Research in Social and Administrative Pharmacy 17 (2021) 1645–1649

Contents lists available at ScienceDirect

Research in Social and Administrative Pharmacy

journal homepage: www.elsevier.com/locate/rsap

Cognitive load theory: Implications for assessment in pharmacy education

Kyle John Wilby [a,*], Bridget Paravattil [b]

[a] School of Pharmacy, University of Otago, PO Box 56, Dunedin, 9054, New Zealand
[b] College of Pharmacy, Qatar University, PO Box 2713, Doha, Qatar

ARTICLE INFO

Keywords
Mental workload
Cognitive load
Assessment

ABSTRACT

The concept of mental workload is well studied from a learner's perspective but has yet to be better understood from the perspective of an assessor. Mental workload is largely associated with cognitive load theory, which describes three different types of load. Intrinsic load deals with the *complexity* of the task, extraneous load describes *distractors* to the task at hand, and germane load focuses on the development of *schemas* in working memory for future recall. Studies from medical education show that all three types of load are relevant when considering rater –based assessment (e.g. Objective Structured Clinical Examinations (OSCEs), or experiential training). Assessments with high intrinsic and extraneous load may interfere with assessors' attention and working memory and result in poorer quality assessment. Reducing these loads within assessment tasks should therefore be a priority for pharmacy educators. This commentary aims to provide a theoretical overview of mental workload in assessment, outline research findings from the medical education context, and propose strategies to be considered for reducing mental workload in rater-based assessments relevant to pharmacy education. Suggestions for future research are also addressed.

It is Friday afternoon (1pm) and you are frantically trying to organize a patient discharge that requires communication with the patient's community pharmacy. You also need to hand over all monitoring for the weekend clinical staff, including the complicated patient on aminoglycosides with the serious pseudomonal abscess. As you are searching for the right pharmacy to phone, you feel a tap on your shoulder and it is your final-year pharmacy student wanting to know at what time you would like to go over their evaluation (that you have been meaning to start all week …). You look at your watch and tell them to meet you in the meeting room at 2:30pm (this should give you enough time, right?).

Fast-forward to 2:15pm and you have managed the patient discharge and handover. You sit down at your computer to log into the student evaluation system and open up the form. A document consisting of over 40 rateable items is presented before you. As you begin to move through the items, you wonder when these evaluations got so complicated? Eight items alone for professionalism? You think to yourself, 'How do I distinguish between these?' and 'I should have paid more attention to remember some examples'. Your student was 'good' and largely independent, so you begin to just move through the form and give passing scores while thinking, 'I'll be more organized next time'.

The concept of 'workload' is a common topic of discussion in pharmacy education.[1,2] Preceptors and academic staff are attempting to cope in a world with increasing demands of time, energy, attention, and patience. While 'workload' is often conceptualized as tasks or activities required by an individual to complete, the concept of mental workload rarely enters our conversations. Mental workload refers to the mental effort and cognitive processing demands required to complete specific tasks or activities.[3] It is a function of attention and memory and can be influenced by multitude of factors. Specific theories, such as cognitive load theory, have attempted to explain the mental workload associated with learning.[4,5] Studies largely show that learning is a complex process and must account for the mental workload associated with activities and tasks, in order to be effective.[6] More recently, however, research has begun to explore the concept of *assessors'* mental workload associated with rating tasks and the mental effort required by clinical educators when completing student performance assessments.[7–12]

Despite an increase in the literature relating to mental workload and assessment in medical education, the topic is not yet explored in pharmacy education. Although findings from medical education are likely relevant to pharmacy, assessment structures (especially within experiential training) differ, which may influence considerations pertaining to mental workload.[13] Future scholarly activity within pharmacy may therefore complement existing knowledge and lead to a greater understanding of the influence of mental workload on assessment across

* Corresponding author.
E-mail addresses: kyle.wilby@otago.ac.nz (K.J. Wilby), bridgetj@qu.edu.qa (B. Paravattil).

https://doi.org/10.1016/j.sapharm.2020.12.009
Received 15 December 2019; Received in revised form 9 November 2020; Accepted 15 December 2020
Available online 22 December 2020

disciplines. This commentary aims to introduce the concept of mental workload, provide an overview of the central theoretical perspective, identify potential strategies to reduce mental workload, and highlight implications for future research and practice. The authors of this commentary both work in administrative positions that require development and evaluation of assessment procedures for simulation and experiential education. Based on feedback from academic staff and clinical preceptors relating to the complexity of student assessment instruments, the authors began to explore the concept of mental workload in assessment.

Current trends in assessment

Schuwirth and van der Vleuten (2020) published a history of assessment in medical education and the trends noted in this paper align with what has been seen in pharmacy education.[14] The advent of competency-based education shifted assessments from strictly practices of measurement to those of judgement. This shift in thought resulted in development of newer assessment approaches, namely OSCEs and assessments embedded within experiential training settings. These assessments required the assessor to have a more central role (resulting in greater mental workload) in making judgements and exposed concerns that differences in assessors' judgements could have great implications on student progression, unless complemented with multiple assessments across different contexts.[14] These concerns have resulted in re-conceptualizing assessment as a 'system', where efforts throughout the whole course (or program) contribute to progression decisions, rather than being informed by one or few high stakes assessments.

The introduction of Entrustable Professional Activities (EPAs) as an assessment tool in experiential settings is an example of a successful way to help reduce some of the problems encountered in the opening example.[15] EPAs assessed throughout a 4-week rotation, for example, may help alleviate concerns of the preceptor with respect to attention and memory. The use of EPAs in both medicine and pharmacy have shown improved psychometric properties, as compared to direct observation alone, and are generally perceived favorably by students and preceptors.[13,14] Despite these benefits, however, EPAs can be difficult to interpret and the optimal number of EPAs to be assessed in experiential settings (including the number of times each should be assessed) is not yet known. As such, EPAs (in addition to the other competency-based assessments) may also be prone to concerns regarding mental workload and implications of these are highlighted below.

The concept of mental workload

The primary theory driving research related to mental workload is cognitive load theory.[5,16] According to this theory, mental workload consists of intrinsic, extraneous, and germane load.[16,17] Intrinsic load is the complexity of the task, which is a function of the of association between a learner's expertise and the nature of the task. Task complexity is related to the number of elements and the degree of element interactivity necessary to be learned concurrently by the working memory.[18] Element interactivity is the extent to which information can be learned in isolation of each other.[5] High element interactivity occurs when multiple pieces of information are needed to be processed at the same time to learn a task versus low element interactivity occurs when the information can be learned in isolation. Take, for example, the concept of an escape room-based learning task. This could likely be considered a complex learning task but learners' mental workload may vary based on their experience and/or expertise in completing previous escape rooms and solving puzzles. There may also be high interactivity if participants must solve multiple puzzles or piece together different types of information to solve a single task. From the perspective of an assessor, intrinsic load still refers to complexity but now is related to the complexity of the assessment task. Tasks with a high level of element interactivity that require assessors to disentangle multiple components of a learner's performance tend to have higher associated intrinsic load. Examples of these tasks may include experiential training assessments with multiple rateable items across different competency domains, extensive checklists/rubrics for OSCEs or oral presentations, the requirement to assess multiple learners at the same time, or OSCE stations that are blueprinted to have multiple layers of complexity (complex patient characteristics and/or complex problems to solve).

Extraneous load is defined as the load that is imposed due to poor instructional design or other factors that divert the attention away from the learning environment.[16] Extraneous load is commonly considered to be a 'distraction'. This may include times when task instructions are not clear, or perhaps if factors such as noise or conversations interrupt the learner's attention to the task at hand.[3] The same examples can be used for assessment, including poor instructions or suboptimal assessment environments. Other sources of extraneous load may occur when assessors are given an assessment tool that is not clear or when asked to perform secondary assessment tasks, such as assessing the performance of a simulated actor or being responsible for operating timers or audiovisual aids during an assessment.

Germane load refers to the mental effort an individual must dedicate to create a permanent store of knowledge.[3,5,17] It is described as the cognitive effort for learning and differs from the 'interfering' nature of both intrinsic and extraneous load. It is the processing, construction, and automation of 'schemas'. Schemas are described as a mental structure of preconceived ideas, a framework for categorizing knowledge, or a system for organizing and conceptualizing new information. Contrary to intrinsic and extraneous load, germane load is viewed positively and one that may facilitate formation of long-term memories for easy future recall. Consider once again the concept of an escape room. The first time an individual enters the escape room can be daunting, as they may not know what to expect. After a first experience, however, a 'schema' may be developed in their mental structure that represents the concepts and processes associated with an escape room for memory retrieval at a later date. Future exposures may draw on the schema through attention to integrate new information or reprocessing contradictory information. With respect to assessment, training, receiving feedback, or stimulated self-reflection may all contribute to increases in germane load and schema formation.

To fully grasp cognitive load theory, it is important to comprehend that intrinsic and extraneous load are additive.[19] When intrinsic load is low and extraneous load is high (or vice versa) but kept within the working memory limits, impairment to working memory may not be as detrimental compared to when both total intrinsic and extraneous loads exceed the working memory limit. It is crucial, therefore, to ensure the loads imposed on learners and assessors are within working memory limits to avoid overload and burnout. With repeated exposure or with more experience, it is hoped over time that the learner or assessor will have developed a schema from the working memory and be able to better manage increases in loads.[5]

Measuring mental workload

Measurement of mental workload has been attempted using a variety of ways. Mental workload can be measured subjectively using the NASA TLX (National Aeronautics and Space Administration Task Load Index) instrument.[20] The NASA TLX subjective workload questionnaire is a subjective, self-reported, multidimensional assessment that rates the participant's perceived workload on a given task.[20,21] Other subjective instruments also are available to measure workload.[21] Mental workload can also be measured using performance data from an objective secondary task.[3,21] The most commonly reported secondary task is the use of a wireless vibrotactile device attached to the arm of participants. The device vibrates at random intervals between 10 and 90 seconds over the course of the primary task participants were performing (e.g. assessing student performance). Subjects must press the button on the device to cease it from vibrating whenever the device randomly alerted but are

also instructed to prioritize their primary task first. The time it takes the participant to turn off the vibration is measured as a marker of the mental workload of the primary task (e.g. greater amounts of time meaning greater workload on the primary task). Another secondary task measure may include the quantity and quality of narrative comments written during primary rating tasks (e.g. the more and higher quality comments should correlate with lower mental workload of rating).

Mental workload and rater-based assessment

Mental workload and rater-based assessment is an emerging concept within medical education but has not appeared to be discussed or researched in the context of pharmacy. As pharmacy programs around the world are increasingly developing performance-based assessments such as Objective Structured Clinical Examinations (OSCEs) and placing greater emphasis on assessment from experiential training, the concept is very relevant to the academy. In particular, the current trend of competency framework development has led to creation of rater-based assessment tools that are meant to capture a trainee's competency across of a variety of domains and rateable items.[22] Although these assessments and tools may be anchored within the competencies we seek to assess, research from medical education suggests that those that require high mental workload may produce ratings of poor quality (usually measured through perceived accuracy, inter-rater reliability, and quantity/quality of rich data collected through narrative comments).[23,24] If the same holds true for assessments and tools used in pharmacy education, validity and reliability of rater-based assessment exercises could be threatened.

The following text and tables outline load influencers and associated strategies that may be useful at optimizing mental workload across two examples (OSCEs and experiential training). These influencers and strategies may also serve as a starting point for investigators seeking to research the impact of mental workload on rater-based assessments in pharmacy. The strategies are focused on intrinsic and extraneous load, as these are considered the modifiable loads to support assessors' attention and working memory.[3] Germane load, as previously discussed, is perceived as a positive load and it can be speculated that reducing the other types of cognitive load may be beneficial for germane load and the formation of schemas and long-term memory.[16]

Table 1 provides examples of load influencers and suggested strategies to optimize load for intrinsic and extraneous load associated with OSCEs. Intrinsic load (i.e. complexity) in this assessment context may present as a result of the task, assessor's experience, assessment tool, among others.[3,23] Table 1 provides proposed strategies for overcoming these load influencers with the goal of improving assessment quality. To date, research primarily shows that focusing assessors' attention in OSCEs (as demonstrated through videotaped interactions) towards fewer competencies (e.g. two vs. six or seven) may offer the most benefit in terms of inter-rater reliability and provision of rich performance feedback.[11,12] Simply reducing the number of rateable items (without reducing the number of competencies assessed) did not offer benefit in one study.[25] It has been purported, however, that removing checklists and replacing with guidance points and an overall global assessment may improve inter-rater reliability and internal consistency, resulting in the need for fewer stations and assessors.[26] The mental workload associated with this intervention, however, has not yet been studied. The implications of these approaches, including the impact on mental workload and provision of rich and meaningful feedback should be further studied. Further research should also investigate how the approach to OSCE conduction (e.g. total assessor time, number of days required for assessment, live versus videotaped assessment) may influence mental workload and assessment quality.

Table 2 provides examples of load influencers and suggested strategies to optimize load for intrinsic and extraneous load associated with experiential training placements. Intrinsic load in this setting has a higher level of element interactivity as the assessor is responsible for both precepting and assessing the learner. Precepting itself is a complex task that includes teaching, demonstration, observation, feedback, and assessment.[27] Precepting and assessing can therefore not be done in isolation of each other and this may increase the complexity of the assessment task. Extraneous load within this setting is typically high, as the assessor encounters numerous simultaneous distractors that may take their attention away from precepting and assessment. Examples of these distractors may include patient care responsibilities, requests from other healthcare providers, administration duties, and learner interactions, Similar to an OSCE setting, the complexity of the assessment tool and provision of assessment instructions may also contribute to overall load. Proposed strategies for managing these influencers are also provided in Table 2. It should be noted, however, that no research has been identified to provide confirming or disconfirming evidence that these strategies may be beneficial in improving assessment quality in the experiential context. Authors of published studies using videotaped interactions hypothesize that the benefits of focusing the assessor on fewer competency domains when assessing could be extrapolated to experiential settings but this should remain a priority for future research.[12]

Table 1
Example load influencers and proposed strategies to optimize assessors' mental workload within an OSCE context.

Type of Load	Load Influencers	Strategies
Intrinsic	• Complex tasks to assess (e.g. advanced communication/ empathy/motivational interviewing) • Unfamiliarity with topic area before assessment • High numbers of rateable items • Focus on multiple competencies within one interaction (e.g. patient care, communication, and professionalism)	• Assessor training and validation exercises prior to OSCE implementation to decrease perceived complexity or unknown nature of tasks • Ensure rateable items are of a reasonable number and are clearly stated (e.g. piloting assessment tools to investigate usability) • Blueprint/focus stations in terms of competency and/or complexity (e.g. communication, ethics, patient assessment, simple versus complex patients and problems)
Extraneous	• Unclear assessment instructions that distract from assessment procedures (e.g. tasks required to be completed by assessor) • Unfamiliar venue, travel delays, pressing work or family commitments • Poorly worded assessment tools (e.g. rubrics or checklists) • Managing timing device, electronic resource, or paper-based references that may distract focus from the interaction being assessed	• Provide clear instructions both written and verbally; consider pilot or role play as part of training/orientation exercises • Scheduling of breaks for assessors between students or student groups • Peer review of assessment tools with target users; use of lay language for writing of assessment tools • Use of consistent assessment tools across similar assessments or similar EPA/ competency assessments (e.g. design, wording, instructions) • Reduce extraneous tasks assessor must complete in addition to rating tasks (e.g. timing, clearing history of electronic resources, directing student flow)
Germane	• Self reflection on assessment experience	• Pair experienced assessor with inexperienced assessor • Provide practical training sessions in same format/ context of actual exam • Provision of feedback and/or audit data

EPA = entrustable professional activity.

Table 2
Example load influencers and proposed strategies to optimize assessors' mental workload within an experiential training context.

Type of Load	Load Influencers	Strategies
Intrinsic	• Number of learners per rotation to evaluate (e.g. one versus two or more) • Numerous rateable items across multiple competencies for evaluation • Recalling student-specific scenarios and examples for feedback and evaluation (e.g. dependent on rotation length) • Engaging in active learning strategies during topic/patient discussions • Managing learner's cognitive load when assigning tasks and assignments	• Negotiate manageable number of learners per practice experience • Reduce complexity of assessment tools (e.g. number of rateable items, number of competencies, number of EPAs) • Take notes of student examples during practice experiences; use technology (e.g. apps to capture real-time examples) • Provide preceptors with a framework on how to seek the objective in assessment (e.g. base assessment on actions versus personal relationship bias) • Seeking learner feedback before and during the rotation to tailor activities accordingly
Extraneous	• Preceptor workload in practice site (e.g. number of patients, other activities outside of patient care) • Accountability to patients and other team members • Learner dynamics that distract from the observation (e.g. poor academic performance, unprofessional attitude) • Poorly designed evaluation tools (e.g. unclear or repetitive)	• Schedule set times for student discussions outside of work duties (e.g. put into a calendar/agenda) • Provide detailed examples/ case studies on how to manage learner dynamics as part of orientation/training purposes • Use of consistent assessment tools across contexts or similar competency/EPA assessments (e.g. design, wording, instructions); peer review with target users
Germane	• Self reflection on practice experience	• Review learner's feedback • Peer observation in practice • Engage in preceptor development activities • Provision of program feedback and/or audit data

EPA = entrustable professional activity.

The way forward

It is common to read conclusions from pharmacy education studies that call for increased assessor training, anchoring, or calibration when assessment markers (e.g. inter-rater reliability) are determined to be suboptimal. While this may present as the logical solution, the literature shows that despite our best efforts to train assessors and standardize expectations, these exercises likely have little value.[23,28] It is therefore time we supplement with a different approach and instead of attempting to shape the assessor to the assessment, we must rather begin to shape our assessments to the assessor.[23] The concept of mental workload is just one example that may be relevant and worth exploring to help improve assessment quality that may complement or enhance assessor training initiatives. Research shows that other areas, such as differences in how assessors conceptualize competencies and how assessors' own idiosyncrasies and expertise may influence assessment, could also contribute to a perceived lack of quality in rater-based assessment tasks.[29–31] Educational researchers should therefore be encouraged to consider these factors when designing rater-based studies and interpreting assessment results.

From a practical perspective, evidence to date suggests that the instructions and tools we provide to assessors should consider the concept of mental workload. It is known, for example, that when assessors are given multiple competencies to assess, they engage in strategies to overcome the associated higher mental workload.[11] Prioritization and selection are tactics often used to focus only on domains that assessors felt were the most important. Simplification, another approach, occurs when the assessor focuses on behaviors or domains that were the easiest and most obvious to recognize. Until more is known about the most effective ways to optimize mental workload to promote high quality assessment, these considerations dictate that we must strive to provide clear assessment instructions, focus assessors on specific competencies (where applicable), and ensure assessment tasks are not overly complicated by complexity (i.e. intrinsic load) or distractors (i.e. extraneous load).

The relevance of this topic to pharmacy education warrants future research that more clearly addresses the links between mental workload and assessment. Specifically, studies should investigate the effectiveness of the strategies outlined in this paper on reducing workload and increasing assessment quality. Given more data in this area, strategies could be prioritized according to impact. Future research should also seek to investigate the influence of mental workload on other assessments, such as critical reflection and portfolios. Finally, it should be determined if efforts made to optimize the mental workload associated with assessment tasks improves assessor acceptance of the assessment task and overall engagement in the learning process (e.g. propensity to precept more students, return as an OSCE assessor, etc.).

Summary

The preceptor in the opening example was likely experiencing high mental workload associated with his or her assessment task. Intrinsic load may be high due to the complexity of the form, the number of rateable items, or the competencies that were required to be assessed. The length of the rotation may have influenced his or her working memory and challenged the ability to remember student-specific examples. The preceptor's regular clinical duties and patient care pressures could be considered extraneous load as distractors and may have influenced the preceptor's attention away from assessment responsibilities. Moving forward, the academy must consider these pressures when designing assessment structures and strive to determine evidence-based solutions to improve rater-based assessment by optimizing mental workload in assessment design.

Funding

This commentary did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

None.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sapharm.2020.12.009.

References

1. Danielson J, Craddick K, Eccles D, Kwasnik A, O'Sullivan TA. A qualitative analysis of common concerns about challenges facing pharmacy experiential education programs. *Am J Pharmaceut Educ*. 2015;71, 06.
2. Wilborn TW, Timpe EM, Wu-Pong S, et al. Factors influencing faculty perceptions of teaching workload. *Curr Pharm Teach Learn*. 2013;5:9–13.

3. Tavares W, Eva KW. Exploring the impact of mental workload on rater-based assessments. *Adv Health Sci Educ*. 2013;18:291–303.
4. Lord RG, Maher KW. Alternative information-processing models and their implications for theory, research, and practice. *Acad Manag Rev*. 1990;15:9–28.
5. Van Merrienboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;44:85–93.
6. Van Merrienboer JJG, Ayres P. Research on cognitive load theory and its design implications for e-learning. *Educ Technol Res Dev*. 2005;53:5–13.
7. Byrne A, Tweed N, Halligan C. A pilot study of the mental workload of objective structured clinical examination examiners. *Med Educ*. 2014;48:262–267.
8. Byrne A, Soskova T, Dawkins J, Coombes L. A pilot study of marking accuracy and mental workload as measures of OSCE examiner performance. *BMC Med Educ*. 2016;16:191.
9. Wood TJ. Mental workload as a tool for understanding dual processes in rater-based assessments. *Adv Health Sci Educ Theory Pract*. 2013;18:523–525.
10. Tavares W, Eva KW. Impact of rating demands on rater-based assessments of clinical competence. *Educ Prim Care*. 2014;25:308–318.
11. Tavares W, Ginsburg S, Eva KW. Selecting and simplifying: rater performance and behavior when considering multiple competencies. *Teach Learn Med*. 2016;28:41–51.
12. Tavares W, Sadowski A, Eva KW. Asking for less and getting more: the impact of broadening a rater's focus in formative assessment. *Acad Med*. 2018;93:1584–1590.
13. Croft H, Gilligan C, Rasiah R, Levett-Jones T, Schneider J. Current trends and opportunities for competency assessment in pharmacy education – a literature review. *Pharmacy*. 2019;7:67.
14. Schuwirth LWT, van der Vleuten CPM. A history of assessment in medical education. *Adv Health Sci Educ*. 2020. https://doi.org/10.1007/s10459-020-10003-0.
15. Ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ*. 2005;39:1176–1177.
16. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive architecture and instructional design. *Educ Psychol Rev*. 1998;10:251–296.
17. Debue N, van de Leemput C. What does germane load mean? An empirical contribution to the cognitive load theory. *Front Psychol*. 2014;5:1099.
18. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE guide No. 86. *Med Teach*. 2014;36:371–384.
19. Carlson R, Chandler P, Sweller J. Learning and understanding science instructional material. *J Educ Psychol*. 2003;95:629–640.
20. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv Psychol*. 1988;52:139–183.
21. Byrne A. Measurement of mental workload in clinical medicine: a review study. *Anesth Pain*. 2011;1:90–94.
22. Koster A, Schalekamp T, Meijerman I. Implementation of competency-based pharmacy education (CBPE). *Pharmacy (Basel)*. 2017;5:10.
23. Eva K. Cognitive influences on complex performance assessment: lessons from the interplay between medicine and psychology. *J App Res Mem Cogn*. 2018;7:177–188.
24. Paravattil B, Wilby KJ. Optimizing assessors' mental workload in rater-based assessment: a critical narrative review. *Perspect Med Educ*. 2019;8:339–345.
25. Hurley KF, Giffin NA, Stewart SA, Bullock GB. Probing the effect of OSCE checklist length on inter-observer reliability and observer accuracy. *Med Educ Online*. 2015;20:29242.
26. Wilkinson T. Hot not to put the O into an OSCE. *Perspect Med Educ*. 2018;7(Suppl 1):28–29.
27. Worrall CL, Aistrope DS, Cardello EA, et al. Priming the preceptor pipeline: collaboration, resources, and recognition. *Am J Pharmaceut Educ*. 2016;80:S19.
28. Ginsburg S, McIlroy J, Oulanova O, Eva K, Regehr G. Toward authentic clinical evaluation: pitfalls in the pursuit of competency. *Acad Med*. 2010;85:780–786.
29. Gingerich A, Ramlo SE, van der Vleuten CPM, Eva KW, Regehr G. Inter-rater variability as mutual disagreement: identifying raters' divergent points of view. *Adv Health Sci Educ*. 2017;22:819–838.
30. Berendonk C, Stalmeijer R, Schuwirth LWT. Expertise in performance assessment: assessors' perspectives. *Adv Health Sci Educ*. 2013;18:559–571.
31. Wilby KJ, Dolmans HJM, Austin Z, Govaerts MJB. Assessors' interpretations of narrative data on communication skills in a summative OSCE. *Med Educ*. 2019;53:1003–1012.