Application of Programmatic Assessment in the Evaluation of General Practice Clinical Reasoning Curriculum: Postprint

Authors: Zhai Jiayi, Lu Yuan, Shi Jianjun, Yu Dehua, Lu Yuan

Date: 2024-05-23T00:00:00+00:00

Abstract

Background Clinical reasoning in general practice is essential for general practitioners to deliver high-quality diagnostic and therapeutic services. However, as an abstract competence that cannot be directly observed, it is not amenable to assessment using traditional methods. Objective This study aims to enhance the training efficiency of general practitioners' clinical reasoning and enrich medical education evaluation by optimizing curriculum teaching assessment methods. Methods From September to December 2022, 38 master' s students in general practice from Tongji University School of Medicine were enrolled and divided into an on-the-job group (n=16) and a residency training group (n=22) based on student source categories. The curriculum comprised three modules: core, case, and outcome, and was delivered using a blended teaching approach. Quantitative evaluation consisted of skill rating scales, PBL rating scales, and case rating scales, while subjective evaluation content was derived from multisource feedback. Statistical analysis was performed using SPSS 22.0 software. Results No statistically significant difference was observed between the two groups in the core and case module scores during the initial session ($P>0.05$), but statistically significant differences were found in both modules during the final session ($P<0.05$). Comparisons of quantitative evaluation scores between initial and final sessions for the entire class and for both groups in the core and case modules all revealed statistically significant differences ($P<0.001$). Subjective evaluation categories were classified as positive feedback and constructive feedback. As the curriculum progressed, the quantity and percentage of positive feedback for both the core and case modules demonstrated a continuous upward trend, while the quantity and percentage of constructive feedback showed a continuous downward trend. Conclusion Procedural assessment can enrich the general practice reasoning evaluation system and facilitate students' progressive construction of general practice reasoning. Additionally, the flipped classroom was identified as an effective presentation format for procedural assessment.

Full Text

Effects of Programmatic Assessment in Clinical Reasoning Courses in General Practice

ZHAI Jiayi[1],[2],[3], LU Yuan[1],[3],[4]*, SHI Jianjun[1],[3],[4], YU Dehua[1],[3],[4]

[1]Department of General Practice, Yangpu Hospital of Tongji University, Shanghai 200090, China
[2]Community Health Service Center of Jiading Town in Jiading District, Shanghai 201899, China
[3]Academic Department of General Practice, Tongji University School of Medicine, Shanghai 200090, China
[4]Shanghai General Practice and Community Health Development Research Center, Shanghai 200090, China
*Corresponding author: LU Yuan, Chief physician; Email: lussy@126.com

Abstract

Background: Clinical reasoning in general practice is essential for general practitioners to provide high-quality medical services. However, as an abstract ability that cannot be directly observed, it is not suitable for evaluation using traditional methods. **Objective:** This study aims to improve the efficiency of general practitioners' clinical reasoning training and enrich medical education evaluation by optimizing curriculum teaching assessment methods. **Methods:** From September to December 2022, 38 master' s degree students in the general practice track at Tongji University School of Medicine were enrolled and divided into an in-service group (n=16) and a residency training group (n=22) based on their student category. The course comprised three modules—core, case, and outcome—and was delivered using a blended teaching approach. Quantitative evaluation consisted of skill rating scales, PBL rating scales, and case rating scales, while subjective evaluation content was derived from multi-source feedback. SPSS 22.0 software was used for statistical analysis. **Results:** There was no statistically significant difference between the two groups in the first-session scores for the core and case modules ($P>0.05$), but the final-session scores for both modules showed statistically significant differences ($P<0.05$). Comparisons of quantitative evaluation scores between the first and final sessions for the core and case modules were statistically significant for both the entire class and each group separately ($P<0.001$). Subjective evaluations were categorized as positive or constructive. As the course progressed, the number and percentage of positive evaluations for both the core and case modules showed a continuous upward trend, while constructive evaluations showed a continuous downward trend. **Conclusion:** Programmatic assessment can enrich the evaluation system for general practice thinking, promote the sequential construction of general practice thinking among trainees, and we also found that the "flipped classroom" serves as an effective presentation format for programmatic assessment.

Keywords: General practice; Teaching; Education, graduate; Programmatic assessment; Clinical reasoning; Flipped classroom

---

General practice clinical reasoning (hereinafter referred to as "GP thinking") is a complex competency that integrates general practitioners' history-taking, physical examination, and interpretation of auxiliary examinations, serving as the key to providing high-quality clinical care [**?**, **?**]. Currently, training formats for GP thinking have become increasingly diversified [**?**, **?**], such as Problem-Based Learning (PBL) and Case-Based Learning (CBL). However, evaluation of GP thinking, as a crucial means of measuring training effectiveness, also employs various assessment forms. Single assessment methods struggle to comprehensively and objectively reflect trainees' overall GP thinking abilities [**?**, **?**].

According to Miller' s Pyramid theory [**?**], GP thinking development progresses through four stages: "knows," "knows how," "shows how," and "does." The first three stages are termed "standardized assessment," typically evaluated through theoretical exams and skill operation tests. The "does" stage requires "non-standardized assessment" [**?**, **?**] because integrated professional knowledge, skills, humanities, and collaboration—termed "domain-independent competencies" — cannot be comprehensively evaluated by conventional methods and necessitate combined qualitative and quantitative "non-standardized assessment" throughout the course. Programmatic Assessment (PA), first proposed by VAN et al. [**?**] in 2005, addresses both assessment requirements. It follows five principles: "meaningful learning," "longitudinal assessment," "continuous feedback," "learner-centeredness and self-responsibility," and "assessment for learning and as learning" [**?**]. PA combines the advantages of formative assessment (promoting reflection) and summative assessment (summarizing outcomes) [**?**], employing both standardized and non-standardized assessments for GP thinking. Any single assessment method functions as a "pixel" that cannot fully capture trainees' overall GP thinking; only when sufficient "pixels" accumulate to form a clear "image" can pass/fail decisions be made [**?**, **?**]. Beyond optimizing instructor decision-making, PA' s more important function is enabling trainees to actively construct professional knowledge and improve training efficiency [**?**, **?**].

Clinical Reasoning in General Practice is a required competency-enhancement course for general practitioners at Tongji University and a shared compulsory course for two types of graduate students in the general practice track. Through years of practice and exploration, it has achieved certain effectiveness [**?**, **?**]. This study introduced PA into this course to explore how improved assessment methods can enhance training efficiency and provide insights for cultivating high-quality general practitioners.

1.1 Teaching Participants

Thirty-eight master' s students enrolled in the *Clinical Reasoning in General Practice* course in 2022 served as research subjects, including in-service general

practitioners, specialists interested in transitioning to general practice, and residents in standardized general practice training. Participants were divided into two groups based on their learning and work experience: the in-service group comprised 16 first-year academic master' s students (hereinafter "in-service trainees"), while the residency training group comprised 22 second-year professional master's students in the "four certificates integrated" program (hereinafter "residency trainees"). Participant baseline characteristics are shown in Table 1 .

1.2 Course Content

The *Clinical Reasoning in General Practice* course totaled 32 class hours. Excluding the general introduction, it was divided into three major modules: core, case, and practice, delivered through lectures, discussions, and practical sessions. Specific hour allocations are shown in Table 2 . The core module covered history-taking, physical examination, auxiliary examinations, doctor-patient communication, and SOAP note writing. The case module included five categories of typical general practice cases: health management, chronic disease management, multimorbidity, psychosomatic disorders, and undifferentiated diseases. The outcome module assessed trainees' GP thinking through case presentations.

1.3.1 PA Design Philosophy

PA comprises three modules and ten principles [**?**, **?**, **?**, **?**], detailed in Table 3 . The standardized assessment module includes four principles: "competency-specific, not generic," "objectivity differs from reliability," "stimulus matters more than response," and "validity is built-in." The non-standardized assessment module also includes four principles: "bias is inherent in professional judgment," "assessment validity depends on the tool user," "subjective content matters," and "effective feedback use."

1.3.1.1 Standardized Assessment Module: Principles, Content, and Application

(1) **"Competency-specific, not generic" principle:** Performance in one competency test does not necessarily predict performance in others. For instance, trainees with strong theoretical scores may not excel in practical skills—this is termed "content specificity" of clinical competence [**?**]. PA combines formative and summative assessment characteristics. The framework design establishes longitudinal assessment with summative evaluation at course completion and formative assessment during the course. Evaluation content must be comprehensive, assessing not only core GP thinking competencies but also doctor-patient communication skills.

(2) **"Objectivity differs from reliability" principle:** Even structured, objective assessments like OSCE face reliability challenges. Test comprehensiveness is key, and under these circumstances, subjective assessments

can also be reliable [**?**]. Therefore, the course includes direct observation assessments and real-time commentary as subjective evaluation methods.

(3) **"Stimulus matters more than response" principle:** Assessment content is determined by "stimulus" rather than "response." "Stimulus" refers to assessment formats such as questions, examinations, and tests, while "response" refers to trainees' answering processes and results [**?**]. This principle suggests focusing on stimulus quality and format, such as using authentic cases and scenarios. All cases in this course were derived from real general practice clinic cases [**?**, **?**], enhancing stimulus authenticity. In terms of format, core and case modules employed PBL and CBL "flipped classroom" approaches, with classroom performance evaluations further enriching stimulus variety.

(4) **"Validity is built-in" principle:** This emphasizes that evaluation and curriculum quality improvement measures include pre-assessment design, mid-assessment guidance, and post-assessment analysis and summary [**?**, **?**].

1.3.1.2 Non-standardized Assessment Module: Principles, Overall Evaluation Content, and Application

(1) **"Bias is inherent in professional judgment" principle:** Bias exists in all assessments, but its impact on comprehensive professional judgment should be minimized [**?**, **?**]. To reduce bias, the course followed triangulation principles: each quantitative assessment employed two instructors to form a mini decision-making committee, and "flipped classroom" approaches increased instructor-trainee interaction, reducing stereotypical impressions.

(2) **"Assessment validity depends on the tool user" principle:** Instructor competency requires continuous improvement [**?**]. The course established a design team and an instructor team. The design team comprised six associate senior-level experts in general practice and medical education who understood trainees' baseline situations and needs before class, collected feedback afterward, and made appropriate course and assessment improvements within the overall framework while inviting peer experts to review the course and assessment plan to ensure quality. The instructor team comprised six general practice instructors with intermediate-level or higher professional titles and over five years of teaching experience who underwent unified training and trial lectures before class.

(3) **"Subjective content matters" and "Effective feedback use" principles:** When feedback becomes central to assessment, quantitative information has clear limitations if instructor-trainee interaction mediates effective feedback, while subjective evaluation content offers advantages [**?**]. Reviewing evaluation content, core competencies, communication, and classroom performance can all use combined quantitative and subjective assessment. According to PA' s "meaningful learning" principle,

trainees should construct GP thinking themselves, so self-directed learning ability evaluation was added. During framework design, feedback sessions were included after both core and case modules, aiming to improve trainees'next performance. Feedback methods were further improved by incorporating multi-source feedback (MSF) combining self-assessment, peer assessment, instructor assessment, and standardized patient assessment.

(4) **Overall evaluation:** This includes "no perfect assessment method" and "assessment drives learning" principles, suggesting that course assessments should combine quantitative and subjective methods whenever possible, with "scoring" followed by "commentary."

1.3.2 Evaluation Content Design Based on "Pixel Point" Theory

Based on PA' s "pixel point" theory [**?**, **?**] and the course' s three modules (core, case, and outcome), evaluation content was refined to assess GP thinking core competencies, communication skills, classroom performance, and self-directed learning ability, forming "pixel points" combining quantitative and subjective evaluation. Quantitative evaluation comprised 100-point rating scales for each module, while qualitative evaluation consisted of subjective assessments in each module' s MSF, categorized as positive or constructive evaluations [**?**]. Positive evaluations addressed trainees' performance strengths, while constructive evaluations identified areas needing improvement. The final collection of "pixel points" reflects the complete "image" of trainees' GP thinking.

Evaluation content design is detailed in Table 4 . The core module used custom rating scales assessing logic, relevance, completeness, alignment with general practice characteristics, and communication skills. The case module used custom rating scales evaluating diagnostic and treatment skills, management appropriateness, general practice environment fit, resource allocation, and case presentation quality.

1.4 Teaching Practice Process

From September to December 2022, teaching practice was conducted with 38 general practice track master' s students. Before the course, all trainees were fully informed about the research content and provided consent. The first session was a general introduction presenting the course overview to all trainees and instructor team members, explaining the evaluation format and instructor-trainee collaboration to establish a foundation for subsequent PA implementation. After the introduction, a baseline survey of all trainees was conducted.

The core module consisted of knowledge consolidation and practice. After instructor-guided review, trainees played the role of doctors practicing on standardized patients (SPs), with another trainee serving as an assistant to help solve problems when difficulties arose. During practice, instructors, other trainees, and SPs scored the practicing trainee, followed by MSF after the session.

The case module used online delivery, with each case taught in two sessions of one class hour each: the first session involved PBL discussion, and the second session involved trainee presentations. Before class, trainees were divided into six groups of 6-7 members each. PBL cases were distributed 1-2 days in advance. During the first session, each group brainstormed questions, after which the instructor summarized and assigned questions to groups. Trainees discussed assigned questions after class and prepared presentation materials, with group leaders organizing discussions and instructors participating online. Each group selected one trainee to present discussion content during the second session, followed by instructor commentary and knowledge point summary. Instructors and group leaders scored all trainees' discussion and classroom performance, with MSF providing on-site feedback after presentations.

The outcome module involved online case presentations where trainees independently prepared a general practice-characteristic case for presentation. Materials were submitted one week in advance for A-C level evaluation by instructors. Presentations were conducted in three sessions, with presentation order determined by level evaluation results. Level A presented first, while levels B and C reserved time, referenced level A presentations, and modified their presentations based on on-site commentary before presenting. Two instructors scored on-site presentations, followed by MSF after each presentation.

1.5 Statistical Methods

SPSS 22.0 software was used for statistical analysis of quantitative evaluation results. All data were normally distributed. Measurement data were expressed as $(\bar{x} \pm s)$, and count data as constituent ratios. Chi-square tests were used for rate comparisons, and t-tests for between-group and before-after comparisons. $P<0.05$ was considered statistically significant.

2.1 Quantitative Evaluation Results

Quantitative scores from this course were analyzed by comparing the two groups' scores across the core, case, and outcome modules to understand performance differences. Results showed no statistically significant difference in first-session scores for the core and case modules between groups ($P>0.05$), but final-session scores for both modules showed statistically significant differences ($P<0.05$). Comparisons of quantitative evaluation scores between first and final sessions for the core and case modules were statistically significant for both the entire class and each group separately ($P<0.001$). Case presentation scores are shown in Table 5 .

2.2 Subjective Evaluation Results

Transcription results were analyzed for quantity and percentage. As the course progressed, the number and percentage of positive evaluations for both core and

case modules showed a continuous upward trend, while constructive evaluations showed a continuous downward trend, as detailed in Table 6 .

3.1 PA Enriches the GP Thinking Evaluation System

Based on this course' s quantitative and subjective evaluations, PA comprehensively reflected trainees' GP thinking. Previous GP thinking assessments often used case reports, theoretical exams, OSCE, Direct Observation of Procedural Skills (DOPS), and Mini-Clinical Evaluation Exercise (Mini-CEX), with single-form teaching assessment and implementation formats leading to limited evaluation tools and large gaps between evaluation results and trainees' actual thinking levels [**?**, **?**]. This course differed from traditional evaluations by integrating subjective assessment with quantitative evaluation to conduct mixed assessments of teaching effectiveness, enriching the GP thinking evaluation system. Contrary to the stereotype that subjective evaluation lacks reliability, this evaluation used multi-source feedback (MSF) combining self-assessment, instructor-trainee bidirectional evaluation, peer assessment, and SP assessment, conforming to triangulation principles and yielding conclusions consistent with VAN that subjective evaluation is equally reliable when assessing advanced abstract competencies [**?**]. Therefore, issues like theoretical exams being unable to measure skill levels or "gastric tube insertion" operations failing to reflect physical examination competency—problems of item sampling—can be addressed through PA' s mixed evaluation, which completely reflects trainees' GP thinking. Quantitative evaluation provides direct assessment, while subjective evaluation provides supplementary and lateral assessment.

3.2 "Flipped Classroom" as PA' s Presentation Format

Unlike traditional teaching, the "flipped classroom" positions trainees as the main actors, with instructors and peers serving as learning guides and facilitators to help learners internalize knowledge, emphasizing self-directed learning, reflective communication, and collaboration [**?**]. This course' s participants were general practice track master's students with certain professional foundations, so instructors primarily served as guides rather than basic theory lecturers, helping trainees construct their own GP thinking. History-taking and physical examination in the core module are "basic skills" for general practitioners' consultation abilities, laying the foundation for subsequently "connecting" various competencies in GP thinking. The core module focused on skill-based content, so integrating practice allowed trainees to improve rapidly. However, the practice emphasis was not on skill proficiency but on the "thinking" behind skill application, hence case-based practice. To prevent trainees from failing to achieve expected outcomes due to insufficient ability, assistant roles were created where other trainees could help solve problems, serving both as evaluation feedback and training. The case module used representative real general practice cases and was the primary pathway for improving "connection" abilities. From PA theory perspective, it was also the main process for constructing "connection"

abilities. The case module employed PBL format but used formative longitudinal evaluation—assessing after each course content to observe trainees' GP thinking development throughout the module. In PA design, formative evaluation feedback needs adjustment to promote performance in subsequent course content. Finally, case presentations reflected the outcomes of "connection"ability construction. This module also required feedback integration, as only through continuous dynamic evaluation and feedback could trainees gradually develop active learning awareness and construct GP thinking from existing foundational knowledge. The course fully adhered to PA' s five principles— "meaningful learning," "longitudinal assessment," "continuous feedback," "learner-centeredness and self-responsibility," and "assessment for learning and as learning" —ultimately presenting "cases, practice, discussion, and feedback" through the "flipped classroom" format.

3.3 PA Promotes Sequential Construction of GP Thinking

In PA, no single assessment result determines trainees' final outcomes, making individual assessment results "low-stakes" decisions [**?**, **?**], while the collective results of all quantitative and subjective evaluations constitute "high-stakes" decisions. Trainee heterogeneity causing quantitative score differences has been confirmed in previous studies [**?**, **?**]. In this study, the two groups showed no difference in first-session scores for core and case modules, but differences emerged in final-session scores. This may relate to PA decision "stakes." Previously, in-service trainees outperformed residency trainees in both core and case modules before and after courses, primarily due to practical experience. However, from teaching format and course adaptation perspectives, both groups faced unknown situations, resulting in no initial performance differences. As the course progressed, in-service trainees' experience advantages gradually dominated, leading to superior performance across all modules.

Comparisons of quantitative evaluation scores between first and final sessions for core and case modules, along with the number of positive and constructive evaluations across sessions, demonstrated gradual trainee improvement, indicating that PA can promote trainees' gradual internalization of existing knowledge and construction of GP thinking after general practice concept infusion. Previous GP thinking training modes included group discussion, apprenticeship, theoretical teaching, and outpatient instruction [**?**, **?**]. Different teaching modes suit different competency development [**?**], but no single mode can comprehensively improve all trainee abilities. GP thinking is based on theoretical knowledge and clinical skills yet requires "connecting" and flexibly applying theory and skills, making it abstract and multifaceted, requiring integrated teaching modes. In this PA practice, baseline pre-class surveys, in-class formative evaluation, and post-class summative evaluation were combined to ensure course quality. PBL and CBL design, including the innovative assistant role in CBL practice to help trainees, followed PA' s "assessment for learning and as learning" principle.

Feedback is key to constraining course quality, with current issues including

missing "before" and "during" feedback and lack of trainee knowledge achievement [?]. This course' s feedback differed from traditional formative evaluation feedback by not only promoting trainee reflection but also using PA to continue evaluating trainees' improvement on similar issues in subsequent courses, thereby driving learning. Course feedback achieved "during-course" feedback through MSF while also feeding back trainee knowledge achievement to promote continuous course improvement and quality enhancement. Therefore, analyzing subjective evaluation quantities revealed that as the course progressed, positive evaluations for core and case modules continuously increased while constructive evaluations continuously decreased. In PA-guided "flipped classrooms," trainees received feedback through instructor-trainee interaction, trainee practice, and collaboration, ultimately constructing GP thinking based on their own professional knowledge and skills.

This course followed PA and Miller' s Pyramid theoretical principles to optimize evaluation of the required course *Clinical Reasoning in General Practice* for Tongji University' s general practice track master' s students. By introducing PA and conducting teaching practice, we found that PA can enrich the GP thinking evaluation system, promote sequential construction of GP thinking, and that "flipped classroom" serves as an effective presentation format for PA.

This study has limitations: subjective evaluation content could not be fully presented due to space constraints. Future work will further improve this and re-evaluate PA application effectiveness.

Author Contributions: ZHAI Jiayi was responsible for conceptualization, design, feasibility analysis, and manuscript writing. LU Yuan was responsible for English revision, quality control, and manuscript review. SHI Jianjun was responsible for implementation. YU Dehua was responsible for manuscript revision and supervision.

Conflict of Interest: The authors declare no conflict of interest.

ORCID iDs:
ZHAI Jiayi: https://orcid.org/0009-0007-3891-435X
LU Yuan: https://orcid.org/0000-0002-4360-8476
SHI Jianjun: https://orcid.org/0009-0003-3117-779X
YU Dehua: https://orcid.org/0000-0001-7652-938X

References

[?] ZHAI Jiayi, LU Yuan, QIAN Shulei, et al. Curriculum design of clinical reasoning for general practice master's students at Tongji University [?]. Chinese General Practice. DOI:10.12114/j.issn.1007-9572.2022.0715.

[?] ZHAI Jiayi, LU Yuan, GE Xuhua, et al. Competency-based curriculum design and implementation of clinical reasoning in general practice [?]. Chinese Journal of Medical Education, 2022, 42(12): 1079-1082. DOI:10.3760/cma.j.cn115259-20220502-00566.

[?] YI Jiang, WEI Tao, YU Zhiyuan, et al. Preliminary exploration of clinical reasoning training system establishment in standardized general practice residency training [?]. Chinese Postgraduate Medical Education, 2020, 4(4): 311-314, 324. DOI:10.3969/j.issn.2096-4293.2020.04.006.

[?] LU Yuan, YU Dehua, ZHANG Bin, et al. Practical application of PBL teaching model in standardized general practice training [?]. Chinese General Practice, 2014, 17(16): 1880-1883. DOI:10.3969/j.issn.1007-9572.2014.16.019.

[?] ZHAN Jingyan, LOU Jingqiu, WANG Shaokun, et al. Research on teaching models and application effects in China' s standardized general practice residency training [?]. Chinese General Practice, 2021, 24(19): 2401-2407. DOI:10.12114/j.issn.1007-9572.2021.00.201.

[?] VAN DER VLEUTEN C, VERHOEVEN B. In-training assessment developments in postgraduate education in Europe [?]. ANZ J Surg, 2013, 83(6): 454-459. DOI:10.1111/ans.12190.

[?] VAN DER VLEUTEN C P M, SCHUWIRTH L W T. Assessing professional competence: from methods to programmes [?]. Med Educ, 2005, 39(3): 309-317. DOI:10.1111/j.1365-2929.2005.02094.x.

[?] MILLER G E. The assessment of clinical skills/competence/performance [?]. Acad Med, 1990, 65(9 Suppl): S63-S67. DOI:10.1097/00001888-199009000-00045.

[?] VAN DER VLEUTEN C P M, DANNEFER E F. Towards a systems approach to assessment [?]. Med Teach, 2012, 34(3): 185-186. DOI:10.3109/0142159X.2012.652240.

[?] WEN Deliang. Formative assessment methods in medical education [?]. Medical Teaching in Higher Education: Electronic Edition, 2011, 1(1): 36-38.

[?] TORRE D M, SCHUWIRTH L W T, VAN DER VLEUTEN C P M. Theoretical considerations on programmatic assessment [?]. Med Teach, 2020, 42(2): 213-220. DOI:10.1080/0142159X.2019.1672863.

[?] VAN DER VLEUTEN C P, SCHUWIRTH L W, SCHEELE F, et al. The assessment of professional competence: building blocks for theory development [?]. Best Pract Res Clin Obstet Gynaecol, 2010, 24(6): 703-719. DOI:10.1016/j.bpobgyn.2010.04.001.

[?] VAN DER VLEUTEN C P, SCHUWIRTH L W, DRIESSEN E W, et al. A model for programmatic assessment fit for purpose [?]. Med Teach, 2012, 34(3): 205-214. DOI:10.3109/0142159X.2012.652239.

[?] VAN DER VLEUTEN C P M, SCHUWIRTH L W T, DRIESSEN E W, et al. Twelve Tips for programmatic assessment [?]. Med Teach, 2015, 37(7): 641-646. DOI:10.3109/0142159X.2014.973388.

[?] SCHUWIRTH L W T, VAN DER VLEUTEN C P M. A history of assessment

in medical education [?]. Adv Health Sci Educ Theory Pract, 2020, 25(5): 1045-1056. DOI:10.1007/s10459-020-09973-9.

[?] JOHN A. DENT, RONALD M. HARDEN, DAN HUNT. A Practical Guide for Medical Teachers (6th Edition) [?]. WANG Weimin et al., trans. Beijing: Peking University Medical Press, 2020: 282-289.

[?] MONICA VAN DE RIDDER J M, STOKKING K M, MCGAghie W C, et al. What is feedback in clinical education? [?]. Med Educ, 2008, 42(2): 189-197. DOI:10.1111/j.1365-2923.2007.02973.x.

[?] ZHANG Jinlei, WANG Ying, ZHANG Baohui. Research on flipped classroom teaching model [?]. Distance Education Journal, 2012, 30(4): 46-51. DOI:10.15881/j.cnki.cn33-1304/g4.2012.04.008.

[?] DING Caicheng, CHEN Binghe. Diagnosis and improvement strategies for curriculum quality in higher vocational colleges [?]. Vocational and Technical Education, 2017, 38(8): 55-58. DOI:10.3969/j.issn.1008-3219.2017.08.019.

(Received: November 14, 2023; Revised: May 14, 2024)
(Editor: CHENG Sheng)

Note: Figure translations are in progress. See original paper for figures.