the scoring criteria were revised to improve specificity and reduce overlap between distinct performance components. Particular attention was given to the Therapeutic Exercise station. S2-item2 showed inconsistent relationships with the remaining station criteria and a cross-loading pattern in the EFA; therefore, technical performance was separated from dosage, adaptation, and progression. S2-item3 showed limitations in clarity and item–station coherence but a strong factorial loading, supporting retention of the underlying construct with revised wording. In the Orthopaedic Manual Therapy station, dosage and assessment of clinical response were separated from clinical justification to improve criterion specificity. The Physiotherapy Assessment station retained its original conceptual structure, with wording refined for greater clarity. These revisions resulted in four distinct criteria per station, representing selection, technical performance or clinical interpretation, dosage/adaptation where applicable, and clinical justification. The resulting 12-item candidate version of the musculoskeletal OSCE is presented in **Table 6**.

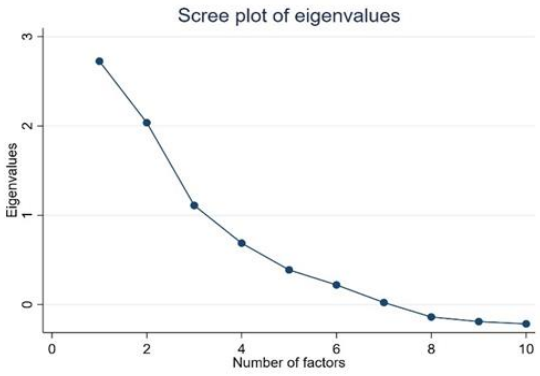

Figure 1. Scree Plot of Eigenvalues for the Exploratory Factor Analysis of the Musculoskeletal OSCE

Table 6. Proposed Revised Version of the Musculoskeletal OSCE

Station	Item	Revised performance criterion	LO
S1 – Physiotherapy Assessment	Item 1. Selection of the physiotherapy assessment	Selects and prioritizes an appropriate physiotherapy assessment according to the health condition, rehabilitation phase, clinical presentation, and information provided in the clinical case.	LO1
	Item 2. Performance of the physiotherapy assessment	Performs the selected assessment accurately and safely, following the relevant technical procedures and instructions.	LO1
	Item 3. Analysis and integration of findings	Accurately interprets and integrates the assessment findings in relation to the patient's clinical presentation and functional status.	LO1
	Item 4. Clinical justification of the assessment	Justifies the selection of the assessment strategy using clinical reasoning and relevant scientific evidence.	LO3
S2 – Therapeutic Exercise	Item 1. Selection of the therapeutic exercise intervention	Selects an appropriate therapeutic exercise intervention according to the health condition, rehabilitation phase, clinical findings, and therapeutic goals.	LO2
	Item 2. Technical performance of therapeutic exercise	Performs the selected therapeutic exercise safely and accurately, using appropriate positioning, movement execution, and clear instructions to the patient.	LO2
	Item 3. Dosage, adaptation, and progression of therapeutic exercise	Determines an appropriate exercise dosage and adapts or progresses the intervention according to the patient's performance, symptoms, clinical response, and therapeutic goals.	LO2

	Item 4. Clinical justification of therapeutic exercise	Justifies the selection of the therapeutic exercise intervention according to the clinical findings, rehabilitation phase, therapeutic goals, and available scientific evidence.	LO3
S3 – Orthopaedic Manual Therapy	Item 1. Selection of the orthopaedic manual therapy technique	Selects an appropriate orthopaedic manual therapy technique according to the health condition, rehabilitation phase, clinical findings, and therapeutic objective.	LO2
	Item 2. Technical performance of orthopaedic manual therapy	Performs the selected technique safely and accurately, considering patient and therapist positioning, hand placement, and the direction of the applied force or movement.	LO2
	Item 3. Dosage and assessment of clinical response	Determines an appropriate dosage for the selected technique according to the clinical presentation and identifies a suitable strategy to assess the patient's response to the intervention.	LO2
	Item 4. Clinical justification of orthopaedic manual therapy	Justifies the selection and proposed implementation of the manual therapy technique according to the clinical findings, therapeutic objective, rehabilitation phase, and available scientific evidence.	LO3

Abbreviations: LO, learning outcome; OSCE, Objective Structured Clinical Examination. Each criterion is scored using a 4-point analytic rating scale ranging from 0 (not achieved) to 3 (totally achieved). Each station comprises four criteria, yielding a maximum score of 12 points per station. The proposed revised version was derived from the combined findings of content-validity assessment, internal-consistency analysis, item-level performance, and exploratory factor analysis obtained during the pilot implementation. This version has not yet undergone independent psychometric evaluation and should therefore be considered a candidate instrument for subsequent validation.

This pilot study provides preliminary evidence from the implementation of a three-station musculoskeletal OSCE in postgraduate physiotherapy education. Four main findings emerged. First, the original 10-item instrument showed favorable preliminary content-validity evidence for relevance and representativeness, although selected criteria—particularly within the Therapeutic Exercise station—showed limitations in clarity and item–station coherence. Second, internal consistency was satisfactory for the Physiotherapy Assessment and Orthopaedic Manual Therapy stations, whereas the Therapeutic Exercise station showed inconsistent relationships among its criteria. Third, student performance was generally high, with marked ceiling effects for several items. Finally, the exploratory factor pattern was broadly compatible with a station-related organization of clinical performance. Together, these findings identified specific areas for refinement and informed the development of a revised 12-item candidate version of the OSCE.

OSCEs in musculoskeletal physiotherapy education

Previous research on musculoskeletal OSCEs has predominantly involved undergraduate students and shows considerable heterogeneity in examination structure, station duration, scoring methods, and psychometric evaluation [11, 12, 20]. Evidence specifically addressing postgraduate musculoskeletal education remains scarce, although OSCE-based approaches have been explored in qualified therapists undergoing further clinical training [13]. Available psychometric evidence is also partial, with greater attention to internal consistency and inter-rater reliability than to internal structure and other measurement properties [8, 11]. The present study therefore extends the available evidence by examining an OSCE specifically designed for advanced musculoskeletal training in a postgraduate setting.

Psychometric performance of the original OSCE

Content-validity findings supported the relevance and representativeness of the clinical domains included in the OSCE. However, lower clarity and item–station coherence for selected Therapeutic Exercise criteria highlight an important distinction between the relevance of a competency and the quality of its operationalization within a scoring rubric. Performance criteria should not only represent the intended competency but also describe sufficiently specific and observable behaviors to facilitate consistent examiner judgments [15, 21, 22]. This is consistent with previous evidence indicating that perceived realism or relevance alone does not constitute comprehensive validity evidence and that station-specific scoring criteria require careful development and evaluation [21, 23].

Internal consistency differed substantially across stations. Physiotherapy Assessment and Orthopaedic Manual Therapy showed coherent relationships among their criteria, with Cronbach's alpha values of 0.807 and 0.789, respectively. In contrast, Therapeutic Exercise yielded negative alpha coefficients and a negative average inter-item correlation. These negative values reflect inconsistent covariance among station criteria rather than meaningful "negative reliability" [18]. One possible explanation is that the Therapeutic Exercise station combined performance components that may not be expressed at the same level within an individual learner. While S2-item1 and S2-item3 primarily addressed intervention selection and clinical justification, S2-item2 combined technical execution, patient instruction, interpretation of findings, and exercise progression within a single score. Thus, a student could demonstrate appropriate clinical reasoning and intervention selection while performing the exercise less consistently, or conversely demonstrate adequate technical execution without an equally developed clinical rationale. This distinction between cognitive-clinical and psychomotor components is consistent with evidence showing that different domains of clinical competence may

exhibit different performance patterns within OSCEs and may be influenced by the specific context and task being assessed [11, 24, 25].

Student performance was generally high, particularly in the Therapeutic Exercise station, where all participants met the predefined passing criterion. However, item-level score distributions suggest that the different performance components varied in their level of challenge. The largest ceiling effects were observed for S1-item1, S2-item1, S2-item3, and S3-item1, which primarily assessed the selection or clinical justification of an appropriate assessment or intervention. Given that participants were already qualified physiotherapists undertaking postgraduate training, these competencies may have been relatively well established, resulting in restricted score variability. In contrast, criteria requiring direct and precise technical execution showed greater dispersion. This pattern was most evident for S3-item2, which assessed patient and therapist positioning, hand placement, and the direction of the applied force or movement and yielded both the lowest mean score and the smallest ceiling effect. A similar, although less pronounced, pattern was observed for S2-item2. These findings are consistent with evidence showing that performance may differ across clinical skills assessed within an OSCE, suggesting that reasoning, procedural performance, and other competency domains are related but not interchangeable constructs [25]. From an assessment perspective, criteria requiring direct demonstration of performance may therefore offer greater discrimination among advanced learners. Importantly, the 100% passing rate in S2 is not inconsistent with its poor internal consistency, as passing rates reflect achievement relative to a predefined standard, whereas internal consistency reflects the relationships among the criteria contributing to the station score [18].

Preliminary internal structure

The exploratory factor analysis provided additional information regarding the organization of clinical performance. The scree plot suggested three factors, and the rotated factor pattern was broadly aligned with the three clinical contexts represented by the examination. Most Physiotherapy Assessment criteria clustered together, Therapeutic Exercise criteria contributed mainly to a second factor, and all Orthopaedic Manual Therapy criteria loaded on a third factor. This pattern suggests that performance may be partly context-specific, with assessment, therapeutic exercise, and manual therapy representing distinguishable domains of clinical performance [24].

The fact that criteria assigned to the same learning outcome did not necessarily cluster together may indicate that learning outcomes operate across clinical contexts rather than functioning as independent psychometric dimensions. For example, clinical justification may represent a common curricular outcome while being expressed differently during assessment, exercise prescription, and manual therapy. Structural validity remains comparatively underexamined in physiotherapy OSCE research, where available psychometric evidence has focused largely on reliability and associations among station scores [11]. Nevertheless, the marginal KMO value of approximately 0.50 and the small sample size substantially limit the strength of this interpretation. Because EFA solutions are sensitive to sample size, factor-retention decisions, extraction methods, and rotation procedures, these findings should be interpreted cautiously [19]. The factorial findings should therefore be considered preliminary and hypothesis-generating rather than confirmation of a three-factor structure.

Refinement of the musculoskeletal OSCE

A major outcome of this pilot study was the development of a revised candidate version of the musculoskeletal OSCE. Refinement was based on convergence across content-validity findings, internal-consistency analysis, item-level performance, and the exploratory factor pattern rather than on any single statistical criterion [23, 26]. Particular attention was given to the Therapeutic Exercise station. The original S2-item2 showed inconsistent relationships with the remaining station criteria and a cross-loading pattern in the EFA, supporting separation of technical execution from dosage, adaptation, and progression. S2-item3, in contrast, showed concerns regarding clarity and item–station coherence despite a strong factorial loading; therefore, the underlying construct was retained while its wording and conceptual boundaries were refined.

A similar conceptual revision was applied to Orthopaedic Manual Therapy by separating dosage and assessment of clinical response from clinical justification. These changes resulted in a more explicit four-criterion structure for each station, with selection, performance, dosage or clinical interpretation, and justification represented as more clearly distinguishable behaviors. The proposed 12-item version therefore has potential conceptual advantages, including greater criterion specificity, reduced multidimensional scoring, improved observability of performance behaviors, and more consistent station architecture [22].

However, the revised version should not be interpreted as psychometrically superior on the basis of the present data. It was derived from findings obtained with the original instrument and has not yet been administered to an independent cohort. Demonstrating improvements in reliability, score discrimination, and structural validity will require prospective evaluation. This iterative approach is consistent with the view that validation is an ongoing process and that modifications to OSCE structure or scoring should be followed by renewed psychometric evaluation [14, 23].

Strengths, limitations, and future directions

This study provides preliminary evidence from an underrepresented postgraduate musculoskeletal setting and integrates complementary psychometric approaches to inform OSCE refinement.

Several limitations should be acknowledged. The small single-programme sample (n = 20) limits generalizability and

particularly constrains the EFA. This was the initial pilot implementation, with no prior formal pilot administration. Each station was scored by a single examiner, preventing assessment of inter-rater reliability, and content validity was evaluated by only three experts. In addition, three clinical scenarios were used without formally examining differences in case difficulty. Other sources of validity evidence, including external associations, measurement error, and generalizability across raters and cases, were not assessed. Finally, student perceptions, satisfaction, and perceived authenticity of the OSCE were not assessed. Given that learner perceptions can provide complementary information regarding the acceptability and educational experience associated with OSCE implementation, this represents an additional area for future evaluation [27].

Future studies should evaluate the revised OSCE prospectively in larger independent cohorts, include multiple raters, examine case-related variability, and reassess its internal structure. With adequate sample size, confirmatory analyses could compare station-based and learning-outcome-based models and determine whether the proposed refinements improve the psychometric performance of the instrument.

Conclusion

This pilot study provides preliminary evidence from the implementation of a three-station musculoskeletal OSCE in postgraduate physiotherapy education. Beyond its psychometric contribution, the OSCE provided a structured approach for translating postgraduate learning outcomes into observable clinical performance across physiotherapy assessment, therapeutic exercise, and orthopaedic manual therapy. Content-validity evidence was favorable overall, whereas internal consistency was satisfactory for the Physiotherapy Assessment and Orthopaedic Manual Therapy stations; specific limitations in the Therapeutic Exercise station informed targeted refinement of the scoring criteria. The exploratory internal structure was broadly consistent with a station-based organization of clinical performance. These findings supported the development of a revised 12-item candidate OSCE, which may contribute to more structured and transparent assessment of advanced musculoskeletal competencies but requires prospective psychometric evaluation in a larger independent cohort.

Acknowledgments: Exercise and Rehabilitation Sciences Institute and School of Physical Therapy, Universidad Andres Bello.

Conflict of interest: None

Financial support: None

Ethics statement: The ethical aspects of the study protocol, including the use of academic assessment data for research purposes, were approved by the Postgraduate Unit and the Faculty of Rehabilitation Sciences, Universidad Andrés Bello, in June 2026 (reference No. 162026). All participants provided informed consent for the use of their academic assessment data for research purposes.

References

1. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease Study 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2021;396(10267):2006-17. doi:10.1016/S0140-6736(20)32340-0
2. Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med. 2020;54(2):79-86. doi:10.1136/bjsports-2018-099878
3. Edwards I, Jones M, Carr J, Braunack-Mayer A, Jensen GM. Clinical reasoning strategies in physical therapy. Phys Ther. 2004;84(4):312-30.
4. Ellis R, Zeng I, Jull G, de Ruijter R, Finucane L, Lascelles R, et al. IFOMPT's educational standards and international monitoring: a member survey and review. Musculoskelet Sci Pract. 2025;75:103239. doi:10.1016/j.msksp.2024.103239
5. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1(5955):447-51.
6. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The objective structured clinical examination (OSCE): AMEE Guide No. 81. Part II: organisation and administration. Med Teach. 2013;35(9). doi:10.3109/0142159X.2013.818635
7. Criscione-Schreiber L. Turning objective structured clinical examinations into reality. Rheum Dis Clin North Am. 2020;46(1):21-35. doi:10.1016/j.rdc.2019.09.010
8. de la Barra-Ortiz HA, Gómez-Miranda LA, de la Fuente-Astroza JI. Objective structured clinical examination (OSCE) to assess the clinical skills of physical therapy students when using physical agents. Rev Fac Med. 2021;69(3). doi:10.15446/revfacmed.v69n3.83545
9. Ladyshewsky R, Baker R, Jones M, Nelson L. Evaluating clinical performance in physical therapy with simulated patients. J Phys Ther Educ. 2000;14(1):31-7. doi:10.1097/00001416-200001000-00008
10. Sandoval-Cuellar C, Alfonso-Mora ML, Castellanos-Garrido AL, Del Pilar Villarraga-Nieto A, Goyeneche-Ortegón RL, Acosta-Otalora ML, et al. Simulation in physiotherapy students for clinical decisions during interaction with people with low back pain: randomised controlled trial. BMC Med Educ. 2021;21(1):375. doi:10.1186/s12909-021-02812-7
11. Bobos P, Pouliopoulou DV, Harriss A, Sadi J, Rushton A, MacDermid JC. A systematic review and meta-analysis of measurement properties of objective structured clinical examinations used in physical therapy licensure and a

structured review of licensure practices in countries with well-developed regulation systems. PLoS One. 2021;16(8). doi:10.1371/journal.pone.0255696

12. Wessel J, Williams R, Finch E, Gémus M. Reliability and validity of an objective structured clinical examination for physical therapy students. J Allied Health. 2003;32(4):266-9.
13. Sakurai H, Kanada Y, Sugiura Y, Motoya I, Wada Y, Yamada M, et al. OSCE-based clinical skill education for physical and occupational therapists. J Phys Ther Sci. 2014;26(9):1387-97. doi:10.1589/jpts.26.1387
14. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ. 2003;37(9):830-7. doi:10.1046/j.1365-2923.2003.01594.x
15. Mokkink LB, Herbelet S, Tuinman PR, Terwee CB. Content validity: judging the relevance, comprehensiveness, and comprehensibility of an outcome measurement instrument—a COSMIN perspective. J Clin Epidemiol. 2025;185:111879. doi:10.1016/j.jclinepi.2025.111879
16. Gwet KL. Large-sample variance of Fleiss generalized kappa. Educ Psychol Meas. 2021;81(4):781-90. doi:10.1177/0013164420973080
17. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth. 2019;22(1):67-72. doi:10.4103/aca.ACA_157_18
18. Doval E, Viladrich C, Angulo-Brunet A. Coefficient alpha: the resistance of a classic. Psicothema. 2023;35(1):5-20. doi:10.7334/psicothema2022.321
19. Goretzko D, Pham TTH, Bühner M. Exploratory factor analysis: current use, methodological developments and recommendations for good practice. Curr Psychol. 2021;40(7):3510-21. doi:10.1007/s12144-019-00300-2
20. Swift M, Spake E, Gajewski BJ. Student performance and satisfaction for a musculoskeletal objective structured clinical examination. J Allied Health. 2013;42(4):214-22.
21. Nyangeni T, Ten Ham-Baloyi W, van Rooyen DRM. Strengthening the planning and design of objective structured clinical examinations. Health SA. 2024;29:2693. doi:10.4102/hsag.v29i0.2693
22. Homer M, Fuller R, Hallam J, Pell G. Shining a spotlight on scoring in the OSCE: checklists and item weighting. Med Teach. 2020;42(9):1037-42. doi:10.1080/0142159X.2020.1781072
23. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. Adv Simul (Lond). 2016;1:31. doi:10.1186/s41077-016-0033-y
24. Monteiro S, McConnell MM. Evaluating the construct validity of competencies: a retrospective analysis. Med Sci Educ. 2023;33(3):729-36. doi:10.1007/s40670-023-01794-z
25. Sim JH, Abdul Aziz YF, Mansor A, Vijayananthan A, Foong CC, Vadivelu J. Students' performance in the different clinical skills assessed in OSCE: what does it reveal? Med Educ Online. 2015;20:26185. doi:10.3402/meo.v20.26185
26. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: a review of metrics—AMEE Guide No. 49. Med Teach. 2010;32(10):802-11. doi:10.3109/0142159X.2010.507716
27. de la Barra-Ortiz HA, Gómez-Miranda LA, de la Fuente-Astroza JI. Level of satisfaction and correlation between the performance and self-evaluation of physical therapy students in the objective structured clinical examination (OSCE) when using physical agents. Rev Fac Med. 2022;70(3). doi:10.15446/revfacmed.v70n3.92473