



OPEN ACCESS

EDITED BY

Predrag Jovanovic,
University Clinical Center Tuzla, Bosnia
and Herzegovina

REVIEWED BY

Enver Zerem,
Academy of Sciences and Arts of Bosnia
and Herzegovina, Bosnia and Herzegovina
Ziyi Yan,
Sichuan University, China

*CORRESPONDENCE

Yun Li
✉ liyun@gzhmu.edu.cn

†These authors have contributed equally to
this work and share first authorship

RECEIVED 15 May 2025

ACCEPTED 22 September 2025

PUBLISHED 27 October 2025

CITATION

Zhang H, Li J, Zhong L, Zheng J, Chen X,
Wen S, Li Y and Li J (2025) Exploring
the localization effects of organ
system-based curriculum: a comparative
study of different teaching programs
in medical education.
Front. Educ. 10:1629192.
doi: 10.3389/feduc.2025.1629192

COPYRIGHT

© 2025 Zhang, Li, Zhong, Zheng, Chen, Wen,
Li and Li. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Exploring the localization effects of organ system-based curriculum: a comparative study of different teaching programs in medical education

Huiqun Zhang^{1†}, Jiaying Li^{2†}, Lujie Zhong², Jianmin Zheng³,
Xiaoyun Chen¹, Shenyu Wen⁴, Yun Li^{2,5*} and Jianhua Li³

¹Practical Teaching Center, Guangzhou Medical University, Guangzhou, China, ²School of Health Management, Guangzhou Medical University, Guangzhou, China, ³Academic Affairs Office, Guangzhou Medical University, Guangzhou, China, ⁴Nanshan School, Guangzhou Medical University, Guangzhou, China, ⁵Guangdong-Hong Kong-Macao Greater Bay Area Medical and Health Industry High Quality Development Rule of Law Guarantee Research Center, Guangzhou, China

Background: Organ system-based curriculum (OSBC) has been introduced to improve traditional discipline-based teaching in Chinese medical schools. This study evaluates the localization effects of OSBC by comparing short term, i.e., final exam, objective structured clinical examination (OSCE) and long term i.e., national medical licensing examination (NMLE) outcomes across different teaching classes. Findings aim to inform future curriculum reform and faculty development in localized medical education.

Methods: This study employed a quasi-experimental design and obtained 111 undergraduate medical students through cluster sampling. Medical students were grouped by teaching schemes, gender and grade with short-term outcomes assessed via final exams and OSCE, and long-term outcomes via post-graduation NMLE scores.

Results: The results of Wilcoxon rank sum test showed that the OSCE scores of students in Nanshan class were higher than those in traditional teaching class, especially in the modules of Medical History Collection and Physical Examination ($Z = 1.979$, $p = 0.048$; $Z = 2.405$, $p = 0.016$). Yet, the comparison of NMLE items revealed no significant differences between the two student groups. Males and earlier cohorts exhibited slight advantages in OSCE/NMLE, though non-significant.

Conclusion: The localized OSBC demonstrates early prospects in clinical skills highlighting the need for optimized resource allocation personalized guidance and holistic student development in medical education reform, offering insights for the exchange and practice of international medical education.

KEYWORDS

organ-system-based curriculum, medical education, clinical skills training, OSCE, PBL

1 Introduction

1.1 Challenges in Chinese conventional medical education model

Chinese conventional medical education model has long followed the Soviet three-stage model (“general education - clinical knowledge education - clinical skills training”) and its “8 + 2” teaching mode (Zhou, 2013). In recent years, in order to cope with the impact and challenge of higher education development, colleges are not only facing the situation of expanding the number of higher education places, optimizing the scale and resources of colleges and universities, but also taking on the responsibility of avoiding mediocrity in higher education and providing high-level, high-quality innovative talents for a diverse society (O’Connell and Pascoe, 2004; Yang, 2013; Smithson et al., 2020). Medical universities need to respond to the demands of national scientific and technological progress and economic growth, and strive to cultivate high-level innovative talents and promote the reform of clinical medical education (Xiao et al., 2011, 2022; Yang et al., 2012; Yan, 2023). However, this traditional, teacher-centered approach often results in a disconnect between students’ basic knowledge and clinical skills, passive learning, and a low level of critical thinking (An, 2022; Stan et al., 2022; Yan et al., 2024; Song et al., 2017; Eleftheriou et al., 2022), significantly hindering to efficiently boosting the competency of medical professionals and overcome above challenges (Wijnen-Meijer, 2023).

1.2 Introduction and localization reform of OSBC

Based on international experience, the adoption of an Organ System-Based Curriculum (OSBC) may be a breakthrough of traditional Chinese medical education model. OSBC is an innovative medical education model that focuses medical education on the human organ system. This model was first proposed by Case Western Reserve University in the U.S. and was widely recognized by the international medical education community at the World Medical Education Summit in Edinburgh in 1993 (Wang et al., 2024). The OSBC model integrates the knowledge of basic and clinical medicine both horizontally and vertically, enabling students to develop a comprehensive and in-depth understanding of medical knowledge (Sakles et al., 2006). It guides them through a holistic learning process, starting from anatomy and physiology, progressing through pathology and epidemiological characteristics, and extending to clinical manifestations, diagnosis, and treatment of diseases.

However, given the cultural, educational levels, and talent reserves of different countries, the application of the OSBC solution is not a complete copy but requires localized reforms. Localization refers to the transfer, adaptation and development of relevant values, knowledge, technologies and behavioral norms from the local environment (Cheng, 2005). Localization can significantly enhance the relevance of education to local development in terms of educational reform. At the school level, localization is specifically reflected in adapting external economic, political, social and cultural measures to local resources, such

as cooperation with communities and hospitals, adjustment of curriculum structure and student quotas, to promote local cultural development, cooperation among educational institutions and enhance educational effectiveness (Welch and Mok, 2003).

Based on the application and reform of the bio-psycho-social medical model (Bolton, 2023) and OSBC, the five-year “Nanshan Class” project was launched, aiming to cultivating innovative talents who combine excellent clinicians with medical scientists. In implementing the OSBC localization reforms, a number of distinctive measures have been taken with the objective of rendering the teaching model more pertinent to the specific context of our school, while simultaneously differentiating it from the models employed by other medical schools. Firstly, we have implemented an unprecedented comprehensive curriculum integration, which not only encompasses the core curriculum of basic and clinical medicine but also incorporates the knowledge of the humanities, thereby establishing a multi-dimensional and interdisciplinary knowledge system. Secondly, the institution has meticulously devised 9 modularized curricula, each systematically organized around a specific organ system. This modular approach facilitates a holistic understanding of the subject matter, promoting deeper learning and knowledge retention. In the formation of teaching teams, we have established 13 interdisciplinary teams, integrating faculty from basic medical sciences, clinical practices, and humanities. Teachers from interdisciplinary teams should hold seminars every semester to discuss the implementation effects of courses, adjust teaching focuses based on students’ final performance, enrich teaching forms and evaluation methods, and update the design of teaching materials. This comprehensive integration ensures the coherence and depth of the educational content. Furthermore, we have adopted innovative pedagogical methods, including problem-based learning and case studies, which foster student engagement and nurture independent critical thinking and problem-solving abilities. The development of clinical skills is another pivotal aspect of our OSBC reform. By integrating essential clinical competencies, we have developed a comprehensive practical skills curriculum. This curriculum allows students to learn and practice in simulated clinical environments that closely mimic real-world settings, significantly enhancing their clinical operation and reasoning skills. Regarding teaching evaluation, we have implemented a multifaceted assessment system that combines both summative and formative evaluations. This system focuses not only on exam results but also on students’ ongoing performance and progress, enabling a more holistic appraisal of learning outcomes.

Despite the fact that this innovative medical education model has been in place for some time, there has been a paucity of comprehensive and rigorous evaluation of its educational outcomes. It is imperative that effective methods be employed to conduct a comparative analysis of the impact of traditional and innovative models on the enhancement of medical students’ theoretical knowledge and clinical practice skills.

1.3 Application of OSCE and NMLE in medical model reform effect evaluation

In order to verify the effectiveness of the localization of the OSBC model, Objective Structured Clinical Examination (OSCE)

and National Medical Licensing Examination (NMLE) were used as important indicators of the evaluation in this study.

OSCE has emerged as an alternative method for assessing clinical skills, gaining recognition in the academic and professional communities (Harden et al., 1975). OSCE comprehensively evaluates students' theoretical knowledge, practical skills, and analytical thinking by requiring them to go through a series of assessment stations simulating real diagnostic scenarios within a limited time. Due to its excellence, reliability, and flexibility, this method has been widely adopted, especially suitable for assessing various clinical skills such as communication skills, physical examination, and problem-solving abilities.

Notably, the requirements of NMLE in China consistent with the components of the OSCE, including medical history collection, cardiopulmonary auscultation, etc., covering areas such as internal medicine, surgery, obstetrics and gynecology, pediatrics, and emergency medicine, aiming to objectively assess whether candidates meet the required qualification standards (Yang, 2013). Medical graduates are eligible to take the NMLE exam 1 year after completing their undergraduate degree. Since its formal introduction in 1998, the NMLE has served as a pivotal instrument in assessing the professional knowledge and competencies medical professionals (Wang, 2021, 2022). It not only ensures that physicians possess the requisite clinical skills and foundational competencies but also functions as a critical measure of the educational standards of medical institutions. The exam covers fundamental and clinical knowledge, medical ethics, and legal regulations, thoroughly assessing candidates through both a written test and practical skills evaluation. Administered under the supervision of the Ministry of Health, the NMLE adheres to principles of fairness, scientific precision, and standardization, ensuring the reliability and validity of its outcomes.

In conclusion, in order to verify the effect of OSBC localization, this study compared the innovation and clinical literacy of medical students under the two educational models (Nanshan class and traditional clinical medicine class) after localization. Taking the theoretical scores before graduation and the practical scores of the OSCE module as the short-term indicators on campus, and the NMLE scores in the first year after graduation as the long-term indicators of the localization effect, the effectiveness of the OSBC teaching model in shaping the abilities of medical students is evaluated, and suggestions are provided for the localization and integration of medical education.

2 Materials and methods

2.1 Participants

Prior to the survey, the researcher had ensured that the participants were informed about the purpose and risks of the study and that the data were stored anonymously and confidentially. All participants voluntarily agreed to participate, and provided written informed consent. The study has obtained ethical approval from Guangzhou Medical University (No. 202305008).

A complete cluster sampling approach was adopted with intact teaching classes serving as the sampling units. According to the following inclusion and exclusion criteria, we included 51

undergraduate medical students (20 males and 31 females) from the Nanshan class, constituting the experimental group. Of these, 24 students were enrolled in grade 2015 and 27 in grade 2016. A control group of 60 undergraduate medical students from the traditional clinical medicine program (31 males and 29 females) was also included in the study. Of this group, 30 students were in grade 2015 and 30 were in grade 2016. Since Nanshan class is produced through strict selection after admission, its selection conditions need to include any of the following conditions:

- (1) Outstanding college entrance examination results: The students from Guangdong province should be in the top 15 of the medical students admitted by our university in the same year, and the students from other provinces should be in the first place of the medical students admitted by their province;
- (2) Excellent academic performance: the average grade point of required courses and restricted courses in the first semester is among the top 8% of the major in the grade;
- (3) Outstanding foreign language performance: entered class A of graded English teaching, and ranked top 8% in the general score of Common English course in the first semester among the students in the class.

Therefore, the students in Nanshan class are the group of excellent students with high comprehensive quality. In order to ensure the comparability of the quality of students between the control group and the control group, we sorted the students in the First Clinical Medical school according to their student numbers (student numbers were sorted according to the college entrance examination scores), and considered the top 30 students in each grade as the sample range of the control group. Participants in the Nanshan and traditional clinical medicine classes were required to meet the following screening criteria. Inclusion criteria: (1) Age ranging from 17 to 25 years old; (2) Complete 5 years of undergraduate medical study; (3) Participate in the final, OSCE and NMLE assessments; (4) Voluntarily participate in this survey. Exclusion criteria: (1) Having experiences such as changing majors or taking a leave of absence that interrupted on-campus learning; (2) Past or current severe psychological disorders or mental illnesses.

This study adopted a quasi-experimental design and used *DEFF* to evaluate the sample efficacy [$DEFF = 1 + (m - 1) \times ICC$] (Campbell et al., 2004). The experimental group and the control group each included one natural class (51 and 61 students respectively, $m = 56$). Limited by the actual class sample size, based on the intraclass correlation coefficient ($ICC = 0.03$) and the preset medium to large effect size Cohen's $d = 0.7$, slightly lower than the standard 0.8 but acceptable (Fritz et al., 2012), the power analysis was conducted by calculating the design effect ($DEFF = 2.65$). The results show that the current sample size can provide 70% statistical test power.

2.2 Research design

This study conducted at Guangzhou Medical University, China, from September 2015 to July 2022. The study compared the academic and clinical performance of undergraduate medical

students enrolled in a localized OSBC with those following a traditional discipline-based curriculum. The study duration covered the full 5 years undergraduate education and one-year post-graduation period, allowing both short-term and long-term outcomes to be assessed.

The control group adhered to the traditional clinical course teaching model, which sticks to the system-to-body, localized-to-body, and later gradual deeper forms of the disease. Taking the professional courses as an example, the control group's basic teaching is systematic anatomy, histology and embryology and other courses, followed by basic medical courses such as local anatomy, biochemistry and pathophysiology, and finally sub-specialties such as obstetrics and gynecology, internal medicine and so on.

Whereas the experimental group designed mandatory courses in nine major organ system disease modules, such as the skeletal muscle system, skin, blood immune system, cardiovascular system, respiratory system, digestive system, nervous system, endocrine system, and urinary reproductive system. The goal was to develop integration across several disciplines including basic, clinical, preventative, and humanities. Additionally, the experimental group fully implemented problem-based learning (PBL) as a guided deep learning model. Furthermore, the experimental group embraced the concept of "early clinical exposure, multi-clinical exposure, repeated clinical exposure," conducting a one-month internship each semester. In terms of structure, the experimental group added mentorship, dual-mentorship, and international training compared to the control group. The overall difference between the Nanshan class and the traditional clinical medicine class is briefly presented in [Figure 1](#). The specific training program of Nanshan class and the traditional clinical medicine class can be found in the [Supplementary material](#).

2.3 Observation indicators

2.3.1 Theoretical results of the final exam

The final theoretical examination for medical students before graduation is a testing procedure to assess their mastery of core medical theories during their undergraduate studies. In terms of overall structure, this exam refers to the module form of NMLE and can be regarded as a mock exam of NMLE. However, the specific content of each examination area is not as comprehensive as that of NMLE. The setting of examination content is based on a comprehensive consideration of the high degree of learning difficulties reflected by students during their undergraduate studies, the low pass rate, and the importance of the subject in NMLE. This also reflects that the assessment is more targeted among the students in school. The specific areas, contents and proportions of the final exam are shown in [Table 1](#).

The test consists entirely of objective questions, utilizing formats that are designed to evaluate both theoretical knowledge and practical reasoning. These formats include:

1. A1-Type Question: the question is based on a narrative single sentence, examines basic knowledge, and selects one of five options.

2. A2-Type Question: it involves brief clinical scenarios followed by multiple-choice answers, with only one correct response of five choices.
3. A3-Type Question: the structure begins with a description of a clinical situation, followed by two or three questions, each related to the initial case but targeting distinct concepts.
4. B1-Type Question: the question starts with five alternatives, followed by at least two questions, with the test taker required to select the most relevant answer for each. Each alternative may be chosen once, multiple times, or not.

The exam is conducted as a closed-book assessment, lasting 120 min, with a total score of 100 points. It effectively prepares students for the NMLE by simulating its structure, providing an accurate measure of their readiness to embark on a professional medical career.

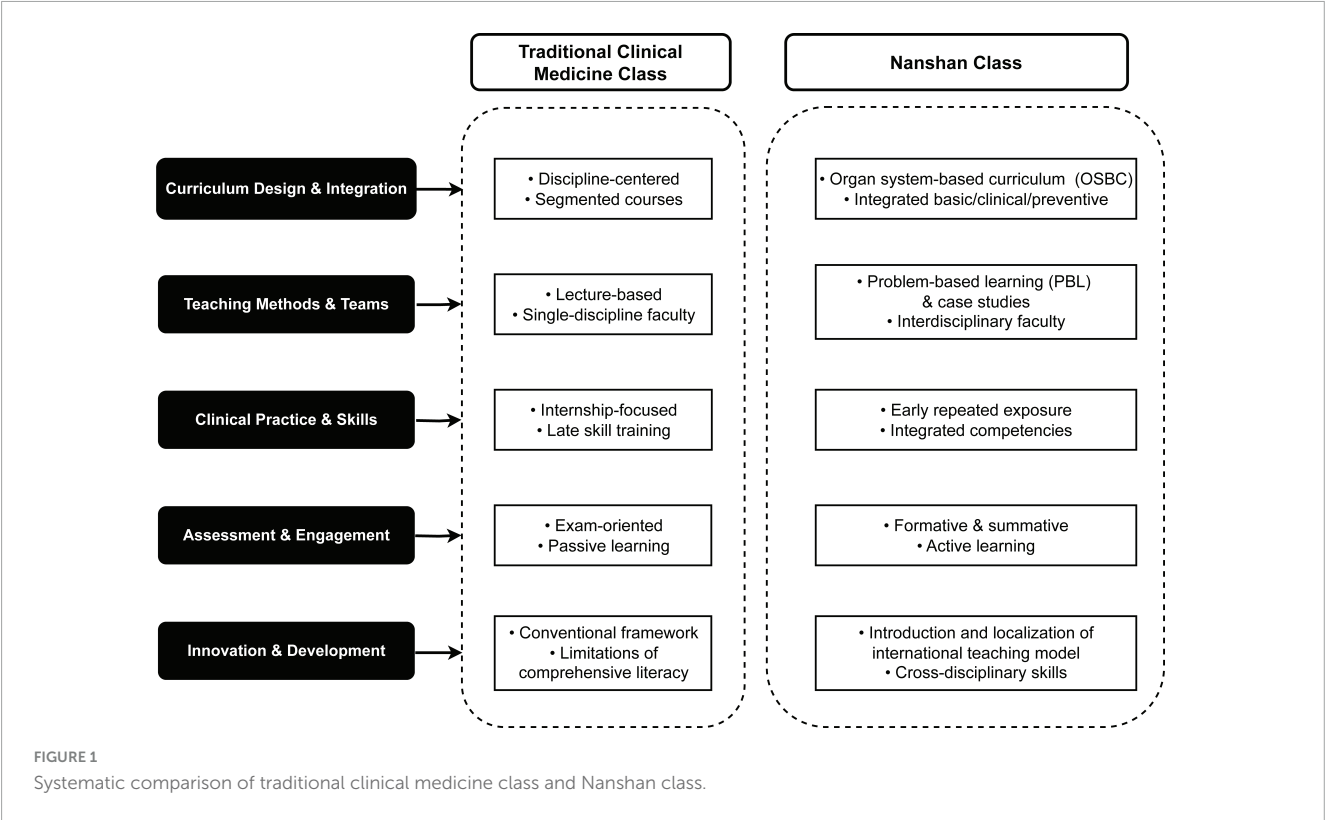
2.3.2 Objective structured clinical examination (OSCE) in school

Objective structured clinical examination comprehensively assesses clinical students' theoretical knowledge, practical skills, and clinical thinking abilities. According to the talent cultivation objectives, teaching syllabus, and internship syllabus of the clinical medical specialty, and with reference to the national qualification examination for clinical practicing physicians, the graduation skills assessment of the university has set up nine stations, with the examination time for stations 1–8 being 15 min each and 35 min for station 9, making a total of 155 min ([Zhang et al., 2025](#)). 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, nursing, etc., The specific stations are set up as shown in [Table 2](#). The annual examination is scheduled after the fifth-year student has completed all clinical rotations and prior to graduation. It is conducted in a uniform manner in the Clinical Skills Laboratory Center.

2.3.3 National medical licensing examination (NMLE)

National Medical Licensing Examination (NMLE), serving as a crucial standard for evaluating physicians' professional literacy and comprehensive abilities, covers a wide range of in-depth topics encompassing the four core modules of Basic Medicine Science, Medical Humanities and Regulations, Preventive Medicine, Clinical Medical Sciences ([Wang, 2022](#)). This examination is not only a test of physicians' knowledge reserve, but also a comprehensive assessment of their clinical thinking and practical abilities (See [Table 3](#)).

In terms of the examination format, NMLE adopts a module-cognition level design. The four modules can be divided into three levels: memory, understanding, and application, to comprehensively assess physicians' memory ability, understanding ability, and application ability of knowledge points. This design makes the examination more scientific and reasonable, enabling a more accurate evaluation of physicians' professional literacy and comprehensive abilities.



2.4 Data analysis

We used R studio 4.3.3 (R Studio, Inc., Boston, Massachusetts) “tidyverse” and “dplyr” packages for data collation, the “psych” package for descriptive statistics. And Wilcoxon rank sum test was performed on non-normal distribution data using “stats” package, the “ggpubr” package for violin plot visualization, with $p < 0.05$ considered statistically significant.

3 Results

3.1 Wilcoxon rank sum test in theory and OSCE performance between different teaching groups

To examine the class difference between the scores of theoretical scores and each OSCE module results, we conducted a Wilcoxon rank sum test (Figure 2 and Table 4). The results indicated that the two classes score comparably in the theoretical examination ($Z = 1.713, p = 0.087$). It is worth noting that in terms of OSCE module implementation, students from Nanshan class were significantly higher than those of Clinical Medicine in Medical History Collection (Clinical Medicine: $Me = 76.69, IQR = 13.58$; Nanshan: $Me = 83.19, IQR = 8.75; Z = 1.979, p = 0.048$) and Physical Examination (Clinical Medicine: $Me = 72.26, IQR = 11.10$; Nanshan: $Me = 77.80, IQR = 11.42; Z = 2.405, p = 0.016$). However, no significant differences were found in the Basic Operational Skills, Auxiliary Examination, Case Analysis modules, and the total score ($p > 0.05$). A comparison by gender

TABLE 1 Settings of final exam.

Areas	Contents	Proportion (%)
Basic medical sciences	Anatomy	15–20
	Pathophysiology	
	Immunology	
	Pharmacology	
Clinical disciplines	Internal medicine	65–70
	Surgery	
	Obstetrics and gynecology	
	Pediatrics	
Medical humanities and prevention	Medical ethics	10–20
	Medical regulations	
	Medical psychology	
	Preventive medicine	

(Supplementary Figure 1 and Supplementary Table 1) revealed that male and female students differed significantly in their performance, with male performing better at theoretical test (Clinical Medicine: $Me = 71.00, IQR = 7.00$; Nanshan: $Me = 69.00, IQR = 7.00; Z = 2.129, p = 0.033$), and female scoring higher in the Medical History Collection (Clinical Medicine: $Me = 76.69, IQR = 18.44$; Nanshan: $Me = 81.44, IQR = 10.88; Z = 2.204, p = 0.028$), Basic Operational Skills (Clinical Medicine: $Me = 77.31, IQR = 7.76$; Nanshan: $Me = 83.88, IQR = 5.79; Z = 2.270, p = 0.023$), and total score (Clinical Medicine: $Me = 74.00, IQR = 7.00$;

TABLE 2 Settings of OSCE examine.

Contents		Time (minute)	Proportion (%)
Station 1	Medical history collection I	15	8
	Abnormal abdominal palpation I		2
Station 2	Normal physical examination	15	10
Station 3	Internal pediatrics skill operation	15	10
Station 4	Auxiliary examination	15	10
Station 5	Medical history collection II	15	8
	Abnormal abdominal examination II		2
Station 6	Surgical skill operation	15	10
Station 7	Obstetrics and gynecology skill operation	15	10
Station 8	Nursing skill operation	15	10
Station 9	Case analysis	5	2.5
	Cardiac auscultation	5	2.5
	Pulmonary auscultation	25	6
	Clinical medical skills examination	25	9

Nanshan: Me = 78.00, IQR = 5.00; $Z = 2.079$, $p = 0.038$). The results of grade difference show that students of grade 2015 have higher scores (Supplementary Figure 2 and Supplementary Table 2), which are significant in Medical History Collection (Clinical Medicine: Me = 85.22, IQR = 9.55; Nanshan: Me = 80.56, IQR = 14.53; $Z = 3.254$, $p = 0.001$), Physical Examination (Clinical Medicine: Me = 87.75, IQR = 5.64; Nanshan: Me = 73.92, IQR = 11.17; $Z = 7.478$, $p < 0.001$) and total score (Clinical Medicine: Me = 83.00, IQR = 6.00; Nanshan: Me = 75.00, IQR = 6.00; $Z = 6.978$, $p < 0.001$).

3.2 Wilcoxon rank sum test in NMLE performance between different teaching groups

In NMLE, a Wilcoxon rank sum test was conducted with the data set derived from the four modules (Basic Medical Sciences, Medical Humanities, and Regulations, Preventive Medicine, Clinical Medical Sciences), three levels (Memory-Comprehension-Application), as well as the combined theory and skill scores. These scores were compared between two distinct classes, and the results are presented in Figure 3 and Table 5. It revealed that NMLE results were almost consistent across different classes ($p > 0.05$). Similarly, differences in NMLE results were not evident

TABLE 3 Settings of NMLE.

Modules	Contents
Basic medicine science	Pathology
	Pathophysiology
	Anatomy
	Physiology
	Biochemistry
	Pharmacology
	Immunology
	Microbiology
Medical humanities and regulations	Medical ethics
	Medical regulations
	Medical psychology
Preventive medicine	Epidemiology
	Medical statistics
	Preventive services and health management
	Public disease prevention and control
Clinical medical sciences	Internal medicine
	Surgery
	Obstetrics and gynecology
	Pediatrics

between genders, with the exception of Basic Medical Sciences, in which female participants demonstrated superior performance (Clinical Medicine: Me = 51.00, IQR = 7.00; Nanshan: Me = 48.00, IQR = 9.00; $Z = 1.965$, $p = 0.049$, Supplementary Figure 3 and Supplementary Table 3). It is noteworthy that the students of grade 2016 showed significantly higher in Basic Medical Sciences (Clinical Medicine: Me = 51.00, IQR = 9.00; Nanshan: Me = 47.00, IQR = 8.00; $Z = 3.847$, $p < 0.001$), Medical Humanities and Regulations (Clinical Medicine: Me = 31.00, IQR = 2.00; Nanshan: Me = 32.00, IQR = 4.00; $Z = 2.483$, $p = 0.013$), Preventive Medicine (Clinical Medicine: Me = 20.00, IQR = 4.00; Nanshan: Me = 24.00, IQR = 5.00; $Z = 4.840$, $p < 0.001$). The remarkable aptitude of the students of grade 2016 is exemplified by their proficiency in memorization (Clinical Medicine: Me = 71.50, IQR = 9.00; Nanshan: Me = 53.00, IQR = 9.00; $Z = 6.721$, $p < 0.001$) and comprehension (Clinical Medicine: Me = 95.00, IQR = 11.00; Nanshan: Me = 125.00, IQR = 21.00; $Z = 6.631$, $p < 0.001$); other programs were not significantly different (Supplementary Figure 4 and Supplementary Table 4, $p > 0.05$).

4 Discussion

This study conducted a longitudinal evaluation of the localized OSBC by comparing the Nanshan class with the traditional discipline-based medical education model. Unlike prior cross-sectional studies limited to a single time point, this research examined both short-term indicators (e.g., graduation theoretical examination scores and OSCE performance) and long-term indicators (e.g., NMLE scores

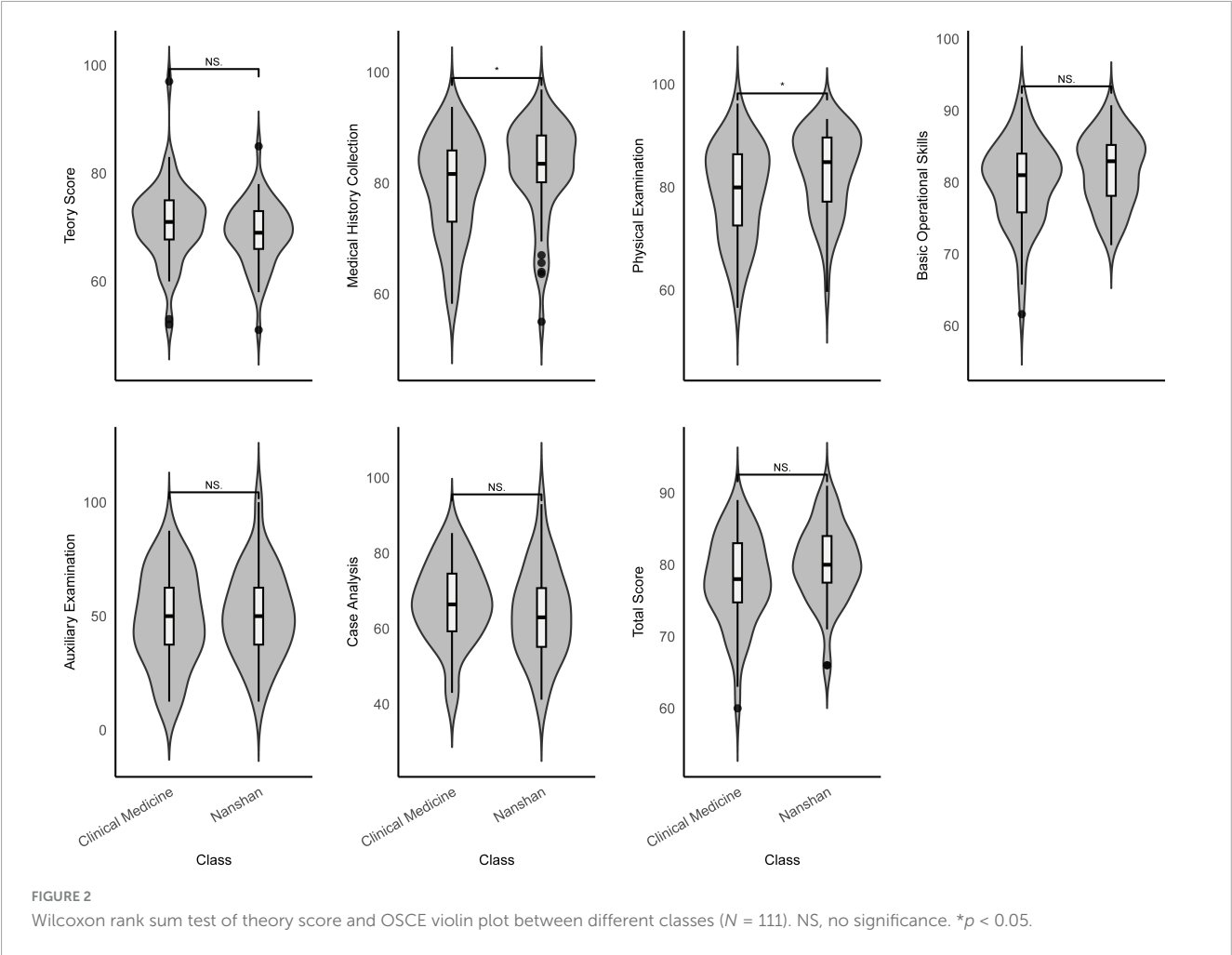


TABLE 4 Wilcoxon rank sum test results of theory score and OSCE module between different classes ($N = 111$).

Groups	Clinical medicine($n = 60$)		Nanshan($n = 51$)		Z	p
	Me	IQR	Me	IQR		
Theory score	71.00	7.75	69.00	7.00	1.713	0.087
OSCE module						
Medical history collection	76.69	13.58	83.19	8.75	1.979	0.048
Physical examination	72.26	11.10	77.80	11.42	2.405	0.016
Basic operational skills	81.00	9.11	82.95	9.18	1.435	0.151
Auxiliary examination	50.00	25.00	50.00	25.00	0.366	0.714
Case analysis	66.42	16.54	63.00	16.50	1.103	0.270
Total score	78.00	9.00	80.00	7.00	1.942	0.052

Me, median; IQR, interquartile range.

1 year after graduation) to comprehensively assess the sustained outcomes of curriculum reform (Leen et al., 2010). Subgroup analyses by gender and academic year were also introduced to further explore the interaction between student characteristics and educational pathways. Through this design, the study not only asks whether OSBC is effective, but also addresses for whom and under what circumstances it may be more effective, thereby offering an empirical

basis for future policy and instructional decisions in medical education.

4.1 The effectiveness of OSBC teaching

This study found that the Nanshan class demonstrated a statistically significant advantage over the traditional teaching class in two OSCE modules: Medical History Collection and

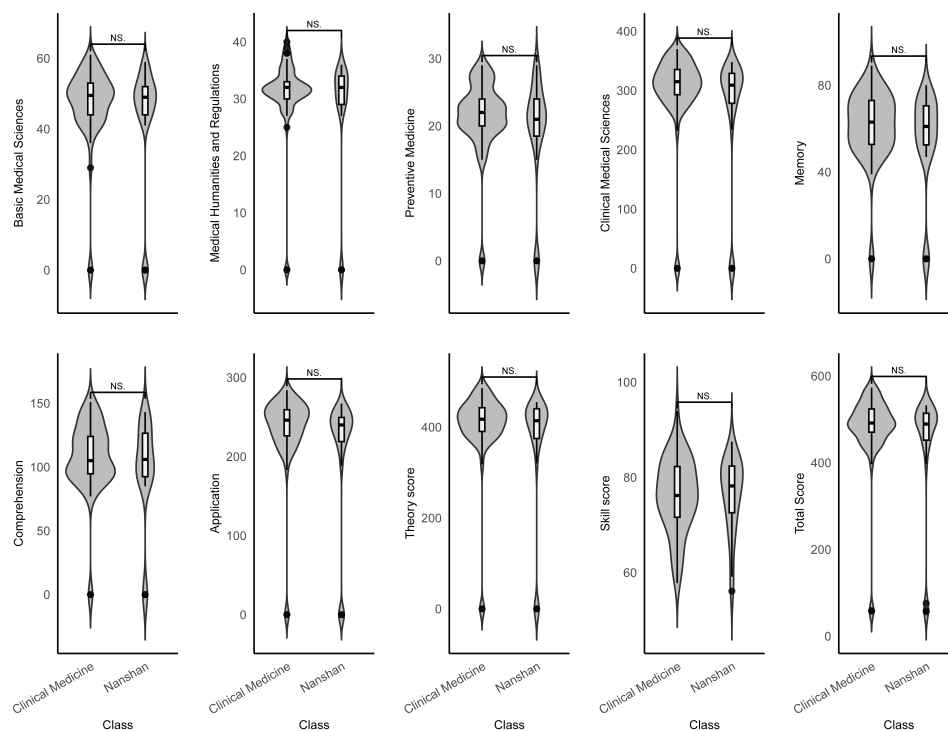


FIGURE 3

Wilcoxon rank sum test of NMLE violin plot between different classes ($N = 111$). NS, no significance.

Physical Examination. Although the score differences in these two modules were not large, the results still suggest that the OSBC teaching model may help enhance students' performance in structured clinical examinations, which is consistent with previous empirical findings on the benefits of systems-integrated curricula in improving practical skills (Feldacker et al., 2014; Hale et al., 2023).

A retrospective examination of the curriculum design suggests that the performance of the Nanshan cohort in structured assessments may be closely related to the systemic architecture of its teaching model. This cohort adopted an organ system-based modular structure that sought to integrate foundational theory, clinical knowledge, and procedural skills into a coherent instructional framework, emphasizing a longitudinal trajectory from conceptual understanding to applied competencies. Supported by an interdisciplinary teaching team, students progressively developed an organized knowledge system within a structured framework, enabling them to flexibly mobilize prior experience and formulate integrative responses under standardized assessment conditions (Wu et al., 2021). Existing empirical studies have also indicated that students with well-developed cognitive integration and reasoning capabilities tend to demonstrate more coherent and structured thinking when confronted with complex clinical scenarios (Yin et al., 2022). However, it is important to acknowledge that system-integrated teaching is not a panacea. Its effectiveness largely depends on the efficiency of interdisciplinary collaboration, the standardization of curriculum content, and the sustained availability of qualified teaching resources. If these supporting mechanisms are underdeveloped, the curriculum may lead to fragmented knowledge and inconsistencies in instructional delivery, thereby undermining the coherence and

integrity of students' cognitive structures (Zhang et al., 2021). As such, the current educational outcomes of the Nanshan Class should be interpreted with caution, and future evaluations should incorporate longer-term and more comprehensive indicators for ongoing assessment.

The formative assessment mechanisms embedded in practical training also warrant attention. During the clinical internship phase, the Nanshan program incorporated tools such as Mini-Clinical Evaluation Exercise and Direct Observation of Procedural Skills to guide students in receiving immediate feedback during hands-on practice and continuously adjusting their strategies accordingly (Wu et al., 2021). Studies have shown that such tools are effective in enhancing students' reflective awareness and adaptability in clinical settings (Norcini and Burch, 2007). In the modules of medical history taking and physical examination, where the assessment objectives are clearly defined, training procedures are standardized, and practice occurs with high frequency, students are more likely to achieve proficiency through focused training. This may partly explain why the Nanshan cohort demonstrated a more pronounced advantage in these specific modules. However, concerns remain regarding the subjectivity and inter-rater consistency of these assessment tools. Some studies have questioned their validity in high-stakes examination contexts (Moonen-van Loon et al., 2013).

In addition to structured training, the curriculum also incorporated real case discussions and bedside teaching, offering students a more authentic clinical context for learning (Wu et al., 2021). Research has shown that context-based learning approaches can help stimulate intrinsic motivation and facilitate the transfer of knowledge to practical settings (Yardley et al., 2012). However, the

TABLE 5 Wilcoxon rank sum test results of NMLE between different classes ($N = 111$).

Groups	Clinical medicine($n = 60$)		Nanshan($n = 51$)		Z	p
	Me	IQR	Me	IQR		
Basic medical sciences	49.50	9.00	49.00	8.00	−0.308	0.758
Medical humanities and regulations	32.00	3.00	32.00	5.00	−0.610	0.542
Preventive medicine	22.00	4.00	21.00	6.00	−0.659	0.510
Clinical medical sciences	315.00	50.00	309.00	55.00	−1.106	0.269
Memory	63.00	21.00	61.00	19.00	−0.776	0.438
Comprehension	105.00	30.00	106.00	37.00	−0.459	0.646
Application	246.00	35.00	240.00	33.00	−1.574	0.115
Theory score	417.50	57.00	414.00	71.00	−0.835	0.404
Skill score	76.20	11.15	78.20	10.65	−0.216	0.829
Total score	491.60	57.19	489.35	72.10	−0.601	0.548

Me, median; IQR, interquartile range.

effectiveness of such teaching in enhancing complex diagnostic and integrative analytical skills tends to rely more heavily on students’ prior knowledge and the overall quality of instruction. This may partly explain why the Nanshan class did not demonstrate a distinct advantage in other OSCE modules, as this type of training is more susceptible to individual differences. Moreover, the success of this teaching model is also contingent on resource allocation and the quality of clinical supervision. Without adequate support, students may fall into a passive state of “seeing much but learning little” (Spencer, 2003).

From the perspective of student experience, the Nanshan class received widespread recognition. The satisfaction survey was conducted anonymously via the “Teaching Evaluation” app embedded in Enterprise WeChat and the universities’ student services system, covering all 51 students with a 100% response rate. Using a 0–5 rating scale, the results indicated a consistently high level of overall satisfaction, with the lowest average score reaching 4.24 (Zhang et al., 2021). Some students specifically mentioned that the course provided practical support in areas such as communication and collaboration, clinical reasoning, and critical thinking. Such feedback indirectly suggests that the course may have had a positive impact on the development of students’ comprehensive competencies (Eva et al., 2016).

4.2 Multiple influencing factors of NMLE measurement

The comparable performance between Nanshan class students and those in the standard clinical medicine track on the NMLE largely reflects the structural characteristics of this standardized assessment. Designed to evaluate whether medical graduates meet baseline requirements in knowledge and clinical skills, the NMLE enforces a tightly regulated content framework. Although Nanshan class implemented a systematic curriculum reform emphasizing scientific literacy and integrative competencies, the instructional process was still aligned with the national examination blueprint to ensure coverage of essential content. As a result, both groups received similar test-oriented training, which made it difficult for

differences in pedagogical approaches to translate into measurable performance gaps.

In addition, the students included in this study were still in the early stages of their professional development and had not yet undergone extended, independent clinical practice. At the time of the examination—one-year post-graduation—their capabilities were concentrated at a foundational level. Higher-order skills such as clinical reasoning, problem-solving, and innovative thinking were not yet fully developed, and individual differences in these domains remained minimal (Zhou et al., 2023). These developmental constraints likely limited the external expression of reform effects, rendering them difficult to detect through a standardized examination format.

More critically, the NMLE itself presents significant limitations in evaluative scope. As it primarily relies on written tests and standardized procedural assessments, the exam fails to capture essential domains of clinical competence, including scientific inquiry, interdisciplinary integration, and patient communication. Prior research has shown that standardized examinations have limited validity in predicting long-term clinical performance, adaptive capacity, and professional growth (Lurie et al., 2009; Pangaro and ten Cate, 2013). Particularly in complex, real-world settings, standardized assessment tools have been found insufficient to capture learners’ cognitive flexibility and developmental potential, revealing inherent structural limitations within these models (Gong and Marion, 2006; Hohl and Dolcos, 2024).

In light of these concerns, a growing body of scholarship calls for a shift from static, score-centric evaluation systems to multidimensional, developmental, and process-based assessment frameworks. On one hand, emerging tools, such as AI-assisted scoring systems, cross-cultural case-based assessments, and growth-oriented reflective portfolios—offer opportunities to enhance both depth and ecological validity (Wang, 2022; Wu et al., 2022; Lin et al., 2025). On the other hand, sustained tracking of higher-order cognitive domains, such as critical thinking, clinical judgment, and integrative reasoning—must be systematically embedded into assessment design (Qureshi et al., 2022; Supper et al., 2023). Against this backdrop, it is imperative to develop a composite evaluation model that integrates clinical performance with longitudinal developmental trajectories, so

as to more comprehensively capture the actual effectiveness of competency-based education. Particularly in contexts where high-stakes examinations and rating-based assessments are widely relied upon, issues such as rater bias and construct-irrelevant variance have posed substantive threats to the validity and reliability of assessment outcomes. While standardized tests like NMLE serve essential gatekeeping functions, they are clearly insufficient for reflecting the multifaceted impact of curricular reform (Downing, 2005). To mitigate these validity threats and to ensure meaningful alignment between instructional intentions and assessment practices, the establishment of a more comprehensive, dynamic, and theoretically grounded evaluation system is urgently required.

4.3 Limitations and future research

However, it should be noted that the study is not without limitations, which could be improved upon in future research. Firstly, it should be noted that the OSCE results of different grades could not be compared due to the failure of students in grade 2015 to complete the Auxiliary Examination and Case Analysis module as a result of the impact of the epidemic. Consequently, the scores for this module were not available for inclusion in the analysis. Additionally, the sample size of this study is relatively limited, and the number of OSCE modules exhibiting notable differences in gender and grade is not substantial. In the NMLE difference results, there was a tendency for the scores of males and females to be consistent. It would be advisable to consider expanding the sample size further in order to avoid the potential for data bias. When selecting different groups of students for comparison, this study lacked baseline measurements to better understand the status of students before the intervention with different teaching methods, despite having been admitted with the same class score. When the background information of the students was investigated, such as identity (master's degree students/standardized training for resident doctors/primary hospital doctors), self-learning ability, class resources, family concept and economic conditions should also be included and controlled to make the survey results more complete. Moreover, this study is primarily focused on one medical college in South China, which may limit the generalizability of the discussed issues and solutions. It is possible that the Nanshan class teaching method may yield different outcomes in other medical schools or in broader educational systems. Further research is therefore required to validate its applicability in these contexts. Finally, educational reform processes frequently encounter a number of challenges, including inadequate teaching staff, limited teaching resources, and diverse student capabilities. Future research should employ a mixed-methods approach, control for various confounding factors, conduct focus group interviews to collect qualitative data, and combine cross-sectional and longitudinal tracking data to assess students' clinical literacy and innovation ability.

5 Conclusion

This study conducted an in-depth analysis of the OSBC approach in the Nanshan class, confirming that it showed early

promise in skill development, while long-term benefits remain unproven. Furthermore, this study indicates that the strategic development of teachers can be achieved through measures such as establishing interdisciplinary teams, iterative curriculum design, and focusing on formative evaluations of students and course feedback, thereby ensuring that the development of teachers aligns with the educational goals of the institution. OSBC not only provides valuable teaching references and suggestions for medical schools but also helps enhance teaching quality and promote the development of students' comprehensive abilities. Its experiences and concepts have significant implications for driving development in medical education localization.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

HZ: Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing – original draft. JL: Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. LZ: Validation, Writing – original draft, Writing – review & editing. JZ: Conceptualization, Investigation, Resources, Writing – original draft. XC: Conceptualization, Investigation, Resources, Writing – original draft. SW: Writing – original draft. YL: Conceptualization, Software, Supervision, Visualization, Writing – review & editing. JianL: Conceptualization, Funding acquisition, Project administration, Resources, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the 2023 Basic and Applied Basic Research Foundation of Guangdong Province (grant number 2023A1515110350), the 2024 Youth Project of Guangdong Office of Philosophy and Social Science (grant number GD24YXL06), the Plan on enhancing scientific research in GMU (grant number 02-410-2302338XM), Guangdong-Hong Kong-Macao Greater Bay Area Medical and Health Industry High Quality Development Rule of Law Guarantee Research Center (grant number 2024TSZK016) and Ministry of Education 2023 Industry-university Collaborative Education program (grant number J24372010).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- An, Z. (2022). The influence of teacher discipline on teaching effect and students' psychology in universities and the normative suggestions for discipline behavior. *Front. Psychol.* 13:910764. doi: 10.3389/fpsyg.2022.910764
- Bolton, D. (2023). A revitalized biopsychosocial model: Core theory, research paradigms, and clinical implications. *Psychol. Med.* 53, 7504–7511. doi: 10.1017/S0033291723002660
- Campbell, M. K., Thomson, S., Ramsay, C. R., MacLennan, G. S., and Grimshaw, J. M. (2004). Sample size calculator for cluster randomized trials. *Comput. Biol. Med.* 34, 113–125. doi: 10.1016/S0010-4825(03)00039-8
- Cheng, Y. C. (2005). "New paradigm for education reforms: Globalization, localization, and individualization," in *Education in the Asia-Pacific region: Issues, concerns and prospects*, ed. Y. C. Cheng (Dordrecht: Springer Netherlands), 19–44. doi: 10.1007/1-4020-3620-5_2
- Downing, S. M. (2005). Threats to the validity of clinical teaching assessments: What about rater error? *Med. Educ.* 39, 353–355. doi: 10.1111/j.1365-2929.2005.02138.x
- Eleftheriou, A., Rokou, A., Argyriou, C., Papanas, N., and Georgiadis, G. S. (2022). Web-based medical education during COVID-19 lockdown: A step back or a leap to the future? *Int. J. Low Extrem. Wounds* 21, 272–274. doi: 10.1177/15347346211011848
- Eva, K. W., Bordage, G., Campbell, C., Galbraith, R., Ginsburg, S., Holmboe, E., et al. (2016). Towards a program of assessment for health professionals: From training into practice. *Adv. Health Sci. Educ.* 21, 897–913. doi: 10.1007/s10459-015-9653-6
- Feldacker, C., Chicumbe, S., Dgedge, M., Augusto, G., Cesar, F., Robertson, M., et al. (2014). Mid-level healthcare personnel training: An evaluation of the revised, nationally-standardized, pre-service curriculum for clinical officers in Mozambique. *PLoS One* 9:e102588. doi: 10.1371/journal.pone.0102588
- Fritz, C. O., Morris, P. E., and Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *J. Exp. Psychol. General* 141, 2–18. doi: 10.1037/a0024338
- Gong, B., and Marion, S. (2006). *Dealing with flexibility in assessments for students with significant cognitive disabilities*. Minneapolis, MI: National Center on Educational Outcomes.
- Hale, A. J., Bartsch, J., Stapleton, R. D., and Parsons, P. E. (2023). How the hospital works: An interdisciplinary, systems-based practice medical student elective. *J. Med. Educ. Curric. Dev.* 10:23821205231203908. doi: 10.1177/23821205231203908
- Harden, R. M., Stevenson, M., Downie, W. W., and Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *Br. Med. J.* 1, 447–451. doi: 10.1136/bmj.1.5955.447
- Hohl, K., and Dolcos, S. (2024). Measuring cognitive flexibility: A brief review of neuropsychological, self-report, and neuroscientific approaches. *Front. Hum. Neurosci.* 18:1331960. doi: 10.3389/fnhum.2024.1331960
- Leen, T., Williams, T. A., Campbell, L., Chamberlain, J., Gould, A., McEntaggart, G., et al. (2010). Early experience with influenza A H1N109 in an Australian intensive care unit. *Intens. Crit. Care Nurs.* 26, 207–214. doi: 10.1016/j.iccn.2010.05.005
- Lin, W., Xu, L., Yin, T., Zhang, Y., Huang, B., Zhang, X., et al. (2025). Exploring the role of moxibustion robots in teaching: A cross-sectional study. *BMC Med. Educ.* 25:58. doi: 10.1186/s12909-025-06669-y
- Lurie, S. J., Mooney, C. J., and Lyness, J. M. (2009). Measurement of the general competencies of the accreditation council for graduate medical education: A systematic review. *Acad. Med.* 84, 301–309. doi: 10.1097/ACM.0b013e3181971f08
- Moonen-van Loon, J. M. W., Overeem, K., Donkers, H. H. L. M., Van Der Vleuten, C. P. M., and Driessen, E. W. (2013). Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Adv. Health Sci. Educ.* 18, 1087–1102. doi: 10.1007/s10459-013-9450-z
- Norcini, J., and Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med. Teacher* 29, 855–871. doi: 10.1080/01421590701775453
- O'Connell, M. T., and Pascoe, J. M. (2004). Undergraduate medical education for the 21st century: Leadership and teamwork. *Fam. Med.* 36(Suppl.), S51–S56.
- Pangaro, L., and ten Cate, O. (2013). Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Med. Teach.* 35, e1197–1210. doi: 10.3109/0142159X.2013.788789.
- Qureshi, S. S., Larson, A. H., and Vishnumolakala, V. R. (2022). Factors influencing medical students' approaches to learning in Qatar. *BMC Med. Educ.* 22:446. doi: 10.1186/s12909-022-03501-9
- Sakles, J., Maldonado, R., and Kumari, V. (2006). Integration of basic sciences and clinical sciences in a clerkship: A pilot study. *Med. Sci. Educ.* 16, 4–9.
- Smithson, S., Beck Dallaghan, G., Crowner, J., Derry, L. T., Vijayakumar, A. A., Storrie, M., et al. (2020). Peak performance: A communications-based leadership and teamwork simulation for fourth-year medical students. *J. Med. Educ. Curric. Dev.* 7:2382120520929990. doi: 10.1177/2382120520929990
- Song, F., Yang, G., and Qi, L. (2017). Analysis of clinical thinking ability training of clinical medical students. *China Continuing Med. Educ.* 9, 39–40. doi: 10.3969/j.issn.1674-9308.2017.01.022
- Spencer, J. (2003). Learning and teaching in the clinical environment. *BMJ* 326, 591–594. doi: 10.1136/bmj.326.7389.591
- Stan, M. M., Topală, I. R., Necșoi, D. V., and Cazan, A.-M. (2022). Predictors of learning engagement in the context of online learning during the COVID-19 pandemic. *Front. Psychol.* 13:867122. doi: 10.3389/fpsyg.2022.867122
- Supper, P., Urban, D., Acker, I., Linke, F. S., Kienast, P., Praschinger, A., et al. (2023). A concept for adapting medical education to the next generations via three-staged digital peer teaching key feature cases. *Wien Med. Wochenschr.* 173, 108–114. doi: 10.1007/s10354-022-00990-7
- Wang, J., Wang, B., Liu, D., Zhou, Y., Xing, X., Wang, X., et al. (2024). Video feedback combined with peer role-playing: A method to improve the teaching effect of medical undergraduates. *BMC Med. Educ.* 24:73. doi: 10.1186/s12909-024-05040-x
- Wang, W. (2021). Medical education in China: Progress in the past 70 years and a vision for the future. *BMC Med. Educ.* 21:453. doi: 10.1186/s12909-021-02875-6

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2025.1629192/full#supplementary-material>

- Wang, X. (2022). Experiences, challenges, and prospects of national medical licensing examination in China. *BMC Med. Educ.* 22:349. doi: 10.1186/s12909-022-03385-9
- Welch, A., and Mok, K.-H. (2003). *Globalization and educational restructuring in the Asia Pacific region*. London: Palgrave Macmillan.
- Wijnen-Meijer, M. (2023). Implications of internationalisation of medical education. *BMC Med. Educ.* 23:640. doi: 10.1186/s12909-023-04630-5
- Wu, Q., Wang, Y., Lu, L., Chen, Y., Long, H., and Wang, J. (2022). Virtual simulation in undergraduate medical education: A scoping review of recent practice. *Front. Med.* 9:855403. doi: 10.3389/fmed.2022.855403
- Wu, T., Li, J., Zheng, J., Zhang, H., and Yin, Z. (2021). Exploring the training mode of elite medical talents based on practices of nanshan class. *Med. Educ. Res. Pract.* 29, 342–345. doi: 10.13555/j.cnki.c.m.e.2021.03.002
- Xiao, Y., Wu, X.-H., Huang, Y.-H., and Zhu, S.-Y. (2022). Cultivation of compound ability of postgraduates with medical professional degree: The importance of double tutor system. *Postgraduate Med. J.* 98, 655–657. doi: 10.1136/postgradmedj-2021-139779
- Xiao, A., He, W., Tang, Q., Wen, Y., Liao, H., Ye, S., et al. (2011). How to cultivate high-quality cardiovascular ultrasound diagnostician. *Chin. J. Med. Educ.* 31, 609–611. doi: 10.3760/cma.j.issn.1673-677X.2011.04.046
- Yan, J., Wen, Y., Liu, X., Deng, M., Ye, B., Li, T., et al. (2024). The effectiveness of problem-based learning and case-based learning teaching methods in clinical practical teaching in TACE treatment for hepatocellular carcinoma in China: A bayesian network meta-analysis. *BMC Med. Educ.* 24:665. doi: 10.1186/s12909-024-05615-8
- Yan, Y. (2023). Model construction and path optimization of cultivating top innovative talents in universities. *Front. Educ. Res.* 6:90–95. doi: 10.25236/FER.2023.061916
- Yang, H. (2013). Deal with diversified social needs of elite education. *J. Chifeng Univ.* 19, 162–164. doi: 10.3969/j.issn.1673-260X.2013.19.067
- Yang, J., Guo, Y., and Tang, Y. (2012). Cultivation of medical students' clinical comprehensive ability in clinical teaching of hepatobiliary surgery. *Chin. J. Med. Educ. Res.* 1292–1294. doi: 10.3760/cma.j.issn.2095-1485.2012.12.02
- Yardley, S., Teunissen, P. W., and Dornan, T. (2012). Experiential learning: Transforming theory into practice. *Med. Teach.* 34, 161–164. doi: 10.3109/0142159X.2012.643264
- Yin, Z., Wu, T., and Ma, J. (2022). An empirical study on the construction of index system of the lifelong learning competence of medical students. *Zhejiang Med. Educ.* 21, 65–72. doi: 10.20019/j.cnki.1672-0024.2022.02.065.08
- Zhang, H., Li, S., Zheng, G., Chen, X., Zhong, L., Li, J., et al. (2025). Network analysis of an OSCE-based graduation skills assessment for clinical medical students. *BMC Med. Educ.* 25:605. doi: 10.1186/s12909-025-07091-0
- Zhang, H., Wu, T., Zhang, H., Lin, A., Yang, L., Pan, C., et al. (2021). Construction of top innovative talents training system of “Nanshan Class.”. *Med. Educ. Manag.* 7, 485–490. doi: 10.3969/j.issn.2096-045X.2021.05.004
- Zhou, F., Sang, A., Zhou, Q., Wang, Q. Q., Fan, Y., and Ma, S. (2023). The impact of an integrated PBL curriculum on clinical thinking in undergraduate medical students prior to clinical practice. *BMC Med. Educ.* 23:460. doi: 10.1186/s12909-023-04450-7
- Zhou, R. (2013). *Research on construction of training mode for outstanding doctors majoring in five-year program of clinical medicine*. China: Anhui Medical University.