BMC Medical Education

RESEARCH

Open Access

"Development and Reliability of an Objective Structured Clinical Examination (OSCE) for assessing clinical skills of audiology undergraduate students in Brazil"

Caroline Nunes Rocha-Muniz[1*], Maria Cecília Greco[1], Maria Beatriz Baruzzi[1], Elisiane Crestani de Miranda Gonzalez[1,2], Kátia de Almeida[1], Samira Yarak[2] and Jose Knopfholz[3]

Abstract

The aim of this study was to introduce and validate an Objective Structured Clinical Examination (OSCE) as a robust assessment tool for evaluating clinical skills in audiology among third-year audiology and speech-language students. Drawing on guidelines for OSCE development, key competencies and clinical skills in audiology were identified through expert consultation. Four OSCE stations were designed comprising one theoretical and three practical stations covering essential clinical tasks. Content validity was measured using the Content Validity Coefficient (CVC), while inter-rater reliability was assessed using the Kappa coefficient and Intraclass Correlation Coefficient (ICC). The study involved 33 audiology students and eight audiologist expert evaluators. The study demonstrated high content validity of the OSCE, with CVC scores ranging from 0.97 to 1 across all stations. Inter-rater reliability analysis revealed substantial to almost perfect agreement among evaluators, with ICC values ranging from 0.94 to 0.99. Comparison of scores between evaluators revealed minimal statistically significant differences, indicating overall consistency in assessment. The results provide substantial evidence supporting the content validity, inter-rater agreement, and reliability of the Objective Structured Clinical Examination (OSCE) as an effective tool for assessing the clinical skills of audiology and speech-language pathology students in the area of audiology.

Keywords OSCE, Objective Structured Clinical Examination, Audiology, Clinical Competence, Undergraduate

Practice points

- Implementing the Objective Structured Clinical Examination (OSCE) in audiology education offers a structured and reliable method for assessing clinical skills among audiology students.
- The OSCE demonstrated strong content validity and inter-rater reliability, ensuring consistency and fairness in evaluating students' clinical competencies.
- Use of the OSCE can address limitations in traditional assessment methods by providing a comprehensive evaluation of both theoretical knowledge and practical skills in audiology.
- By incorporating the OSCE into audiology programs, educators can promote standardized training, better preparing students for the challenges of clinical practice.

*Correspondence:
Caroline Nunes Rocha-Muniz
caroline.muniz@fcmsantacasasp.edu.br
[1] School of Medical Science, Santa Casa de São Paulo, Rua Dr Cesario Mota Jr 61, Sao Paulo, SP 01221-020, Brazil
[2] Universidade Federal de São Paulo, São Paulo, Brazil
[3] Pontifícia Universidade Católica Do Paraná, Curitiba, Brazil

- The study findings support the adoption of OSCE as an effective tool for evaluating clinical skills in audiology education, enhancing the quality and consistency of training for future audiologists.

Introduction

Assessment plays a fundamental role in the teaching–learning process, actively contributing to shaping students' profiles as outlined by the Brazilian National Curricular Guidelines. This process seeks to endow medical and health science students with a broad critical perspective that is committed to the principles of health care. Specifically, audiology and speech-language pathology (ASLP) undergraduate programs must ensure that students acquire theoretical knowledge and demonstrate clinical competencies essential for delivering effective patient care. Continued assessments during their formative process are vital to achieving these goals.

In Brazil, the ASLP program combines audiology and speech-language pathology into a single undergraduate program. Upon completing the four- or five-year program, professionals receive dual licenses, enabling them to work in both fields, provided they adhere to federal regulatory requirements. During the program's third year (or fifth semester), students typically commence supervised clinical training in audiology. At this stage, they are expected to develop foundational skills, including patient anamnesis, audiological evaluations (e.g., tone and speech audiometry, immittance), interpretation of test results, and clinical reasoning. By the end of the third year, students should be proficient in conducting less complex audiological evaluations autonomously, interpreting results accurately, and demonstrating alignment between diagnostic findings and clinical reasoning.

Traditionally, assessments during supervised clinical practice in Brazil rely on theoretical exams and observational records of students' clinical activities over time. These methods present challenges. Theoretical exams often evaluate only the foundational levels of Miller's pyramid "knows" and "knows how"—failing to assess the application of skills in practice [1]. Furthermore, observational assessments lack standardization, as they depend on case variability and instructor interpretation, potentially introducing subjectivity and inconsistency. This variability can lead to disparities in evaluation, even among students performing similar tasks, ultimately compromising the fairness and reliability of the assessment process.

To address these limitations, the Objective Structured Clinical Examination (OSCE) has emerged as a robust alternative. Developed in the 1970s by Harden and Gleeson [2], the OSCE was designed to provide a structured and standardized method for assessing clinical competence in medical education. The OSCE involves students rotating through a series of stations, each presenting a specific clinical scenario or task. At each station, students demonstrate their ability to perform clinical skills, interpret findings, and make decisions, while being evaluated using standardized checklists. This method aligns with the "shows how" level of Miller's pyramid [1, 3, 4], bridging the gap between theoretical knowledge and practical application.

Globally, the OSCE has become the "gold standard" for evaluating clinical performance in health education, with widespread adoption in fields such as medicine, nursing, dentistry, and physiotherapy [4, 5]. Its reliability and validity have been well-documented, making it a valuable tool for both formative and summative assessments [3, 6]. However, despite its global recognition, the use of OSCE in audiology education remains underexplored, particularly in Brazil. To date, there are no studies in the Brazilian context that have applied the OSCE to assess the clinical skills of ASLP students, highlighting a significant gap in the literature.

The broader implications of implementing reliable competency assessment tools like the OSCE extend beyond individual student performance. High-quality audiology education is critical for addressing the global burden of hearing loss, which affects over 1.5 billion people worldwide [7]. Ensuring that future audiologists are well-trained and competent is essential for improving public health outcomes, particularly in low- and middle-income countries where access to audiological services is limited. By adopting standardized assessment methods like the OSCE, educational institutions can enhance the quality and consistency of audiology training, ultimately contributing to better patient care and health system performance [8–10].

Thus, the OSCE can serve as an important resource for ASLP students, as it provides a comprehensive evaluation of both theoretical knowledge and clinical skills [11]. Another advantage of using the OSCE lies in the ability to plan the content to be assessed in advance, ensuring that all students are evaluated under consistent conditions. This is particularly relevant in the context of audiology, where the ability to perform and interpret diagnostic tests accurately is crucial for effective patient management [12].

Given the lack of OSCE-based assessments in audiology education in Brazil, the objective of the present study was to develop and validate an OSCE instrument for assessing the clinical skills of third year ASLP students. This study represents a significant step toward aligning Brazilian audiology education with international best practices, while addressing the limitations of traditional assessment methods. By doing so, it aims to contribute

to the improvement of audiology training and, ultimately, the quality of care provided to patients with hearing and communication disorders.

Methods

Study design

A descriptive study was performed according to the recommendations for developing an OSCE [12, 13]. The study progressed through 3 stages: (1) construction of OSCE assessment scenarios and checklists (OSCE Development Stage); (2) Content Validation Stage; and (3) OSCE Application and Inter-rater Reliability Checking Stage.

The development of the OSCE followed established guidelines for designing structured clinical assessments [2, 12]. These guidelines emphasize the importance of defining clear learning objectives, creating realistic clinical scenarios, and using standardized checklists to ensure consistency in evaluation.

The present study was conducted by the School of Medical Science of the Santa Casa de São Paulo and approved by the Research Ethics Committee (CEP) of Santa Casa de Misericórdia de São Paulo (CAAE: 77,044,823.8.0000.5479). The study results were only used following consent granted by participants via signing of the Free and Informed Consent Form.

OSCE Development stage

First, to select content for the OSCE, a list of the key clinical competencies and skills in audiology which a student on the ASLP course should have acquired by the end of year 3 (or 6th semester) of undergraduate studies was drawn up by sending out a form to all professors with expertise in the audiology area of the ASLP course of the School of Medical Science of Santa Casa de São Paulo. The competencies and skills outlined were: 1) Brief Anamnesis; 2) Conducting audiological assessments (tone and speech audiometry, and immittance) in individuals with no hearing complaints or autonomous individuals with less complex deficits; 3) Basic knowledge of applying the masking technique; 4) Interpretation of audiology results; 5) Decision-taking for performing different procedures and audiological outcomes. Based on the suggestions provided by the expert professors, author developed 4 OSCE stations based on blueprinting method: 1 theoretical station and 3 practical stations.

A description of the areas and knowledge, skills and approaches required for each station are summarized in Fig. 1.

The devising of the evaluation checklist was based on execution of the core elements of each skill. An example checklist is provided in Fig. 2. Other materials are available upon request.

Stations	Type	Areas	Competencies involved at station	Duration	Score	No. of items assessed
1	Practical	**Immittance** Experience in assessing middle ear using immittance test in simulated patient with no hearing impairments	A, B, C, E, F	10 min	20	15
2	Practical	**Pure-Tone and Speech Audiometry** Presentation, effective Communication about procedures to be performed with patient, Performing Audiometry and Speech Audiometry testing	A, B, C, D, E, F	15 min	20	15
3	Practical	**Application of Masking Technique for obtaining bone conduction thresholds** Conclude audiology test by studying bone conductance thresholds and applying masking in test of individual with shifted auditory thresholds.	A, B, C, D, E, F	15 min	20	17
4	Theoretical	**Clinical Case Study** Correctly enter data obtained from an audiology evaluation into the appropriate report (audiology evaluation sheet), correctly showing when the masking technique was needed (frequency, stimulation route - tone and/or speech). Inclusion of audiometry report; Infer aspects of causality for results presented.	A, B, D, G	30 min	40	20

Fig. 1 Stations assessed by the OSCE. A: Clinical skills; B: Knowledge and understanding; C: Technical skill; D: Critical Thinking; E: Confidence; F: Communication; G: Problem solving/Decision-making

The student.. (items evaluated)	Adequate	Partially Adequate	Inadequate	Score	Remarks
1. Introduced themselves to patient				1	
2. Performed Brief Anamnesis: Asked patient to provide information on "reason for the exam", signs,complaints, ear surgery, which is best ear.				1	

Fig. 2 Example checklist of a practical station

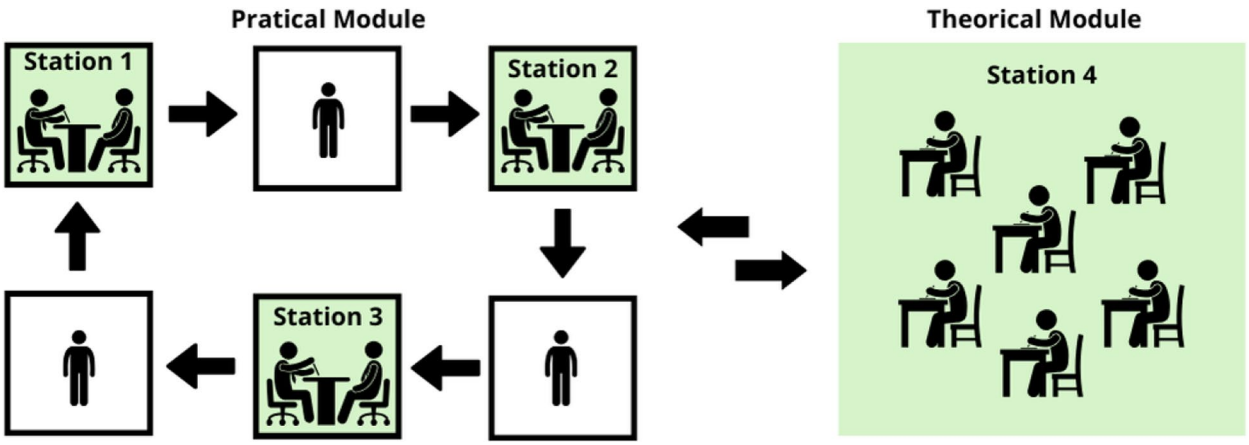

Fig. 3 Schematic depicting rotation performed on OSCE. Students rotated round Stations 1, 2 and 3, with one rest station between each. After concluding the practical module, students moved on to the theoretical module and vice-versa

Content validation stage

In this stage was included 33 third-year audiology students, 8 PhD expert evaluators with a minimum of 5 years of experience in audiology education, and 16 first-year students trained as simulated patients.

The Content Validity Coefficient (CVC) is an index proposed by Hernández-Nieto (2002) 14] to quantify and interpret the judgment of items and scales by a group of experts in the construct that the instrument proposes to measure. The CVC was used to rate each item of the checklist in terms of: a) clarity of assessment records; b) pertinence of each item – whether items are pertinent for assessing final 3rd year ASLP undergraduates; and c) relevance of evaluation – whether items are relevant to the evaluation of final 3rd year ASLP undergraduates.

Following the recommendations of Lynn (1986) [15], who asserts that content validation should involve the evaluation of at least five, we submitted the instrument to five professors on the ASLP course were consulted to rate the items on a 5-point Likert scale ranging from 0 (do not agree) to 4 (fully agree).

The experts rated the items for clarity, pertinence and relevance as follows: (1) "This item is easily understood"; (2) "This item is measuring something pertinent to that expected from a 3rd-year student"; (3) "This item is measuring something relevant for a 3rd-year student". A CVC cut-off of ≥ 0.80 was adopted for the ratings [16].

OSCE Application and inter-rater reliability checking stage

Two expert evaluators were recruited for each station to check the reliability of the checklists devised for each of the 4 scenarios.

A total of 8 PhD Professors with expertise in audiology and education, comprising 6 from the ASLP course of the School of Medical Science of Santa Casa de São Paulo and 2 from other education institutions, participated as evaluators of the OSCE.

Professor assigned as evaluators in the OSCE process were placed in pairs to evaluate each student. Evaluators

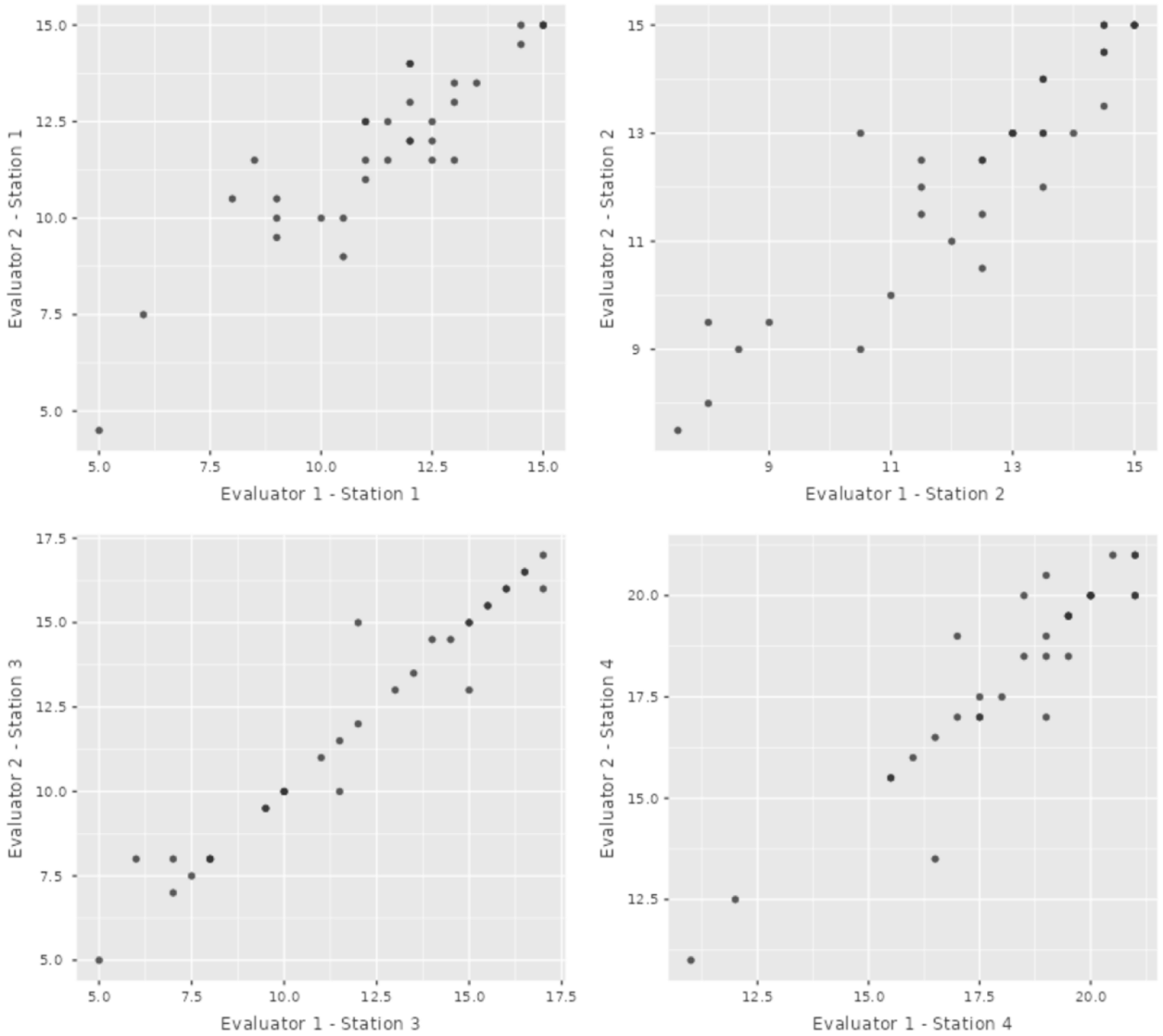

Fig. 4 Scatter plots of association between scores by Evaluators 1 and 2 for each station of OSCE

were given a script for the evaluation together with checklists. Thus, each student had 2 different evaluations.

The simulated patients were 16 students on the 1st or 2nd year of the ASLP course of the School of Medical Science of Santa Casa de São Paulo. Participants were distributed among the practical stations. The simulated patients were given a script of the simulation, guidance instructions, training and were distributed among the practical stations.

For the reliability stage, 33 students were recruited from the 6th semester of the ASLP course of the School of Medical Science of Santa Casa de São Paulo. The sample size of 33 students was determined based on convenience sampling, as it included all third year ASLP students enrolled in the undergraduate program at the time of the study.

Students were divided into small groups and rotated round the series of 4 stations. Each station included simulations of patient consultations for hearing tests, demonstration of technique, clinical reasoning, and decision-making. In order to facilitate the rotation flow, three rest stations were included.

All stations were conducted in sound-treated environments to ensure fidelity for tasks like masking in audiological exams.

During application of the OSCE, some of the students rotated round the practical module stations, while the others were allocated at the theoretical module station.

Upon concluding each module (practical or theoretical), students were directed to the next module (Fig. 3).

To minimize bias, evaluators were paired and blinded to each other's scores, and simulated patients were trained to ensure consistency in their roles.

After concluding the OSCE, the checklists of the pair of evaluators were analyzed for level of inter-rater agreement. Thus, analyses were carried out in an effort to ensure a consistent evaluation process and reduce potential bias in the scores assigned.

Given the data were categorical (Adequate – Partially Adequate – Inadequate), the Weighted Kappa Coefficient was used to analyze the level of inter-rater agreement for each checklist item. Kappa values ranged from −1 to + 1, with higher values indicating stronger agreement. The degree of agreement (reliability and accuracy) of the measures of categorical data established by the Kappa coefficient was defined as follows: Kappa value < 0.00 = no agreement; 0.00–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = substantial; and 0.81–1.00 = almost perfect [16]. In addition to the Kappa coefficient, percentage agreement (% Agreement), z-value (approximated normal test statistic) and *p*-value were determined.

The scores of the 2 evaluators were analyzed using intra-class correlation coefficient—ICC [17] as a measure of degree of agreement of scores given to each student both by station and for final score on the OSCE. The ICC confirmed consistency between the two measures, where ICC values < 0.50 = indicate poor agreement, 0.50–75 = moderate agreement, 0.75–0.90 = good agreement, and > 0.90 = excellent agreement [18].

All statistical analyses of data were performed using the statistical software SPSS, version 20.

Results

Content validation stage

The Content Validation Coefficient (CVC) for each station checklist, based on the mean of total CVC for each of the 3 aspects rated (clarity, pertinence and relevance) was 0.97 for Station 1, 0.99 for Station 2, and 1.0 for Stations 3 and 4. On the individual analysis of each item evaluated by the checklists, the CVC for each item was consistently > 0.80.

Agreement analysis

The results will be presented according to two approaches: 1) Study of inter-rater agreement for checklist items evaluated for each station (Weighted Kappa) (Tables 1, 2, 3 and 4); and 2) Study to check level of agreement of scores attributed by the pair of evaluators for each student, and level of agreement for the final OSCE score for each student (Intra-class Correlation Coefficient – ICC).

The reliability analysis was carried out using Intraclass Correlation Coefficient (ICC) (absolute agreement type; 2-wy mixed model [18]) to assess the agreement of the 2 evaluators for the scores given at Stations 1, 2, 3 and 4 for the 33 students.

Table 1 Inter-rater agreement coefficient for checklist items of Station 1

Station 1				
Checklist items rated	**% Agreement**	**Kappa**	**z**	**p**
1	93	0.73	5.38	< 0.001
2	88	0.79	5.74	< 0.001
3	70	0.43	3.79	< 0.001
4	88	0.71	4.57	< 0.001
5	94	0.77	5.56	< 0.001
6	76	0.43	3.01	0.003
7	94	0.78	5.85	< 0.001
8	82	0.68	5.62	< 0.001
9	88	0.67	4.5	< 0.001
10	82	0.70	5.48	< 0.001
11	85	0.67	5.2	< 0.001
12	76	0.47	3.77	< 0.001
13	88	0.70	5.44	< 0.001
14	91	0.79	5.67	< 0.001
15	97	0.94	5.66	< 0.001

Table 2 Inter-rater agreement coefficient for checklist items of Station 2

Station 2				
Checklist items rated	**% Agreement**	**Kappa**	**z**	**p**
1	100	1.00	5.74	< 0.001
2	91	0.80	6.04	< 0.001
3	97	0.79	5.94	< 0.001
4	100	1.00	5.74	< 0.001
5	88	0.44	2.9	0.004
6	85	0.70	5.06	< 0.001
7	88	0.77	5.02	< 0.001
8	91	0.83	5.26	< 0.001
9	85	0.71	4.46	< 0.001
10	79	0.42	3.66	< 0.001
11	70	0.54	4.32	< 0.001
12	85	0.55	4.23	< 0.001
13	94	0.89	6.64	< 0.001
14	88	0.59	4.51	< 0.001
15	100	1.00	5.74	< 0.001

Table 3 Inter-rater agreement coefficient for checklist items of Station 3

Station 3				
Checklist Items rated	**% Agreement**	**Kappa**	**z**	**p**
1	100	1.00	5.64	<0.001
2	85	0.67	4.85	<0.001
3	91	0.75	5.76	<0.001
4	97	0.93	7.06	<0.001
5	91	0.68	4.57	<0.001
6	91	0.63	4.29	<0.001
7	88	0.61	4.1	<0.001
8	100	1.00	7.13	<0.001
9	100	1.00	5.74	<0.001
10	100	1.00	7.29	<0.001
11	100	1.00	6.69	<0.001
12	94	0.89	5.76	<0.001
13	100	1.00	6.24	<0.001
14	100	1.00	5.74	<0.001
15	91	0.85	6.52	<0.001
16	97	0.95	7	<0.001
17	97	0.95	7	<0.001

Table 4 Inter-rater agreement coefficient for checklist items of Station 4

Station 4				
Checklist Items rated	**% Agreement**	**Kappa**	**z**	**p**
1	100	1.00	7.49	<0.001
2	94	0.78	5.85	<0.001
3	91	0.69	5.06	<0.001
4	94	0.86	6.32	<0.001
5	85	0.69	4.8	<0.001
6	91	0.81	5.55	<0.001
7	79	0.59	4.03	<0.001
8	85	0.70	4.96	<0.001
9	100	1.00	7.49	<0.001
10	100	1.00	7.49	<0.001
11	100	1.00	7.49	<0.001
12	100	1.00	7.49	<0.001
13	91	0.78	5.69	<0.001
14	94	0.81	6.01	<0.001
15	94	0.73	5.89	<0.001
16	100	1.00	5.74	<0.001
17	97	0.82	5.51	<0.001
18	98	0.74	5.43	<0.001
19	94	0.87	6.63	<0.001
20	94	0.88	6.79	<0.001
21	97	0.94	5.66	<0.001

Results showed strong agreement between evaluators across all stations, namely: Station 1 (ICC = 0.94; 95% CI = (0.86 – 0.97), F(32.32) = 18.9. $p < 0.001$); Station 2 (ICC = 0.96; 95% CI = (0.91 – 0.98), F(32.32) = 23. $p < 0.001$); Station 3 (ICC = 0.99; 95% CI = (0.97 – 0.99), F(32.32) = 74.1. $p < 0.001$); and Station 4 (ICC = 0.97; 95% CI = (0.93 – 0.98), F(32.32) = 28.4. $p < 0.001$).

The scatter plot of values given by evaluators 1 and 2 are plotted on the abscissa and ordinate axes for each of the 4 stations of the OSCE. Each point on the graph represents a student at the respective stations of the OSCE.

The graphs reveal a positive correlation between scores given by the evaluators at each station (Fig. 4).

Student´s *t*-test was applied to check for differences in scores given by Evaluators 1 and 2 at each of the 4 stations of the OSCE, and also for final score.

The results of comparison among mean scores given by Evaluator 1 and Evaluator 2 by station (1,2,3 & 4) and for final scores are presented in Table 5. The maximum possible scores on each station were: Station 1: 20; Station 2: 20; Station 3: 20; and Station 4:40. Thus, maximum total score for the 3 practical stations was 60 points and for the theoretical station was 40 points, giving an overall score of 0–100. No statistically significant difference was found on comparisons of mean scores given by Evaluator 1 and 2 at Stations 2, 3 and 4 or of final score. However, a statistically significant difference in scores given by Evaluators 1 and 2 was evident for Station 1.

For Station 1, the small effect size (d = 0.12) indicates that the difference, while statistically significant, is not clinically meaningful.

Nevertheless, the magnitude of this different was very small (−0.61 on a scale of 0–20).

Discussion

This study presents the development and reliability of an Objective Structured Clinical Examination (OSCE) instrument designed to assess clinical skills in audiology among third-year Audiology and Speech-Language Pathology (ASLP) students.

Recognizing the critical role of OSCEs in evaluating healthcare competencies, this investigation addresses a significant gap in the formalization of such assessments within the ASLP field in Brazil. The findings not only contribute to the field but also highlight the importance of integrating structured evaluations into the educational framework for ASLP programs.

Furthermore, the study exemplifies how clinical training practices in ASLP can be effectively implemented in simulated assessment environments like the OSCE. This approach serves as a vital bridge between theoretical knowledge and practical skills, enriching the educational

Table 5 Student´s t-test for comparing means of evaluator scores (Evaluators 1 and 2)

	Evaluator 1	Evaluator 2	Mean of differences	t	*df*	*p*-value
Station 1	15.21 ± 3.22	15.82 ± 3.05	−0.61	−2.47	32	0.02
Station 2	16.46 ± 2.92	16.36 ± 2.86	0.10	0.49	32	0.63
Station 3	14.03 ± 4.24	14.10 ± 4.09	−0.07	−0.42	32	0.68
Station 4	34.78 ± 4.53	34.55 ± 4.70	0.23	0.78	32	0.44
Final score	80.48 ± 11.62	80.83 ± 11.22	−0.34	−0.71	32	0.48

Mean ± standard deviation; df = degrees of freedom

experiences of students and fostering competency-based learning.

While the OSCE framework employed was adapted from internationally recognized methodologies, the results of this study emphasize its successful application and effectiveness in the Brazilian context, where standardized assessments in audiology remain underutilized. The findings demonstrate the adaptability of OSCEs to diverse educational settings and their potential as a replicable model for regions with similar resource limitations.

To the best of our knowledge, this is the first study to develop and evaluate an OSCE instrument specifically tailored to assess audiology skills and competencies among undergraduate ASLP students in Brazil. This pioneering effort provides a foundation for future research and encourages broader adoption of OSCEs as a standardized tool for competency assessment in ASLP education.

Development of OSCEs

OSCEs were first introduced in the 1970s as a final exam on medicine courses in Dundee by Harden et al., assessing over 120 students in a single morning. Today, the number of stations used in an OSCE varies widely [19].

Compared with the traditional model, the OSCE in the present study differs with respect to the lower number of stations and different times between each station, potentially introducing bias: 1) regarding number of stations, studies show that the reliability of the OSCE is positively associated with greater number of stations [20]; and 2) regarding differences in time between stations, where the logistics of transition of students between stations calls for special care, with the inclusion of more rest stations.

Although the current model developed involves a smaller number of stations compared with the traditional model, the current view of the original author is the possibility of flexibility afforded by the OSCE [19]. This feature allowed the development and structuring of the OSCE in accordance with the professional competencies in clinical audiology required for final 3rd-year ASLP undergraduates.

Content validity

Content validation is an important step in developing an assessment instrument such as the OSCE. This step seeks to ensure the instrument accurately and comprehensively captures the skills and competencies it was designed to measure.

In the present study, the content validation stage entailed checking the clarity, pertinence and relevance of items of the checklist and of the 4 evaluation stations of the OSCE.

The results showed high Content Validation Coefficients (CVC), i.e., a high level of agreement was confirmed among the experts on evaluations of clarity, pertinence and relevance for the items making up the checklist devised. This input from the experts consulted in the process of content validation contributes to the reliability of the instrument. In addition, the consistency of the results of the content validation among the evaluators further vouches for the reliability of the instrument, providing a solid foundation for the subsequent stages of implementation and rating of the OSCE.

Thus, the robust content validation reveals a convergence of expectations of the experienced professionals in the field of Audiology and constitutes a key stage to guarantee the validity and reliability of the OSCE as a tool for assessing the clinical skills of ASLP students in the Brazilian milieu.

Inter-rater agreement

The results of the analysis of agreement (Kappa) showed moderate-to-substantial agreement among raters for all items of the standardized checklist across all 4 stations. These results conflict with the findings of some studies which have reported only low-to-moderate inter-rater agreement and indicated the need to train raters prior to assessment [21, 22]. In this respect, all of the raters except one (rater 2, station 1) involved in the stations of the current OSCE held prior knowledge of the application. Rater 2 had the necessary expertise but had never taken part in the development of an OSCE.

The results for Stations 1 and 2 revealed variability in item reliability, particularly for certain checklist items

with lower Kappa values. For example, items 3 and 6 in Station 1 and items 5 and 10 in Station 2 showed moderate agreement, with Kappa values ranging from 0.42 to 0.55. This variability may stem from the inherent subjectivity involved in evaluating clinical reasoning and decision-making tasks, which are emphasized in these stations.

Previous studies have highlighted that task requiring assessors to interpret nuanced student behaviors, such as critical thinking and problem-solving, can yield lower inter-rater reliability [20, 22]. To address this, future iterations of the OSCE could incorporate additional training sessions for evaluators, focusing on standardizing scoring criteria for complex tasks. Enhanced training has been shown to significantly improve inter-rater agreement in similar assessments [23, 24].

The lower inter-rater agreement for certain items in Stations 1 and 2 may reflect the subjective nature of assessing communication skills and clinical reasoning. While these items are essential for evaluating holistic clinical competence, future studies should focus on refining scoring criteria and providing rater training to enhance reliability. Despite these challenges, retaining these items is justified given their importance in preparing students for real-world clinical practice.

The moderate-to-substantial agreement found in the present study, replicating the results of Sobh et al. [23] and Beckett et al. [24], serve to corroborate the importance of having faculty raters who familiar with and trained on the OSCE knowledge. This strong agreement among different raters is essential to ensure fair and reliable evaluation.

The statistical significance of the Kappa values supports the reliability of the assessment process. Moreover, the high percentage agreement among the items rated, in conjunction with significant Kappa values, corroborates the consistency and reliability of the instrument, reflecting the robustness of the assessment of the clinical skills of the students. These results are consistent with previous studies highlighting the role of good inter-rater agreement on the validity and reliability of OSCE in different areas of health [23, 24].

Assessment reliability

Reliability reflects the degree in which scoring of an instrument can reproduce results. In the present study, the analysis of reliability using the Intraclass Correlation Coefficient (ICC) revealed high correlation in scoring of checklist among evaluators across all stations of the OSCE. The difference in evaluator scores was minor, indicating a consistent reliable assessment. These results confirm the inter-rater consistency of the instrument, further confirming the reliability of the OSCE developed.

Analysis of the scatter plots corroborates the ICC results, showing a positive correlation between scores given by the evaluators at each station. This suggests the evaluators have a similar understanding of the clinical competencies of the students assessed.

The reliability of the OSCE developed by this investigation can serve as a solid basis for informing educational decisions and continued improvement of the undergraduate ASLP program.

Comparison among evaluators at each station

Student´s *t*-test was used to compare the scores given by the different evaluators (Evaluator 1 and Evaluator 2) at each station.

Statistical comparison of scores given by the two evaluators revealed that, in general, there were no significant differences in the mean of scores between the evaluators. This lends further support for the reliability and consistency of the assessment process, demonstrating objectiveness in attribution of scores across the different stations of the OSCE.

It is notable that, even where a statistically significant difference was detected (station 1), the magnitude of these differences was very small. This suggests that discrepancies in assessments have a minimal impact on students´ final scores.

Therefore, these results reinforce the evidence of the consistency and validity of the assessment instrument.

Our study addresses a critical gap in audiology education by validating an OSCE for assessing clinical skills in a Brazilian context, where such tools are underexplored.

The findings are relevant to an international readership for some reasons: (1) the global burden of hearing loss underscores the importance of high-quality audiology education [7]; (2) the detailed description of the OSCE stations and checklists are available from the corresponding author on request, allows other institutions to replicate the study in their own contexts; and (3) the OSCE items were developed in alignment with international standards for audiology education, ensuring relevance to a global audience.

Practical implications

The development and consistency in the results found suggest the OSCE is an effective tool for assessing clinical skills of ASLP students, offering several practical benefits [2, 12, 19], including:

- The implementation of the OSCE in audiology education offers a structured reliable method for assessing the clinical skills of ASLP students.

- The OSCE exhibits strong content validity and inter-rater reliability, assuring consistency and fairness in assessments of students´ clinical competencies.
- Use of the OSCE can overcome the limitations of traditional assessment methods, providing a comprehensive evaluation both for theoretical knowledge and practical skills in audiology.
- By incorporating the OSCE into audiology programs, educators can promote standardized training, and better prepare students for the challenges of clinical practice.
- The findings support the adoption of the OSCE as an effective tool for assessing clinical skills in audiology education and identifying gaps in training, thereby allowing improvement in the quality and consistency of training for future audiologists.

Study limitations

Despite the promising results, some limitations of the study should be pointed out. One of the major limitations was the relatively small sample size. Future studies involving larger samples can further validate the use of the OSCE for assessing ASLP students in the Brazilian context. Another relevant limitation is the generalization of results, in as far as the study was conducted at a single institution of higher education in Brazil. Hence, the present study results should be interpreted with caution and the need for further replication of the study in other settings should be recognized. Deeper investigations of construct validity and of other methods for reducing potential assessment bias are warranted.

While our study was conducted in a single institution, the findings have important implications for audiology education globally. The OSCE was designed to assess core competencies in audiology, such as conducting audiological evaluations, interpreting results, and applying masking techniques, which are universally relevant. The strong content validity and inter-rater reliability observed in our study suggest that the OSCE can be adapted to other educational contexts, particularly in low- and middle-income countries, where standardized assessment tools are scarce. However, we acknowledge that differences in institutional resources, rater training, and student motivation may affect the generalizability of our findings [12].

Future research involving larger and more diverse samples across multiple institutions will be essential to confirm the generalizability of these findings and to further enhance the reliability of the OSCE as a competency assessment tool in audiology education.

Conclusion

This study demonstrates the feasibility and reliability of an OSCE for assessing clinical skills in audiology among third year ASLP students in Brazil.

While the observational nature of the study limits the generalizability of the findings, the results provide robust evidence supporting the content validity, inter-rater agreement, and reliability of the OSCE.

Future studies should involve larger, multi-institutional samples to further validate the OSCE and explore its adaptability to diverse educational contexts. By addressing these limitations, the OSCE can serve as a valuable tool for enhancing the quality and consistency of audiology education worldwide.

Acknowledgements

I would like to express my deepest gratitude to my esteemed professors and colleagues at Santa Casa de São Paulo School of Medical Science for their invaluable support throughout this research.

Disclosure Statement

Authors have no conflicts of interest to report.

Authors' contributions

CNRM conceived and designed the study, collected data, analyzed and interpreted the data, and contributed to drafting and reviewing the manuscript. MCG, MBB, EMG, KA collected data and contributed to drafting and reviewing the manuscript. SY and JK conceived and designed the study, interpreted the data, contributed to drafting and reviewing the manuscript.

Funding

Not applicable.

Data availability

All the data is presented as a part of tables or figures. Additional data can be requested from the corresponding author.

Declarations

Ethics approval and consent to participate

Informed consent was obtained from all participants. Ethics approval was obtained Research Ethics Committee for Analyzing Research Projects (CAAE: 77044823.8.0000.5479). The study results were only used following consent granted by participants via signing of the Free and Informed Consent Form.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 17 October 2024 Accepted: 17 March 2025
Published online: 05 May 2025

References

1. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65:S63–7. https://doi.org/10.1097/00001888-199009000-00045.
2. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). Med Educ. 1979;13:41–54.

3. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. BMJ. 1975;1(5955):447–51. https://doi.org/10.1136/bmj.1.5955.447.
4. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & Administration. Med Teach. 2013;35(9):e1447–63. https://doi.org/10.3109/0142159X.2013.818635.
5. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The Objective Structured Clinical Examination: The new gold standard for evaluating postgraduate clinical performance. Ann Surg. 1995;222:735–42.
6. Fleming PR, Manderson WG, Matthews MB, Sanderson PH, Stokes JF. Evolution of an examination: M.R.C.P. (U.K.). Brit Med J. 1974;2(5910):99–107. https://doi.org/10.1136/bmj.2.5910.99.
7. World Health Organization (WHO). World report on hearing. 2021.
8. American Speech-Language-Hearing Association. Standards and implementation procedures for the certificate of clinical competence in audiology. 2020. https://www.asha.org/certification/2020-audiology-certification-standards/.
9. Accreditation Commission for Audiology Education. Accreditation standards for the doctor of audiology (Au.D.) program. 2005. https://acaeaccred.org/wp-content/uploads/2016/02/ACAECurrentStandards.pdf.
10. Diretrizes Curriculares Nacionais do Curso de Graduação em Fonoaudiologia. 2002. http://portal.mec.gov.br/cne/arquivos/pdf/CES052002.pdf.
11. Bhatnagar KR, Saoji VA, Banerjee AA. Objective structured clinical examination for undergraduates: Is it a feasible approach to standardized assessment in India? Indian J Ophthalmol. 2011;59(3):211–4. https://doi.org/10.4103/0301-4738.81032.
12. Daniels VJ, Pugh D. Twelve tips for developing an OSCE that measures what you want. Med Teach. 2018;40(12):1208–13. https://doi.org/10.1080/0142159X.2017.1390214.
13. Jones A, Pegram A, Fordham-Clarke C. Developing and examining an Objective Structured Clinical Examination. Nurse Educ Today. 2010;30(2):137–41. https://doi.org/10.1016/j.nedt.2009.06.014.
14. Hernández-Nieto RA. Contributions to statistical analysis. Mérida: Universidad de Los Andes; 2002.
15. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382–385. Retrieved from https://journals.lww.com/nursingresearchonline/pages/default.aspx.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biome- trics. 1977;33:159–74.
17. Fleiss JL. (Ed.). The design and analysis of clinical experiments. New York: Wiley; 1986.
18. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
19. Harden RM, Lilley P, Patricio M, Norman G. The definitive guide to the OSCE: The objective structured clinical examination as a performance assessment. Edinburgh: Elsevier; 2015.
20. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. Med Educ. 2011;45(12):1181–9. https://doi.org/10.1111/j.1365-2923.2011.04075.x.
21. Ali M, Pawluk SA, Rainkie DC, Wilby KJ. Pass-fail decisions for borderline performers after a summative objective structured clinical examination. Am J Pharm Educ. 2019;83(2): 6849. https://doi.org/10.5688/ajpe6849.
22. Wilby KJ, Black EK, Austin Z, et al. Objective structured clinical examination for pharmacy students in Qatar: cultural and contextual barriers to assessment. East Mediterr Health J. 2016;22(4):251–257. https://doi.org/10.26719/2016.22.4.251.
23. Sobh AH, Austin Z, Izham MIM, Diab MI, Wilby KJ. Application of a systematic approach to evaluating psychometric properties of a cumulative exit-from-degree objective structured clinical examination (OSCE). Curr Pharm Teach Learn. 2017;9:1091–8. https://doi.org/10.1016/j.cptl.2017.07.011.
24. Beckett RD, Gratz MA, Marwitz KK, Hanson KM, Isch J, Robison HD. Development, validation, and reliability of a P1 objective structured clinical examination assessing the national EPAs. Am J Pharm Educ. 2023;87(6):100054. https://doi.org/10.1016/j.ajpe.2023.100054.

Publisher's Note