Zhang et al. BMC Medical Education (2025) 25:605
https://doi.org/10.1186/s12909-025-07091-0

BMC Medical Education

RESEARCH Open Access

Network analysis of an OSCE-based graduation skills assessment for clinical medical students

Huiqun Zhang[1], Shanshan Li[2], Guoquan Zheng[2], Xiaoyun Chen[1], Lujie Zhong[3], Jiaying Li[3] and Yun Li[3*]

Abstract

Background This study investigates OSCE performance among Chinese undergraduate medical students to identify differences and analyze the interconnections among OSCE modules. Employing network analysis, we map the relationships between modules to determine their predictive strength and centrality to overall clinical competence, offering a novel perspective on skill integration and curriculum design.

Methods Analyzing data from 3,710 students across three medical specialities (clinical medicine, anesthesiology, medical imaging) from a leading medical institution in Guangzhou, China (Grade 2011–2016), and the independent samples t-tests, analysis of variance (ANOVA), Pearson's correlation, and network analysis were employed to dissect performance trends, compare demographic group differences, and unearth the correlations among the examination modules.

Results Significant differences were identified between genders, with female students consistently outperforming male students across modules and overall scores except for auxiliary examination analysis. Besides, specialities differed significantly in physical examination, basic operational skills, auxiliary examination analysis, and case analysis. Network analysis revealed that the nodes "Pediatric Skills" and "Nursing Skill" exhibit the strongest linkage, closely followed by "Cardiac Auscultation" and "Pulmonary Auscultation". "General Physical Examination" is the most influential, with "Pediatric Skills" and "Gynecological Obstetrical Skills" also prominent.

Conclusion From the perspective of medical education reform, reinforcing core skills like "General Physical Examination" and integrating modules to enhance skill synergy is pivotal. Implementing targeted educational interventions, including advanced training and clinical practice courses, will not only address performance disparities but also bolster students' clinical acumen and communication skills, laying a robust foundation for superior patient care. The network analysis approach reveals how module performances interrelate, highlighting central modules whose improvement can boost related competencies. It complements traditional statistics, visualizing these relationships to inform OSCE reforms and enrich medical education.

Keywords OSCE, Medical education, Medical students, Network analysis, Clinical skills

*Correspondence:
Yun Li
liyun@gzhmu.edu.cn
[1]Practical Teaching Center, Guangzhou Medical University, Guangzhou 511436, China
[2]Academic Affairs Office, Guangzhou Medical University, Guangzhou 511436, China
[3]School of Health Management, Guangzhou Medical University, Guangzhou 511436, China

Background

The Objective Structured Clinical Examination (OSCE), first introduced by Harden et al. [1], is widely recognized as a robust method for assessing clinical skills. It offers a practical and reliable alternative to traditional assessment methods, providing a structured environment where candidates rotate through various stations that simulate real-world clinical scenarios. These stations assess theoretical knowledge, practical skills, and analytical thinking using standardized criteria. Over time, the OSCE has gained global acceptance due to its ability to assess a wide range of clinical competencies, including communication skills, physical examinations, and clinical decision-making [2].

However, the practice of OSCE also needs to take into account national circumstances and differences in healthcare systems. From an international perspective on medical talent cultivation, western countries like the United States usually begin specialized training after students have completed their undergraduate education and have entered residency programs. In China, the education of specialized clinical medicine sub-disciplines, such as anesthesiology and medical imaging, generally starts at the undergraduate level. This difference is closely related to the historical development of medical education and employment needs in China [3]. To tackle the shortage of highly skilled health professionals, China has adopted various educational strategies at different levels, leading to a unique medical education system. Undergraduate medical education in China typically lasts for five to eight years [4], where students immediately join specialized tracks such as clinical medicine, anesthesiology, or medical imaging. In the United States, students generally enter medical school after completing a four-year undergraduate degree and only start specializing during their residency training [5, 6].

Chinese clinical medicine students undergo comprehensive training in general medical practice, including basic skills like medical history taking, physical examination, and diagnosis. Meanwhile, anesthesiology students concentrate more on perioperative care and anesthetic techniques, while medical imaging students focus on diagnostic tools such as radiology. Comparatively, U.S. medical education follows a consistent structure with two years of pre-clinical education followed by two years of clinical rotations, with specialization only beginning during residency. Chinese medical students tend to focus on acquiring clinical competencies earlier in their academic careers.

Despite its widespread use, there are still gaps in the literature regarding the comparison of OSCE performance between different cultures and student groups. Existing studies tend to rely on small sample sizes and often neglect subgroup comparisons between gender, grade, and major. In addition, while OSCE covers key clinical competencies such as history collection, cardiopulmonary auscultation, and case analysis, few studies have explored the interrelationships between these examination components and their predictive value in assessing overall clinical performance. Traditional statistical methods used in previous studies (such as t-test, ANOVA, correlation analysis) may only show simple associations or differences between modules, while the addition of network analysis can more intuitively present the complex relationship network between modules, reveal which modules are core and which modules are closely connected, and thus provide a more comprehensive perspective for understanding the structure of OSCE. Moreover, the unique role of network analysis in determining the interdependence and predictive value between modules, such as by calculating the centrality index of nodes, can identify the key modules that have a large impact on the overall clinical competence, which is important for curriculum design and teaching intervention, which is difficult to achieve with traditional methods.

In response to these gaps, our study evaluates the performance of students from three clinical majors on the OSCE, focusing on differences by gender, major, and grade based on the Chinese medical background. Additionally, we employ network analysis to investigate the interconnections among OSCE modules. This method allows us to determine the predictive strength of each test item and identify which components are foundational to the exam structure. Specifically, we analyze students' performance in each OSCE module, compare subgroup differences, and evaluate the correlation between each module and the total score. Finally, we assess the centrality of each module within the exam framework, providing insights into the relative importance of different clinical skills for overall performance. We anticipate offering recommendations for the localization reform of the OSCE to enrich the diversity of medical education systems within cultural contexts.

Methods

Participant

This study analyzed a cohort of 3,710 undergraduate students (females: 1,804) from a distinguished medical institution in Guangzhou, China. This cohort, spanning from 2011 to 2016, comprised students majoring in three clinical medicine disciplines(clinical medicine, anesthesiology, medical imaging). Specifically, it included 2,793 students focusing on clinical medicine, 372 in anesthesiology, and 545 in medical imaging. The grade distribution of these students was as follows: 580 from the class of Grade 2011, 452 from Grade 2012, 513 from Grade 2013, 756 from Grade 2014, 666 from Grade 2015, and 743 from Grade 2016, indicating a diverse and extensive sample reflective of the institution's academic breadth during these grades.

Examination organization and implementation

According to the talent cultivation objectives, teaching syllabus and internship syllabus of the clinical medical specialty, and concerning the national qualification examination for clinical practicing physicians, the graduation skills assessment was set up in 9 stations, with the examination time for stations 1–8 being 15 min each, and that for station 9 being 35 min, making a total of 155 min. The content of the examination includes history taking, physical examination, basic operation, auxiliary examination, case analysis, medical humanities, etc. The subjects involved are mainly internal medicine, surgery, obstetrics and gynecology, pediatrics, emergency medicine and nursing, etc. The total score of each item was 100 points, and the standard score was calculated by multiplying the score of the selected item by its corresponding weight. The standard score was also scored with 100 points to the full score. The pass mark for all assessment items and modules was 60 points. The specific stations are shown in Table 1.

The annual examination took place at the Clinical Skills Laboratory Center of the university after fifth-year students completed their clinical internship subjects. Participants were randomly divided into groups, with two groups of 16 students each taking the exam simultaneously. In total, 96 students completed the test in three batches on the same day. Abbreviations for technical terms were explained upon first use, and the language remained objective and value-neutral throughout. The text adhered to common academic conventions, including a formal register, clear structure with logical progression and precise vocabulary. Any bias was avoided, and the grammar, spelling and punctuation were error-free. To achieve a just examination process, each student's exam content is randomly chosen by the graduation examination system on the day of the exam.

Examiners

Examiners are selected annually from each clinical college and affiliated hospital of the institution, all of whom are clinical instructors with long-term involvement in clinical teaching and rich teaching experience. Most instructors have also served as examiners for graduation examinations multiple times. To ensure the examinations are conducted smoothly, orderly and efficiently, examiners undergo training each year before the examinations. This includes organized demonstrations at various examination stations to ensure examiners are thoroughly acquainted with the examination models, procedures and scoring system. The aim is to clarify examination requirements, unify scoring standards and ensure the examinations are fair, just, scientific and reasonable.

Statistical analysis

We first calculated each examination item's mean score, standard deviation, as well as pass rate. Referring to the content of the National Clinical Practitioner Qualification Examination (NCPQE) Practical Skills Examination, the assessment was divided into five modules: history taking, physical examination, basic operation skills, auxiliary examination and case analysis. Mean score, standard deviation and pass rate were counted for each module, and the student's proficiency in each module was

Table 1 OSCE exam station settings

Station	Contents	Duration	Weight	Note
Station 1	Medical history collection I: internal medicine/pediatrics	15	8%	One of two
	Abnormal abdominal palpation I: hepatomegaly/splenomegaly		2%	One of two
Station 2	Normal physical examination: basic four signs + abdominal, spinal examination/cardiac examination /thoracic and pulmonary examination/head and neck extremities neurological system	15	10%	One of four
Station 3	Internal pediatrics operation: bone puncture/lumbar puncture/chest puncture/abdominal puncture	15	10%	One of four
Station 4	Emergency skill operation: cardiopulmonary resuscitation, defibrillation, and electrocardiogram examination	15	10%	Mandatory
Station 5	History taking II: surgery/obstetrics and gynecology	15	8%	One of two
	Abnormal abdominal palpation II: positive Murphy's sign/positive pressure at Mac's point		2%	One of two
Station 6	Surgical skills: gowning, gloving, and disinfecting drapes/surgical instrument recognition, suturing and knotting, and wound dressing	15	10%	One of two
Station 7	Obstetrics and gynecology skills: gynecological examination/obstetric examination	15	10%	One of two
Station 8	Nursing skill operation: gastric tube/urinary catheterization	15	10%	One of two
Station 9*	Cardiac auscultation	5	2.5%	Mandatory
	Lung auscultation	5	2.5%	Mandatory
	Analysis of ancillary findings: laboratory tests, electrocardiogram, imaging	25	6%	Mandatory
	Case analysis		9%	Mandatory

*The grade of 2015 was affected by the epidemic and did not include a ninth examination station compared. To compare performance differences between genders, we utilized an independent samples t-test. Furthermore, we used one-way ANOVA to analyze differences between students of different grades, with a significance level of $p < 0.05$ indicating statistical differences. This study used "tidyverse" and "dplyr" packages for data collation, the "psych" package for descriptive statistics and the "ggpubr" package for ANOVA and visualization. In addition, we investigated the relationship between the different exam modules as well as the total scores using Pearson's correlation coefficient through R Studio, utilizing the "Hmisc" and "tidyr" packages. In order to control the false discovery rate (FDR) and maintain statistical integrity, the p-values of the correlation coefficients were adjusted using the Benjamin Hochberg procedure.

Network estimation

As a final step, we used the "qgraph" package to estimate and visualize two networks to explore the correlation between basic and practical skills [7]. While the EBICglasso and centralityPlot functions from the "qgraph" package were used to visualize the network and calculate the significance of each node [8]. This study calculated four significant neutralization coefficients, namely strength, closeness and betweenness as well as expected influence. And in this study, we focus on the EI values and regard them as the most important centrality indicator [9]. Expected influence (EI) is a measure calculated by adding up the weights of edges connected to a node in a network. Unlike closeness and betweenness, which can be unreliable in some cases [7]. EI is particularly useful for identifying central nodes in networks with both positive and negative connections. Nodes with high EI values have strong and frequent connections with other nodes, i.e. they occupy a central position in the network.

Estimation of network accuracy and stability

The "bootnet" package was employed to evaluate the accuracy and robustness of the network estimation by the bootstrap technique [7]. Specifically, we tested the robustness of two issues related to robustness: (1) the robustness of node weights; (2) the robustness of node centrality. A 95% confidence interval was estimated for node weights to test the robustness of the parameter (i.e., the interval contains a 95% probability of its true value). We then examined differences in edge weights and centrality values based on 1,000 bootstrapping iterations at the 0.05 level for differences in edge weights and centrality values. Secondly, correlation stability (CS) coefficients were estimated via deletion of subjects and re-estimation of the network (e.g., node-dropping and sub-setting bootstrapping). CS (cor = 0.7) represents the maximum acceptable sample reduction, with a minimum of no less than 0.25, and generally above 0.5 is acceptable.

Results

Performance in different assessment items

Each candidate was required to go through nine examination stations for assessment, complete operations, answer questions, and provide online responses as specified by the assessment item cards at each station. The statistical results in Table 2 indicate that among all assessment items, the highest average score was achieved in abnormal abdominal examination II (87.82 ± 12.90), while the lowest was in the interpretation of auxiliary examination analysis (49.56 ± 18.39). The average scores for auxiliary examination analysis, cardiac auscultation and lung auscultation were all below 60. The pass rate for most assessment items was relatively high, with the internal pediatrics skill operation achieving the highest rate at 98.95%. In contrast, lower pass rates were observed in cardiac auscultation (60.38%), lung auscultation (58.91%),

Table 2 Descriptive statistics for different assessment content

Variables		Min	Max	Mean	SD	Pass rate (%)
Station 1	Medical history collection I	0.00	99.88	77.11	9.98	94.82
	Abnormal abdominal palpation I	0.00	100.00	86.50	12.59	96.52
Station 2	Normal physical examination	17.00	100.00	79.69	11.19	96.57
Station 3	Internal pediatrics skill operation	6.00	99.00	81.84	9.00	98.95
Station 4	Emergency skill operation	23.00	99.50	78.59	10.34	96.12
Station 5	Medical history collection II	28.75	100.00	80.66	11.18	94.36
	Abnormal abdominal examination II	0.00	100.00	87.82	12.90	96.55
Station 6	Surgical skill operation	30.00	100.00	82.71	10.04	98.03
Station 7	Obstetrics and gynecology skill operation	26.00	100.00	77.88	10.04	97.71
Station 8	Nursing skill operation	1.00	100.00	71.65	11.76	89.46
Station 9	Auxiliary examination analysis	0.00	100.00	49.56	18.39	36.87
	Case analysis	0.00	98.33	61.79	14.85	56.42
	Cardiac auscultation	0.00	100.00	56.91	25.84	60.38
	Lung auscultation	0.00	100.00	56.01	25.24	58.91

All scores are compared after being converted to a scale of 100

Table 3 Performance of different modules

Variables	Min	Max	Mean	SD	Pass rate (%)
Medical history collection	28.75	99.44	78.85	8.43	97.90
Physical examination	31.08	98.12	76.09	9.70	94.93
Basic operational skill	18.70	97.38	78.68	6.93	98.68
Auxiliary examination analysis	0.00	100.00	49.56	18.39	36.87
Case analysis	0.00	98.33	61.79	14.85	56.42
Total score	23.00	92.00	75.51	6.12	98.68

Results of the difference test

Table 4 Descriptive statistics and ANOVA results of different genders in various modules and total scores

Variables	Gender	Mean	SD	t	df	*p*
Medical history collection	Male	77.30	8.85	11.77	3708	< 0.001***
	Female	80.49	7.62			
Physical examination	Male	75.24	9.94	5.17	3256	< 0.001***
	Female	76.99	9.37			
Basic operational skill	Male	77.37	7.34	11.97	3708	< 0.001***
	Female	80.05	6.18			
Auxiliary examination analysis	Male	49.11	18.39	1.38	3041	0.169
	Female	50.03	18.38			
Case analysis	Male	60.76	15.15	3.90	3041	< 0.001***
	Female	62.86	14.45			
Total score	Male	74.50	6.40	11.81	3256	< 0.001***
	Female	76.98	5.54			

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

case analysis (56.42%), and auxiliary examination analysis (36.87%). These results highlight the students' inadequate mastery of cardiac auscultation, lung auscultation, and auxiliary examination analysis, indicating a weaker case analysis capability and a need to further strengthen clinical thinking skills.

Descriptive statistics of different module grades and total scores.

Regarding the practical skills examination of the National Clinical Practitioner Qualification Examination, the items were categorized according to five modules: history taking, physical examination, basic operation skills, auxiliary examination analysis and case analysis. The statistical analysis of the average scores for each module in Table 3 revealed that the highest average score was in medical history collection (78.85 ± 8.43), followed by basic operational skill (78.68 ± 6.93), physical examination (76.09 ± 9.7), and case analysis (61.79 ± 14.85). The lowest average score was observed in the auxiliary examination analysis (49.56 ± 18.39). The highest pass rate was observed in basic operational skills (98.68%), whereas the pass rate for auxiliary examination analysis was significantly lower, at only 36.87%.

An independent samples t-test was conducted to evaluate the differences in scores across various modules, and the total score between male and female students revealed significant differences (all $p < 0.001$) except for auxiliary examination analysis. Female students scored significantly higher than male students on all modules and overall scores except for auxiliary examination analysis (see Table 4; Fig. 1).

According to Table 5; Fig. 2, analysis of variance (ANOVA) of students from different majors in each module as well as in total scores revealed that specialties differed significantly in physical examination ($F(2,3255) = 5.75$, $p = 0.003$), basic operational skill ($F(2,3707) = 7.43$, $p < 0.001$), auxiliary examination analysis ($F(2,3040) = 3.77$, $p = 0.02$) and case analysis($F(2,3040) = 3.32$, $p = 0.04$). Specifically, physical examination scores were significantly higher in clinical medicine than in anesthesiology ($p = 0.01$); however, the differences between clinical medicine and medical imaging and between anesthesiology and medical imaging were not significant ($p > 0.05$). For basic operational skills, anesthesiology scored significantly lower than clinical medicine ($p = 0.01$) and medical imaging ($p < 0.001$). Besides, students majoring in clinical medicine scored significantly higher than anesthesiology on the auxiliary examination ($p = 0.02$) and case analysis ($p = 0.04$).

Network analysis of raw OSCE items

Figure 3 depicts the network structure and expected influence (EI) of nodes for OSCE. Of the 55 edges, 44 (80%) have non-zero weights, indicating a relatively dense OSCE network. As illustrated, the nodes "Pediatric Skills" and "Nursing Skill" share the most substantial

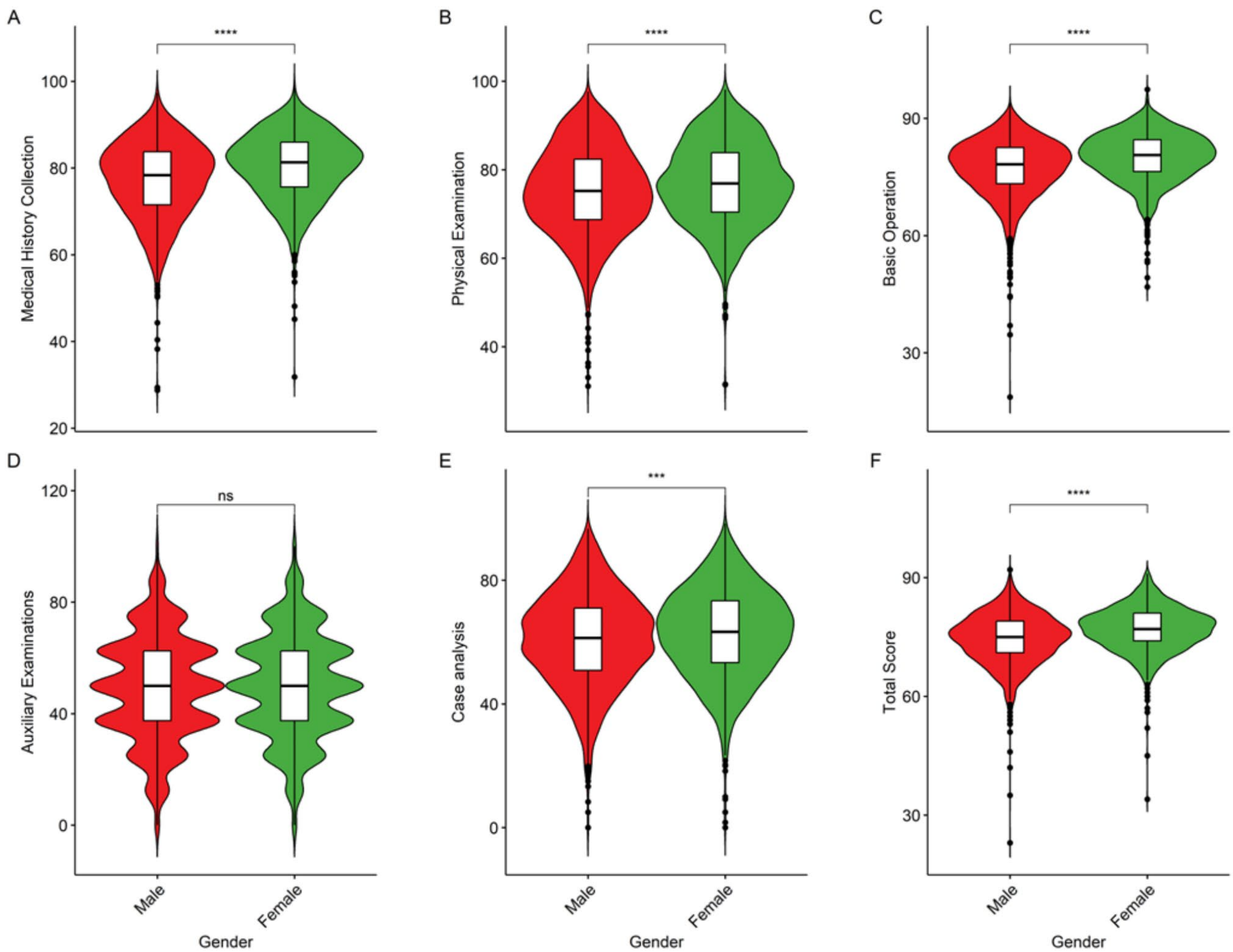

Fig. 1 Violin figures of different genders in various modules and total scores. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 5 Descriptive statistics and ANOVA results of different majors in various modules and total scores

Variables	Major	Mean	SD	df	F	*p*	Pairwise comparison
Medical history collection	Clinical medicine	78.95	8.50	3707	0.82	0.44	
	Anesthesiology	78.44	8.42				
	Medical imaging	78.63	8.02				
Physical examination	Clinical medicine	76.41	9.72	3255	5.75	0.003**	1 > 2: $p = 0.01$*
	Anesthesiology	74.70	10.15				
	Medical imaging	75.40	9.22				
Basic operational skills	Clinical medicine	78.73	6.95	3707	7.43	< 0.001***	1 > 3: $p = 0.01$* 2 > 3: $p < 0.001$***
	Anesthesiology	79.54	6.91				
	Medical imaging	77.79	6.73				
Auxiliary examination analysis	Clinical medicine	49.95	18.40	3040	3.77	0.02*	1 > 2: $p = 0.02$*
	Anesthesiology	46.91	18.02				
	Medical imaging	49.39	18.47				
Case analysis	Clinical medicine	61.98	14.87	3040	3.32	0.04*	1 > 2: $p = 0.04$*
	Anesthesiology	59.75	15.26				
	Medical imaging	62.24	14.37				
Total score	Clinical medicine	75.83	6.24	3255	1.68	0.19	
	Anesthesiology	75.45	5.94				
	Medical imaging	75.32	5.57				

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

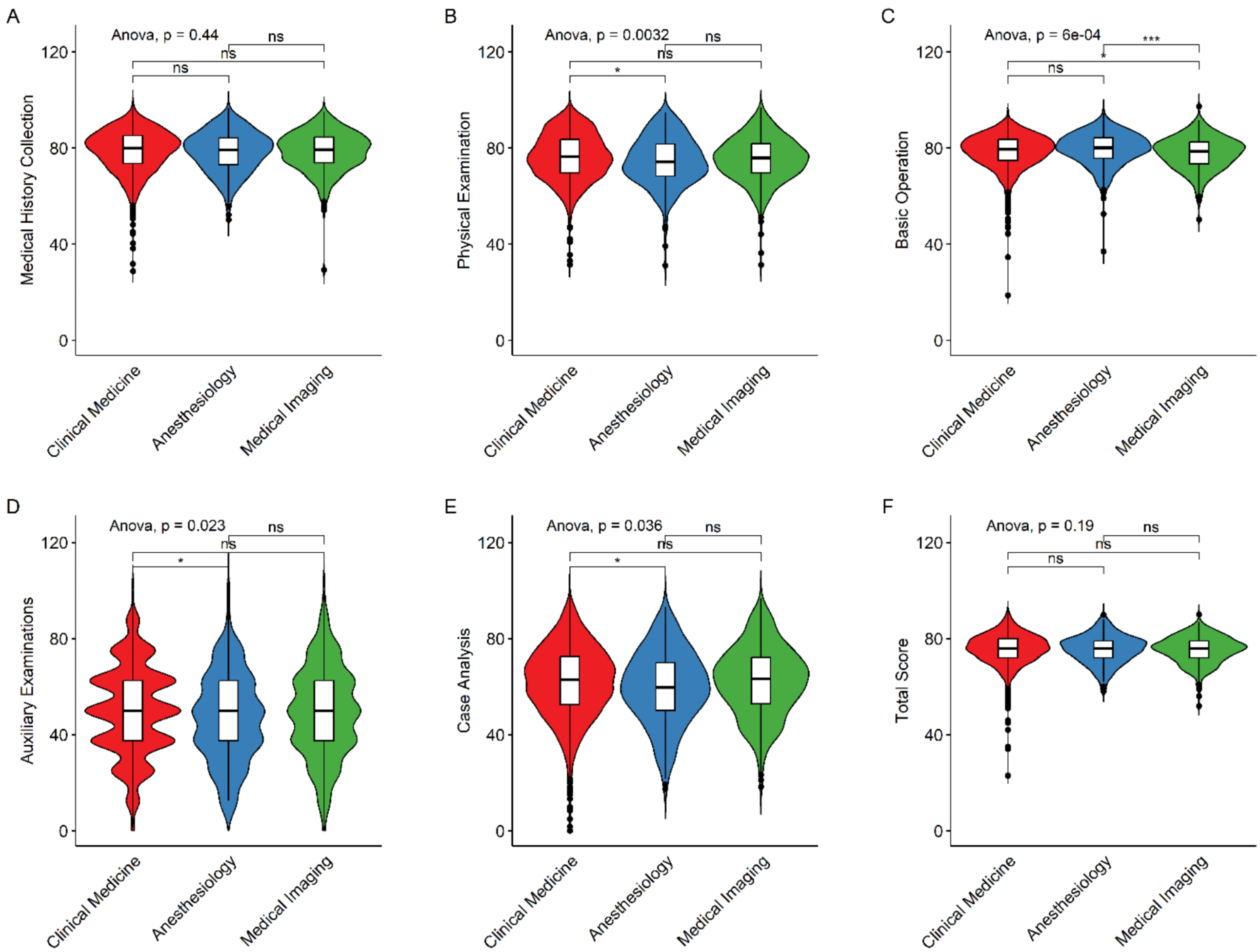

Fig. 2 Violin figures of different majors in various modules and total scores. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

connection, characterized by a weight of 0.19, indicating a notably strong relationship. Following closely is the connection between "Cardiac Auscultation" and "Pulmonary Auscultation" with a weight of 0.15, signifying a noteworthy association (see Table A.1). Bootstrapped difference tests on edge weights can be seen in Figure A.1. In terms of node predictability, approximately 11.63% of the variance can be explained by the influence of neighboring nodes. "General Physical Examination" ($R^2 = 0.20$) demonstrates the highest predictability, followed by "Pediatric Skills" ($R^2 = 0.19$). Predictability values of all nodes refer to Table A.2.

Figure 4 displays normalized measures of the centrality index, including strength, closeness, betweenness, and Expected Influence (EI) within the OSCE network. Among these, "General Physical Examination" stands out as the most influential node (EI = 0.79), followed by "Pediatric Skills" (EI = 0.65) and "Gynecological Obstetrical Skills" (EI = 0.63). Specific values are shown in Table A.3. Bootstrapped difference tests on node EI reveal significant differences for central symptoms compared to the majority of nodes, indicating robust results (as seen in Figure A.2). Figure 5 shows the estimate of network stability and accuracy for the 11 OSCE items. The 95% confidence intervals for edge weights overlap significantly. Specifically, the CS (cor = 0.7) coefficient of expected influence is 0.75 in the subsample bootstrap test, indicating that 75% of participants could be excluded from the analysis without significantly changing the network structure (see Fig. 6). Generally, the CS coefficient should not be lower than 0.25 and preferably higher than 0.5 for stable estimation of the centroid metric.

Discussion

While OSCE has been widely implemented as a primary assessment methodology in medical education, empirical investigations into the interrelationships among examination components remain notably limited in the literature. This study is the first to use network analysis to investigate these relationships within the OSCE, aiming to identify the most influential nodes and connections.

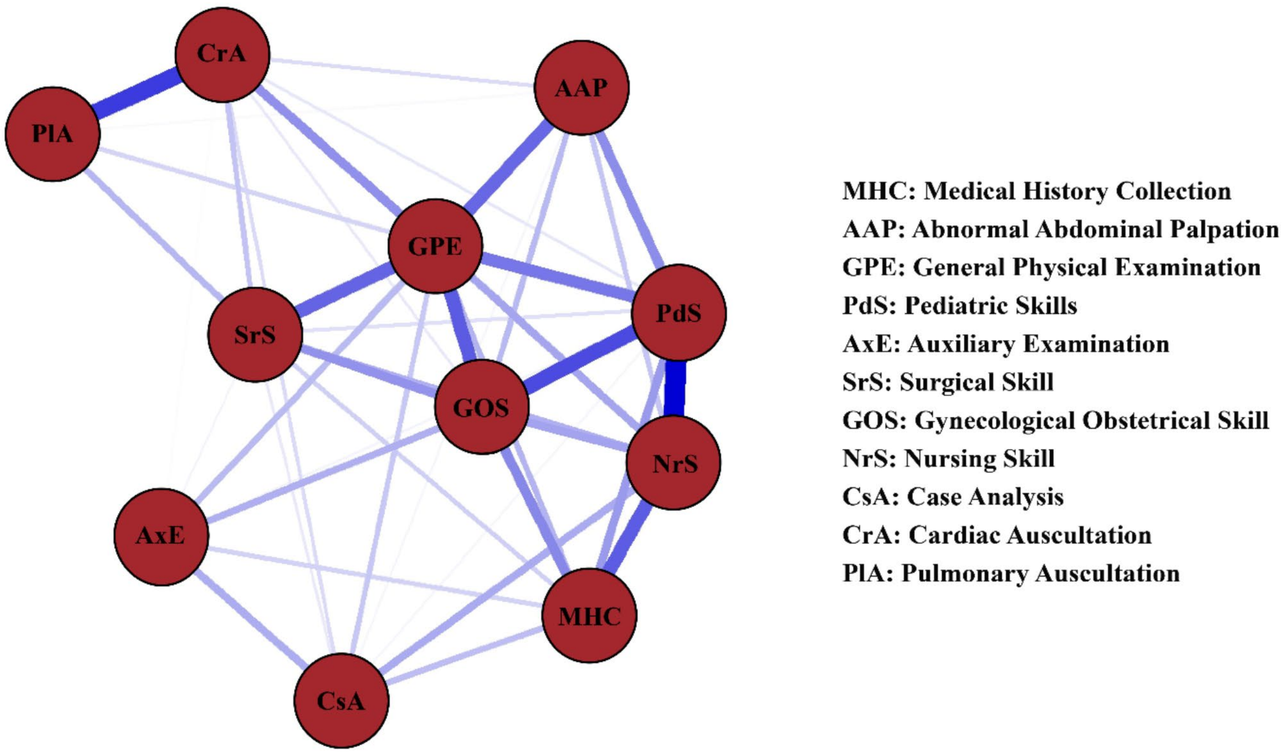

Fig. 3 Registered at OpenNetwork structure of OSCE in clinical medicine students of expected influence for each node. Edge thickness reflects relational strength; thicker edges denote stronger correlations. Red edges denote negative correlations, while blue edges signify positive correlations. "Expected Influence (EI)" is a metric employed in network analysis to measure the anticipated impact of a node within the network. A higher expected influence indicates that the node exerts a more significant influence over the entire network

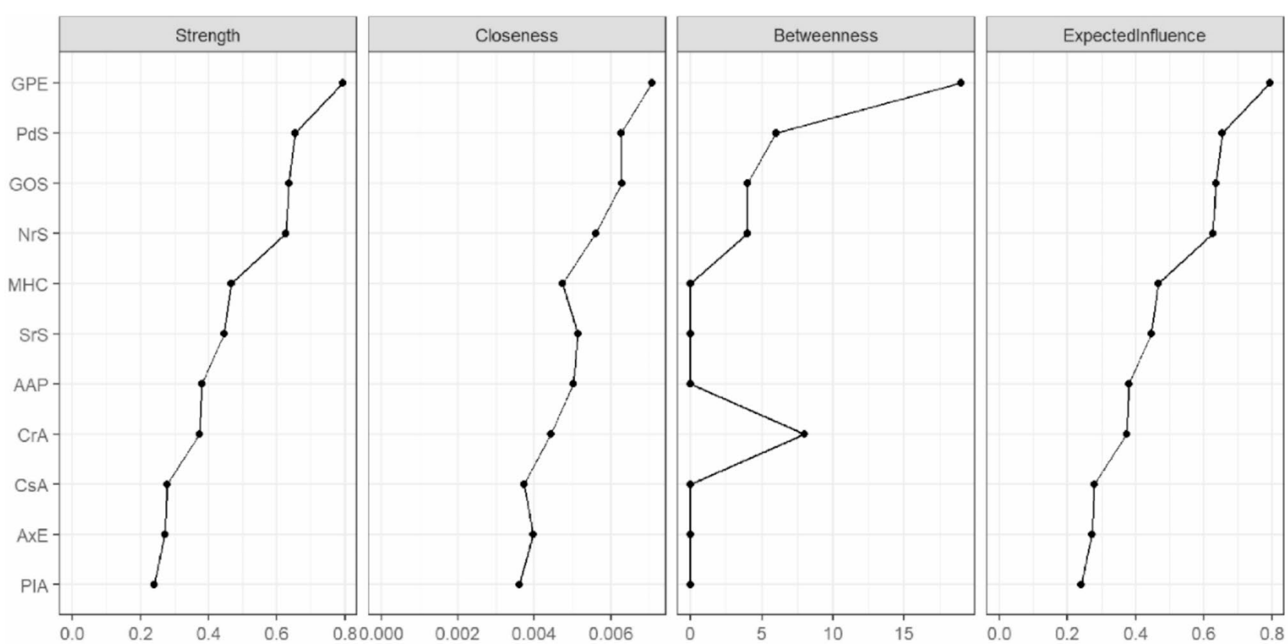

Fig. 4 The centrality index for each node. "Expected Influence (EI)" is a metric employed in network analysis to measure the anticipated impact of a node within the network. A higher EI value indicates that the node exerts a more significant influence over the entire network

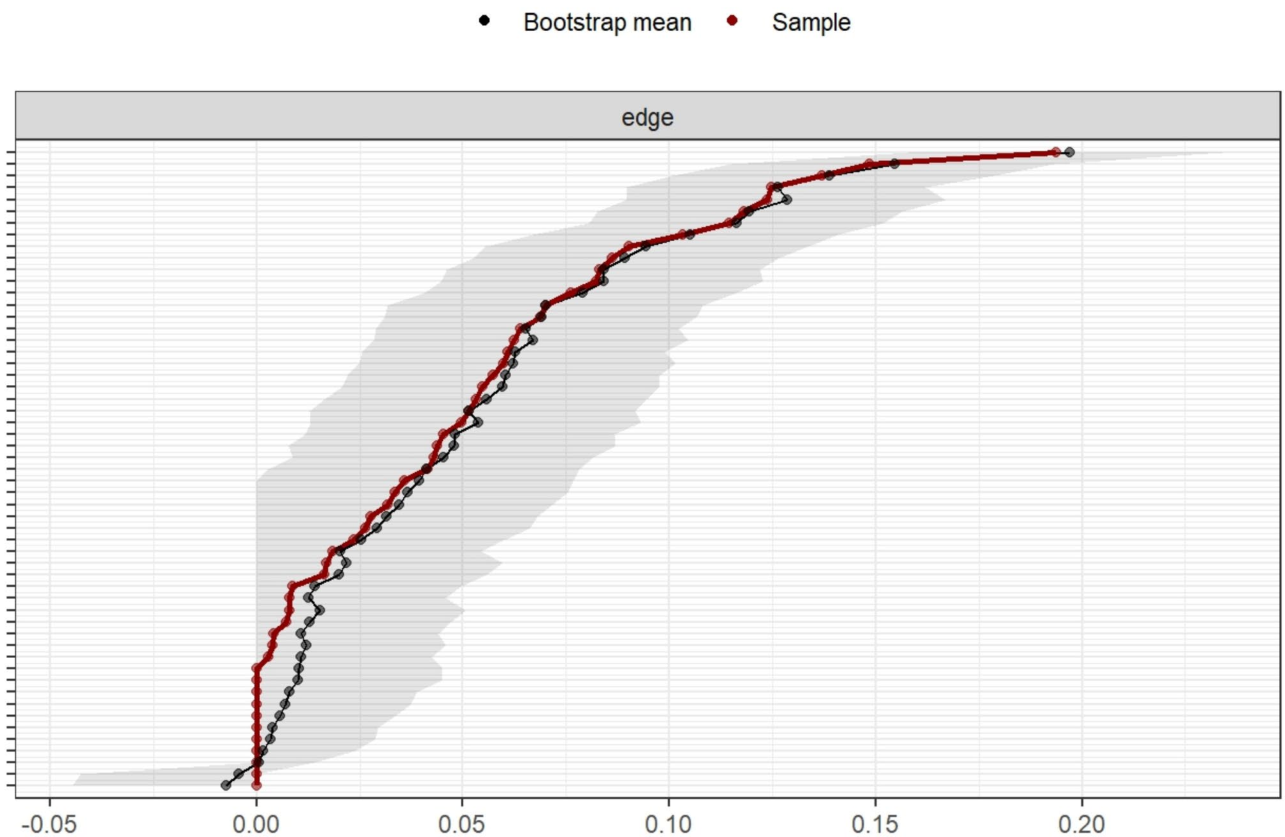

Fig. 5 The 95% confidence interval of the network edge weight of the OSCE items

The analysis revealed significant performance differences across various OSCE examination modules. History taking emerged as the module with the highest average score, suggesting that students are generally more proficient in this area. Following closely were basic operation and physical examination, both of which also received relatively high average scores. These results indicate that students are adequately prepared for these fundamental clinical skills. The strong performance in basic operation is further reflected in the highest pass rate observed for this module, suggesting a high level of competence and confidence among students in executing basic clinical procedures. In contrast, case analysis and auxiliary examination had lower average scores, with auxiliary examination having the lowest. The relatively low pass rates for the auxiliary examination and case analysis sections of the OSCE can be primarily attributed to the complexity of these tasks and the high demand for clinical reasoning skills. Auxiliary examination requires students to master a range of clinical skills and accurately interpret examination results in the context of the patient's specific condition. Case analysis, on the other hand, demands that students synthesize medical knowledge and clinical thinking to conduct in-depth analysis of complex clinical scenarios and formulate appropriate treatment plans. The complexity of these tasks leads to higher demands on students' time management skills. Moreover, as more flexible items, auxiliary examination and case analysis greatly test clinical reasoning abilities. Similar studies have shown that students' clinical reasoning skills, communication skills, and case assessment are indeed the weak points in the OSCE assessment literacy [10–13]. This observation is supported by the study of Schuwirth and Van der Vleuten [14], which pointed out the complexities involved in mastering auxiliary examination techniques and the need for more effective teaching methods. Overall, the results suggest that while students excel in history taking and basic operations, the identified deficiencies in auxiliary examination and case analysis competencies necessitate targeted interventions to enhance comprehensive clinical proficiency. It is essential to cultivate medical students' clinical reasoning and patient communication skills through clinical teaching by physicians and the establishment of more clinical internship courses, allowing students to experience and participate in real hospital work scenarios such as rounds, examinations, handovers, and discussions [11, 15].

Our analysis revealed significant gender disparities in OSCE performance, with female students demonstrating

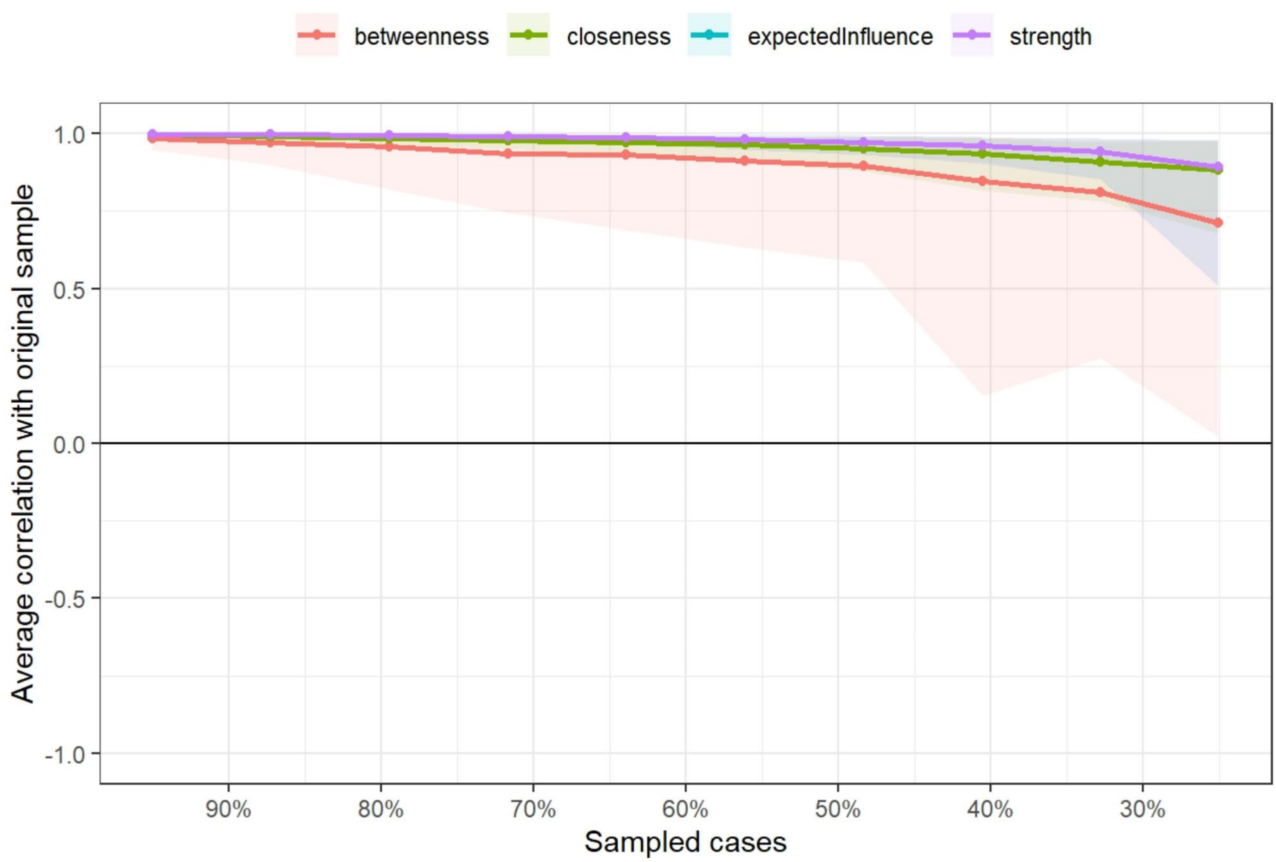

Fig. 6 Subsample bootstrap results of OSCE network

superior achievement across most modules and overall scores, except for auxiliary examination. These findings align with previous research [16–18], demonstrating female medical students' consistently higher performance in clinical assessments, particularly in communication skills and patient interaction. The observed gender differences likely reflect not only general academic achievement but also specific clinical competencies where female students excel. Notably, Graf's research across international medical schools [19] indicates that such gender-based performance differences transcend institutional contexts, suggesting broader implications for medical education. Male students' relative underperformance, particularly in auxiliary examinations and clinical reasoning, indicates a need for targeted educational interventions, including supplementary clinical reasoning workshops and structured mentorship programs. Meanwhile, in the course development, the tutorial system can be carried out to let excellent students serve as model teachers and carry out flipped classroom to help the weak students in this part. Also, structured mentorship programs pairing students with experienced clinicians to provide guidance and feedback on clinical decision-making processes.

Analysis of variance demonstrated significant major-specific differentials in OSCE performance, with clinical medicine students achieving superior scores in physical examination, basic operational skills, auxiliary examination, and case analysis compared to their anesthesiology counterparts. This performance disparity aligns with findings by Norman [20] and Bland et al. [21], who suggested that clinical medicine curricula place more emphasis on foundational medical skills than other specialities. The observed underperformance of anesthesiology students in basic operational skills, despite their advanced specialty-specific training, suggests the need for enhanced instruction in fundamental clinical competencies while maintaining their specialized technical focus. Further curriculum adjustments could include: Integrating more comprehensive basic skills training into the anesthesiology curriculum, possibly through dedicated modules or workshops. Developing interdisciplinary training programs that allow anesthesiology students to collaborate with clinical medicine students on certain procedures, fostering skill exchange and mutual learning. Incorporating more case-based learning scenarios that require anesthesiology students to apply basic operational skills in the context of anesthesia care, bridging the gap between specialized knowledge and general clinical competencies.

Notably, the observed gender and major-specific performance differentials may be modulated by local cultural, educational, and institutional factors, and variations in curricular emphasis likely influence discipline-specific outcomes. In Chinese culture, there may be differences in expectations of male and female roles, which may affect students' learning styles and performance. Female may be seen as better suited for communication and detailed work, while male may be seen as better suited for technical or analytical tasks [22]. This stereotype of gender roles may subtly affect students' learning focus and self-confidence. While undergraduate medical education structures differ globally [23], comprehensive multi-institutional and cross-cultural studies are needed to validate these findings and assess the broader applicability of proposed interventions across diverse educational contexts. Our study thus contributes not through systemic comparisons but by offering empirically derived insights from localized medical education reforms.

Regarding the results of the network analyses, the strong link between "Pediatric Skills" and "Nursing Skills" highlights their interdependence in clinical practice. This connection likely indicates the comprehensive nature of pediatric care, where nursing competencies are crucial for effectively addressing the needs of young patients. This significant relationship is supported by Nickel et al. [24], who emphasized the essential role of comprehensive nursing skills in providing quality pediatric care. Similarly, the strong association between "Cardiac Auscultation" and "Pulmonary Auscultation" reflects their complementary roles in assessing cardiopulmonary function. Accurate auscultation skills are essential for diagnosing and managing heart and lung conditions. This interrelation is supported by contemporary studies, such as those by Mangione et al. [25], highlighting the importance of integrated cardiopulmonary examination skills in medical education. These findings suggest that integrating practical scenarios that combine these skills may better prepare students for real-world clinical settings, ultimately improving patient care outcomes.

Furthermore, the prominence of the "General Physical Examination" as the most influential node suggests its central role in the OSCE, likely due to its foundational importance in clinical practice since physical examination plays a critical role in medical training [20]. "Pediatric Skills" and "Gynecological Obstetrical Skills" also show high influence, reflecting the necessity of specialized skills in these areas. This supports the notion that proficiency in these specific skills is crucial for comprehensive clinical competence, as highlighted by Kaufman [26]. These results indicate that strengthening training in these key areas could enhance overall clinical competence and performance in the OSCE.

The apparent discrepancy between the low pass rates in cardiac and pulmonary auscultation stations versus the high pass rate in general physical examination stations, despite their anticipated correlation, can be explained by differences in task complexity and the specific skill sets required for each type of assessment. Auscultation requires advanced technical competencies in interpreting subtle auditory cues, whereas general physical examinations primarily involve more readily mastered visual and tactile skills. This complexity differential is supported by research demonstrating persistent challenges in auscultation skill acquisition, even with high-fidelity simulation training [27], and documented difficulties in mastering auditory diagnostic cues for cardiac and pulmonary conditions [28]. These inherent differences in skill complexity and acquisition trajectories may explain the weak correlations observed between these stations in network analysis.

Several study limitations warrant consideration. First, our single-institution design may limit generalizability due to institution-specific instructional approaches. For example, another institution might place greater emphasis on communication skills, potentially leading to different performance outcomes. Additionally, we should fully account for variations in student selection criteria (different admission processes or prior educational backgrounds). For instance, some institutions may admit students with more clinical exposure, which could impact their OSCE readiness. Second, while network analysis revealed OSCE module relationships, this methodology may not fully capture the complexity of clinical skill interdependencies. Finally, our study did not account for potential confounding variables like student background factors and learning preferences that could influence OSCE performance. In terms of student background factors, family economic conditions may affect students' ability to obtain additional learning resources, while parents' professional background (such as whether they are medical practitioners) may have a potential impact on students' medical interest and learning motivation. In terms of learning preference, visual learners may perform better in the image diagnosis module, while students with strong hands-on ability may predominate in the skill operation module. In addition, students who actively acquired internships or internships, spent more time in clinical exercises, or had more patient contact were likely to perform better in the history acquisition and physical examination modules. Therefore, future multi-institution studies should add these potential confounders to demographic data collection and baseline control.

Conclusion

Based on the network analysis findings, medical institutions should emphasize the "General Physical Examination" as a foundational component in OSCE curricula, given its demonstrated centrality in skill interconnections. The significant correlations observed between "Pediatric Skills" and "Nursing Skills," as well as between "Cardiac Auscultation" and "Pulmonary Auscultation," necessitate the development of integrated teaching modules that capitalize on these skill synergies.

Furthermore, addressing identified gender-based performance disparities and specialty-specific variations requires the implementation of targeted interventions, specifically advanced skill development programs for female students and specialized communication and pediatric skill workshops for male students, complemented by auxiliary examination training for anesthesiology students. Enhancement of clinical reasoning and patient communication competencies, critical for both OSCE performance and clinical practice, demands an expansion of experiential learning opportunities through structured patient interactions, ward rounds, and case discussions under expert clinical supervision. These evidence-based curricular modifications aim to foster comprehensive clinical competence development and optimize patient care outcomes in contemporary medical practice.

List of abbreviations

OSCE	Objective Structured Clinical Examination
NCPQE	National Clinical Practitioner Qualification Examination
ANOVA	Analysis of Variance
EI	Expected Influence
CS	Correlation Stability

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-025-07091-0.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Author contributions

H.Q.Z. supervision, writing reviews and editing, funding acquisition, and resources. S.S.L. data curation, writing-original draft preparation. G.Q.Z. data curation. X.Y.C. data curation. L.J.Z. writing review and editing. J.Y.L. software, data curation. Y.L. methodology, software, investigation, writing-review and editing, funding acquisition. All authors reviewed the manuscript.

Funding

This work was supported by the 2023 Ministry of Education Industry-Education Cooperation Collaborative Education Project [grant numbers 230806052074117], the 2023 Guangdong Basic and Applied Basic Research Foundation [grant numbers 2023A1515110350], the 2024 Youth Project of Guangdong Philosophy and Social Sciences Planning Project [grant numbers GD24YXL06]; the Plan on enhancing scientific research in GMU [grant numbers 02-410-2302338XM]; 2021 Guangdong Provincial Higher Education Teaching Reform Project of Undergraduate Colleges and Universities "Research on the Reform of Ability-oriented Clinical Internship Exit Assessment and Evaluation" [grant numbers 462]; Guangzhou Educational Science Planning 2021 Project "Practical Research on Online and Offline Blended Teaching in Clinical Skills" [grant numbers 202113657].

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the ethical standards of the Medical Ethics Committee of Guangzhou Medical University (Ref No.: 202404010) and with the principles outlined in the Declaration of Helsinki. Additionally, informed consent was obtained from all participants involved in the study prior to their involvement.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 22 August 2024 / Accepted: 1 April 2025
Published online: 24 April 2025

References

1. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. BMJ. 1975;1:447.
2. Barnes KN, Hardinger KL, Graham MR. Review of objective structured clinical examination practices within pharmacy programs throughout the united States. Am J Pharm Educ. 2024;88:100686.
3. Lam T, Wan X, Ip MS. Current perspectives on medical education in China. Med Educ. 2006;40:940–9.
4. Comparative Study on the Education System of Traditional. Medicine in China, Japan, Korea, and Taiwan. EXPLORE. 2016;12:375–83.
5. Wang W. Medical education in China: progress in the past 70 years and a vision for the future. BMC Med Educ. 2021;21:453.
6. Yichi Zhang. How It's Made — Doctor's Edition: Comparing American and Chinese Medical Education. In: in-Training. 2020.
7. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behav Res. 2018;50:195–212.
8. McNally RJ. Networks and nosology in posttraumatic stress disorder. JAMA Psychiatry. 2017;74:124–5.
9. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths. Social Networks. 2010;32:245–51.
10. Al-Hashimi K, Said UN, Khan TN, Al-Hashimi K, Said UN, Khan TN. Formative objective structured clinical examinations (OSCEs) as an assessment tool in UK undergraduate medical education: A review of its utility. Cureus. 2023;15.
11. Sim JH, Aziz YFA, 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.
12. Wingelaar TT, Wagter JM, Arnold AER. Students' educational needs for clinical reasoning in first clerkships. Perspect Med Educ. 2012;1:56–66.
13. Amin Z, Seng CY, Eng KH. Practical guide to medical student assessment. WORLD SCIENTIFIC; 2006.
14. Schuwirth LWT, Vleuten CPMVD. The use of clinical simulations in assessment. Med Educ. 2003;37:65–71.
15. Raziani Y, Nazari A, Raziani S. Nursing students' experiences with the objective structured clinical examination (OSCE): A qualitative study. Int J Afr Nurs Sci. 2022;16:100433.
16. Alkhateeb NE, Al-Dabbagh A, Ibrahim M, Al-Tawil NG. Effect of a formative objective structured clinical examination on the clinical performance of undergraduate medical students in a summative examination: A randomized controlled trial. Indian Pediatr. 2019;56:745–8.
17. Riese A, Rappaport L, Alverson B, Park S, Rockney RM. Clinical performance evaluations of Third-Year medical students and association with student and evaluator gender. Acad Med. 2017;92:835.

18. Schleicher I, Leitner K, Juenger J, Moeltner A, Ruesseler M, Bender B, et al. Examiner effect on the objective structured clinical exam – a study at five medical schools. BMC Med Educ. 2017;17:71.
19. Graf J, Smolka R, Simoes E, Zipfel S, Junne F, Holderried F, et al. Communication skills of medical students during the OSCE: Gender-specific differences in a longitudinal trend study. BMC Med Educ. 2017;17:75.
20. Norman G. Research in clinical reasoning: past history and current trends. Med Educ. 2005;39:418–27.
21. Bland CJ, Meurer LN, Maldonado G. Determinants of primary care specialty choice: a non-statistical meta-analysis of the literature. Acad Med. 1995;70:620.
22. Wen D, Ma X, Li H, Liu Z, Xian B, Liu Y. Empathy in Chinese medical students: psychometric characteristics and differences by gender and year of medical education. BMC Med Educ. 2013;13:130.
23. Van Zanten M, Norcini JJ, Boulet JR, Simon F. Overview of accreditation of undergraduate medical education programmes worldwide. Med Educ. 2008;42:930–7.
24. Nickel B, Gorski L, Kleidon T, Kyes A, DeVries M, Keogh S, et al. Infusion therapy standards of practice. J Infusion Nursing: Official Publication Infusion Nurses Soc. 2024;44:S1–224.
25. Mangione S, Sullivan P, Wagner MS. Physical diagnosis secrets: physical diagnosis secrets E-Book. Elsevier Health Sciences; 2020.
26. Kaufman DM. Teaching and Learning in Medical Education: how theory can inform practice. In: Understanding Medical Education: Evidence, Theory, and Practice. 3rd edition. 2018. pp. 37–69.
27. Jewett L, Clarke S, Griffin E, Danielson A. Core and cluster or head to toe? A comparison of two types of curricula for teaching physical examination skills to preclinical medical students. BMC Med Educ. 2024;24:337.
28. Osborne C, Brown CW, Mostafa A. Effectiveness of high and low-fidelity simulation-based medical education in teaching cardiac auscultation: a systematic review and meta-analysis. Int J Healthc Simul. 2022;1:75–84.

Publisher's note