Semantic competence and prototypical verbalizations are associated with higher OSCE and global medical degree scores: a multi-theory pilot study on year 6 medical student verbalizations

Pedro Grilo Diogo*, Vítor Hugo Pereira, Frank Papa, Cees van der Vleuten, Steven J. Durning and Nuno Sousa

Semantic competence and prototypical verbalizations are associated with higher OSCE and global medical degree scores: a multi-theory pilot study on year 6 medical student verbalizations

https://doi.org/10.1515/dx-2021-0048
Received April 2, 2021; accepted February 20, 2023;
published online March 15, 2023

Abstract

Objectives: The organization of medical knowledge is reflected in language and can be studied from the viewpoints of semantics and prototype theory. The purpose of this study is to analyze student verbalizations during an Objective Structured Clinical Examination (OSCE) and correlate them with test scores and final medical degree (MD) scores. We hypothesize that students whose verbalizations are semantically richer and closer to the disease prototype will show better academic performance.

Methods: We conducted a single-center study during a year 6 (Y6) high-stakes OSCE where one probing intervention was included at the end of the exam to capture students' reasoning about one of the clinical cases. Verbalizations were transcribed and coded. An assessment panel categorized verbalizations regarding their semantic value (Weak, Good, Strong). Semantic categories and prototypical elements were compared with OSCE, case-based exam and global MD scores.

Results: Students with Semantic 'Strong' verbalizations displayed higher OSCE, case-based exam and MD scores, while the use of prototypical elements was associated with higher OSCE and MD scores.

Conclusions: Semantic competence and verbalizations matching the disease prototype may identify students with better organization of medical knowledge. This work provides empirical groundwork for future research on language analysis to support assessment decisions.

Keywords: assessment; clinical reasoning; dual process theory; knowledge structures; OSCE; prototype theory; semantics

Introduction

One of the facets of expert performance in Medicine is the quality of an individual's clinical reasoning. In the past decades, research on clinical reasoning has had a cognitive emphasis, with structural theories of medical expertise arising from this focus on the organization of knowledge (e.g., how information is believed to be stored and intertwined) on physician's brains [1–3]. Medical knowledge is believed to be organized as knowledge structures (KS), which are domain-specific constructs in long-term memory that are created and refined by cumulative clinical experience, allowing physicians to process multiple elements of clinical information [3–5]. Research has shown that KS are critical features of expert performance in several domains [6–8].

Medical students' KS can be analyzed from their verbalizations about clinical cases. In fact, think-aloud protocols are commonly used in the study of clinical reasoning [7, 9] and student verbalizations have been used for decades in everyday educational practice to infer about their competence during oral exams or case-based discussions. However, the value of student verbalizations to the field of assessment is limited by the fact that assessment methodologies often rely on subjective assessor opinions on the quality of verbalizations instead of

***Corresponding author: Pedro Grilo Diogo**, MD, MHPE, Escola de Medicina da Universidade do Minho, Campus de Gualtar, Braga, 4710-057, Portugal, E-mail: pgrilodiogo@gmail.com. https://orcid.org/0000-0002-1028-4890

Vítor Hugo Pereira and Nuno Sousa, Escola de Medicina da Universidade do Minho, Braga, Portugal

Frank Papa, University of North Texas Health Science Center, Fort Worth, TX, USA

Cees van der Vleuten, School of Health Professions Education, Maastricht, The Netherlands

Steven J. Durning, Center for Health Professions Education, Uniformed Services University, Bethesda, MD, USA

being grounded on established theories about knowledge organization.

In the field of Medical Education, semantics and prototype theory offer different viewpoints to study knowledge organization. The analysis of student verbalizations in the light of these frameworks opens new possibilities in the field of assessment.

Lexical semantics studies the meaning of words [10]. In Medicine, discourse includes several words describing patient signs, symptoms and whole clinical presentations. These words include encapsulations and semantic qualifiers. Encapsulations are high-order clinical terms with the same explanatory value than several small-order elements (e.g. heart failure encapsulates dyspnea on exertion, orthopnea and peripheral edemas). Semantic qualifiers are abstract and often binary concepts that organize and provide meaning to clinical presentations (e.g., acute and chronic, ipsilateral and contralateral, rest and exertion) [3]. Semantics has been used to study physician verbalizations [11–14] and diagnostic success associated with appropriate use of encapsulations and qualifiers, which was named semantic competence [5]. Recent research has integrated semantic competence on written case summaries and post-encounter forms on Objective Structured Clinical Examination (OSCE) exams [15–17].

Dual process theory [18] posits that categorization tasks are performed using two different modes of reasoning: system 1 (e.g., rapid and low effort) or non-analytical and system 2 (e.g., slow and high effort) or analytical. Regarding clinical tasks, non-analytical reasoning consists of a rapid process of pattern-recognition based on comparison and contrast of clinical findings with pre-existing KS. This allows physicians to deal with multiple environment stimuli and carry out complex clinical activities with minimal cognitive demands [19]. Prototype theory asserts that KS correspond to disease prototypes [20], which can be defined as averaged and abstracted portrayals of a given disease, developed through clinical experience by repetitive contact with different patients with a given disease. Prototypes include several clinical findings that are common in a particular disease.

The purpose of this study is to analyze student verbalizations about a clinical case on a high-stakes OSCE at the light of semantics and prototype theory and correlate them with OSCE, case-based exam and global medical degree (MD) scores. We hypothesize that students whose verbalizations are semantically richer and closer to the disease prototype will show higher test scores.

Methods

Study context

This study was conducted at the School of Medicine – University of Minho in Braga (Portugal) from January 2017 to December 2018. At this school, a 6-year MD admits 120 students each year. Year 6 (Y6) includes several clinical clerkships where students work as members of medical teams. The assessment program includes multiple-choice case-based exams, clerkship tutor assessments and an OSCE. This is a high-stakes exam which students need to pass for graduation, which includes six three-step 20-min stations. The pre-encounter introduces the clinical case and physical examination tasks before students enter the exam room. The encounter includes history taking and physical examination tasks on a standardized patient. After 15 min, students leave the exam room and complete the post-encounter form, including one question about the clinical case. OSCEs are conducted in two days where successive student groups are divided in three parallel rotations (A, B, C). Students are randomly assigned by the staff to one of the groups and distributed by one of the three rotations before the OSCE in a short briefing session. The exam includes four components which have different weights on the final score: clinical history (55%), physical exam (25%), communication (10%) and post-encounter form (10%). Standards are set for the clinical history and physical exam components using the borderline regression method.

Study subjects

Our target population were Y6 medical students at our institution. This study was conducted in all student groups from one of the three parallel OSCE rotations, which was chosen randomly. The Y6 OSCE took place in January 2017. All students were informed about the purpose of this study at preparatory skills sessions. In the day of the exam, student groups from the chosen rotation were openly invited to participate during the OSCE briefing session. Students signed an informed consent document and were free to excuse from participation. SPs and assessors were informed about the purpose and the design of the study before the exam.

Study design

One probing intervention was integrated at the end of the OSCE, regarding one of the OSCE stations (study station), in order to elicit students' verbalizations about that clinical case. We chose the 'chest pain' station which presented a middle-aged woman with cardiovascular risk factors and history of coronary artery bypass surgery presenting with chest pain on exertion in the past months, representing a typical case of stable *angina pectoris*. The probing intervention was conducted by one researcher and verbalized in the same manner to all participants. Students where asked to "*Imagine that the patient with chest pain was observed by you in the emergency room*" and to hand-over the case as if the researcher was the next physician responsible for the patient. 45 s were allowed for reflection and note-taking and there was no time limit for their answers. Answers to the probing intervention was audio-recorded.

Elicited verbalizations were used to make students' clinical reasoning visible to researchers, inspired in open-ended nature of the think-aloud methodology [7, 9]. The probing intervention was designed to capture structured reasoning in an authentic clinical format: the hand-over.

The OSCE assessment criteria were not changed for this study and all verbalizations were not considered for formal assessment purposes.

The wording and timing of the probing intervention was tested and fine-tuned with the help of student feedback after a small pilot in December 2016. An OSCE station was simulated and the intervention was conducted. Eight year 4 and year 5 students participated and an SP portrayed a clinical case no longer used in high-stakes exams.

Data analysis

Transcription and coding: All student verbalizations were transcribed verbatim. We analyzed total word count and total time of all verbalizations. Textual elements were identified based on semantics and prototype theory to create a coding framework:

(1) Semantics: one researcher (PGD) analyzed all student transcripts to create a list of *semantic transformations*, which included all encapsulations and semantic qualifiers found (Supplementary Material Table S1). Semantic transformations were considered to exist at growing orders of semantic value. For example, while 'Strong' encapsulations encompass several 'smaller' clinical concepts (e.g., heart failure includes dyspnea, orthopnea and peripheral edemas), 'Weak' encapsulations might only represent simple word transformations into medical terminology (e.g., breathlessness into dyspnea, high blood pressure into hypertension). Semantic qualifiers typically included opposing bipolar terms (e.g., acute vs chronic) but also words that were part of a meaningful clinical scale (e.g., dyspnea on minimal, mild, and moderate exertion). The use of semantic transformations was considered an indicator of semantic competence.

(2) Prototype theory: a disease prototype was created (Table 1) corresponding to the most likely diagnosis for the patient in the study station, which is stable *angina pectoris*. The prototype included *angina pectoris*' cardinal signs and symptoms [21] found in this patient's clinical presentation and was revised by two of the researchers (PGD and VH) who are practicing cardiologists.

After creation of the coding framework, one researcher (PGD) analyzed all student verbalizations (Table 2).

Categorization: We defined three semantic categories (Weak, Good and Strong), similarly to Dory et al. [16]. Four physicians involved in assessment of Y6 students in our school were invited to categorize student verbalizations. Categorization was guided by semantic cards (Table 3) created from the analysis of student transcripts (Supplementary Material Table S1), where semantic transformations were displayed at growing orders of semantic value.

A modified Delphi method [22] was used to build consensus. A face-to-face meeting was summoned where the study purpose, theoretical background and categorization method were introduced. Consecutive rounds were conducted where each person categorized all student verbalizations. In each round there was open discussions and results were grouped by the research team and fed-back to the group until consensus was reached. Consensus was reached at the end of three rounds.

Statistical methods

Descriptive statistics were calculated for all textual elements identified on student verbalizations and for each semantic category, as well as for total OSCE scores, partial OSCE scores (history taking and communication), study station history taking and communication scores, case-based exam score and global MD score (average score of the entire MD). One-way ANOVA with post-hoc Tukey tests were calculated to test for differences in textual elements between categories (e.g., number of semantic transformations and prototypical elements in each semantic category) and to explore associations between OSCE, case-based exam and MD scores and semantic categories. Pearson's correlations were calculated between continuous variables.

Table 1: Elements of the disease prototype for stable angina pectoris.

Disease prototype		
Previous medical history	Smoking habits	
	Dyslipidemia	
	History of coronary heart disease	
	Family history of coronary heart disease	
Current symptom: chest pain	Location	Retrosternal
		Radiating to the neck
	Character and intensity	Opressive
		Intensity of 5 (0–10 subjective scale)
	Duration	10 min
	Relationship to exertion	Appears after moderate exertion (climbing the stairs)
		Does not appear at rest
		Disappears after stopping
	Associated symptoms	Breathlessness/dyspnea on exertion

CABG, coronary artery bypass graft.

Results

In total, 118 Y6 students participated in the OSCE. We obtained 42 student verbalizations; the mean verbalization time and word number was, respectively, 58.64 s (16–123)

Table 2: Example of a coded verbalization: bold represents semantic transformations and underlined represents prototypical elements from the patient's clinical case.

Student 10
Woman, 54 years-old, with history of **coronary bypass surgery** six years ago, **dyslipidemia** and infarction in her family in young ages, and she is a smoker. She presented with **retrosternal** pain classified in 6 out of 10, worsening **on exertion** and not related with her meals … she also shows breathlessness on **mild**, **medium exertion** namely some of her daily tasks and she has no **edemas**.

Table 3: Semantic card. Semantic transformations found on student verbalizations are presented at growing orders of semantic value.

	Semantic value		
	Weak		**Strong**
Previous medical history	Infarction	Ischaemic/coronary event	Acute coronary syndrome
	Heart surgery	Coronary stenosis/occlusion	(Acute) myocardial infarction
	Catheterization	Cardiac intervention/interventioned disease	Revascularization
		Deobstructed	CABG, Angioplasty
	Cholesterol	–	Dyslipidemia/hyperlipemia
	High blood pressure		Arterial hypertension
	High blood sugar		Diabetes
	Aspirin, Ramipril, Bisoprolol, Atorvastatin	Antiaggregation, ACEi/antihypertensive	Dual antiplatelet therapy
	Furosemide	Beta-blocker, Statin/antidyslipidemic Diuretic	
Current disease/clinical presentation	Chest pain	Thoracalgia	Angina
		Central, retrosternal, precordial pain	Stable, unstable angina
	Climbing the stairs	On exertion/physical exercise	Minimal, mild or moderate exertion
	Resting/Stopping	Alleviates at rest	
	Breathlessness/tiredness	Dyspnea, dyspnea on exertion	Heart failure
	Swollen feet	Fatigue	

CABG, coronary artery bypass graft.

and 84.12 words (30–208). In average, verbalizations included eight semantic transformations (1–12) and 5 prototypical elements (1–9).

A total of 14 student verbalizations were categorized as Weak (33.3%), 22 as Good (52.4%) and 6 as Strong (14.2%). Examples of verbalizations from different categories are shown in Table 4.

Semantic 'Strong' verbalizations were longer than Good or Weak verbalizations and made use of more semantic transformations (Table 5). Differences between semantic categories in word count [$F(2.39)=3.30$, $p=0.048$, $\eta_p^2=0.145$] and semantic transformations [$F(2.39)=18.09$, $p<0.001$, $\eta_p^2=0.481$] showed significant results. Post-hoc testing showed differences in word count between Good and Strong

Table 4: Examples of semantic categories: weak and strong verbalizations.

Student 14	**Weak**
A lady, I can't remember her age exactly, with history of heart disease and who had had an **ischemic event** six years ago … she was admitted and followed-up in cardiology. She started with symptoms recently and … she started with **retroesternal pain** three months ago that radiated superiorly to the neck … and she was on medication	2 semantic transformations
Student 37	**Weak**
Woman, 59 years old, she came to the emergency room with a chest pain for several months now and more **on exertion** … she had a heart surgery because of a similar chest pain in the past and had two stents … the pain is better **at rest** … she also felt tired in the last days, she was on several medications and had cholesterol	2 semantic transformations
Student 15	**Strong**
So, she was a 59-year old woman who started with an oppressive **retrosternal** pain two months ago, classified as a 5 out of 10, that radiated to the neck and occurred on **exertion**, such as climbing stairs. She also had **dyspnea** and **fatigue** on **mild exertion** since three to four months ago. She is **dyslipidemic**, had been interventioned, she had a **heart bypass,** and she's a smoker, on medication with aspirin, ramipril and a **statin**	8 semantic transformations
Student 29	**Strong**
Woman with an oppressive **thoracalgia**, irradiating to the base of the neck, associated with **moderate exertion** and also **dyspnea**. She is a **coronary** patient who was **revascularized** six years ago	6 semantic transformations

Bold represents semantic transformations.

Table 5: Descriptive statistics of word count, time and elements per semantic category.

	Time, s (SD)	Word count (SD)	Semantic transformations (SD)	Prototypical elements (SD)
Weak	56.71 (31.87)	88.07 (32.72)	2.79[c] (1.19)	4.64 (1.34)
Good	56.73 (34.89)	74.23[a] (20.71)	5.23[c] (1.77)	5.73 (1.91)
Strong	70.17 (39.56)	111.17[a] (58.40)	7.83[c] (2.86)	6.50 (1.76)

[a]Represents a pair with statistically significant differences (p=0.05).
[c]Represents that all combinations show statistically significant differences.

verbalizations (p<0.05) and semantic transformations among all three categories (p<0.01). Semantic 'Strong' verbalizations showed more prototypical elements than 'Good' and 'Weak' verbalizations, but these differences were not statistically significant (F(2.39)=2.95, p=0.064, η_p^2=0.131).

Test scores

Global OSCE, OSCE components, case-based exam and global MD scores were obtained for the student sample. The mean global OSCE score in our sample was similar to the mean score in the population (13.45; SD 1.85). Descriptive statistics are shown in Table 6.

Semantic 'Strong' verbalizations showed higher test scores. Significant differences were found on total OSCE (F(2.36)=4.06, p<0.05, η_p^2=0.204), OSCE history taking (F(2.36) =5.06, p<0.05, η_p^2=0.226), study station history taking (F(2.36) =4.14, p<0.05, η_p^2=0.055), study station communication (F(2.36)=3.51, p<0.05, η_p^2=0.056), case-based exam (F(2.36) =3.88, p<0.05, η_p^2=0.179) and MD scores (F(2.36)=3.37, p<0.05, η_p^2=0.180). Differences between groups after post-hoc testing are shown in Table 7.

Table 6: Descriptive statistics of test scores in the student sample (scale from 0 to 20).

Mean OSCE score – all students (SD, range)	13.53 (2.01, 9.52–17.05)
Mean OSCE history taking score (SD, range)	11.31 (3.16, 4.19–16.97)
Mean OSCE communication score (SD, range)	17.21 (1.26, 13.60–19.13)
Mean study station history taking score (SD, range)	13.85 (1.74, 10.80–17.71)
Mean study station communication score (SD, range)	17.87 (1.60, 12.22–20.00)
Mean case-based exam score (SD, range)	13.59 (1.16, 11.79–16.43)
Mean global MD score (SD, range)	14.42 (0.89, 12.79–16.37)

Table 7: Descriptive statistics of test scores per semantic category (scale from 0 to 20).

	Semantics		
	Weak	Good	Strong
Mean OSCE score (SD)	12.67[a], SD=1.87	13.57, SD=1.90	15.29[a], SD=1.68
Mean OSCE history taking score (SD)	9.71[a], SD=3.16	11.50, SD=2.92	14.17[a], SD=1.70
Mean OSCE communication score (SD)	16.95, SD=1.29	17.21, SD=1.38	17.76, SD=0.50
Mean study station history taking score (SD)	12.86[a]; SD=1.34	14.22; SD=1.51	14.84[a]; SD=2.38
Mean study station communication score (SD)	18.09, SD=1.27	17.35[a], SD=1.77	19.16[a]; SD=0.77
Mean case-based exam score (SD)	13.10[a]; SD=0.90	13.61; SD=1.26	14.59[a]; SD=0.65
Mean global MD score (SD)	14.01[a], SD=0.85	14.44, SD=0.83	15.11[a], SD=0.82

[a]Represents pairs with statistically significant differences (p=0.05).

Regarding prototype theory, significant correlations were obtained between the number of prototypical elements and total OSCE (r=0.412, p<0.05), OSCE history taking score (r=0.371, p<0.05) and mean global MD score (r=0.428, p<0.05).

Discussion

This study showed that Y6 medical students' semantic competence and prototypical verbalizations were associated with higher exam and global MD scores.

A multi-theory coding framework was developed including elements from semantics and prototype theory to analyze student verbalizations. Analysis based on semantics required the creation of a semantic transformations list, which represents the full vocabulary of clinical concepts (or lexicon) used by the students about the clinical case. A panel of medical educators categorized all verbalizations according to their semantic competence. While the coding process described above is useful for research purposes, the categorization of student verbalizations is closer to the implementation of semantics in real-life assessment.

We observed that semantic 'Strong' verbalizations identified a small subset of students in our sample that voiced a larger number of words and semantic transformations than the majority of students in the other two categories (Good and Weak).

We had hypothesized that semantic competence and verbalizations matching the disease prototype would identify students with better knowledge organization. OSCE,

case-based exam and global MD scores were considered the gold-standard.

Semantic competence was in fact associated with higher scores in all three, while prototypical verbalizations were associated with higher OSCE and MD scores.

The association of semantic competence with higher scores in the case-based exam is coherent with better knowledge organization, since the exam tested diagnostic reasoning in case-based questions. This finding has been shown previously [5]. Regarding the OSCE, semantic competence may be associated with higher scores if students with better knowledge organization ask more key questions during history taking [23], which are valued in assessment checklists. On the other hand, the association between verbalizations matching the disease prototype and higher OSCE scores suggests that students with better knowledge organization find a match between the patient presentation and their pre-existing prototype for stable *angina pectoris*. The chest pain station is therefore more easily remembered at the end of the OSCE and verbalizations matching the prototype are visible on their hand-over verbalizations.

The association of semantic competence and prototypical verbalizations with higher global MD scores raises the possibility that these properties of clinical verbalizations identify students with better medical knowledge organization at the end of the MD.

In fact, we can hypothesize that semantics and prototypes and related: developed disease prototypes may be organized semantically, including encapsulations that encompass several smaller-order prototypical clinical findings. This encapsulation process might enable the process of pattern-recognition (non-analytical reasoning) that is necessary in everyday clinical practice to establish working diagnoses and decide on further testing and therapeutic actions. From this perspective, both semantic competence and prototypical verbalizations are likely to be found in students with better medical knowledge organization. Also, our findings highlight that there might be common learning mechanisms or behaviors by which some students develop better knowledge organization as well as other key competences throughout the MD, thus explaining the association with the global MD score.

Overall, our research suggests that established theories in Medical Education can be integrated in real-life assessment programs, namely in the categorization of student verbalizations or written texts at the light of semantics and prototype theory. Our results suggest that semantic competence is a particularly promising concept that can be explored with larger student cohorts, where the integration of semantic scales can be tested in high-stakes OSCEs and a validity argument for semantic assessment created [24].

The inclusion of semantic competence and prototype theory in assessment programs may be helpful not only to examine student knowledge organization, but also to infer about their clinical competence and support decisions about increasing levels of autonomy in clinical practice. Since developed disease prototypes are crucial elements for diagnostic reasoning [20], representing the basis for both non-analytical and analytical reasoning, this is particularly promising in relation to the assessment of diagnostic skills. From this perspective, students with underdeveloped disease prototypes might be selected for further tutoring or supervised clinical practice, in order to accumulate more experience with different patients and their varied clinical presentations.

This pilot work provides a foundation for further research on language analysis relying on theories of clinical reasoning and knowledge organization. Many research opportunities arise. Student verbalizations or texts may be analyzed in the setting of other assessment formats such as case-based discussions, where longer discourses provide better subtract for language analysis. Analysis can be expanded to clinical cases at different levels of complexity, typicality and other aspects of verbalizations such as treatment plans [25]. Different student and physician populations can be used to study the development of knowledge organization at different stages of training. Finally, automated speech recognition technology and the advent of machine learning may herald a new field of exciting research in medical knowledge organization and development at the interface of Education and linguistics.

Limitations

Regarding the study sample, we chose to conduct the study in only one of the three OSCE rotations (including 42 students). Our small sample size may limit the generalizability of these findings to other medical schools with different assessment programs, hindering extrapolations about the value of semantic competence in identifying students with better knowledge organization.

Regarding the study's theoretical background, although the concept of semantic transformation is grounded on theory, its specific definition in our work is the product of consensus among researchers and thus subjective, amenable to critique and refinement. The semantic value of some clinical concepts is unclear (Supplementary Material Table S1), which suggests that longer verbalizations are needed for further clarification. Also, we assumed a specific identity of KS as disease prototypes for the purpose of the study, but there is ongoing

debate in the Medical Education field with a competing view of KS as disease exemplars [20].

On the study's methodology, the use of verbalizations to infer about student's knowledge organization may be confounded by different factors such as linguistic constructs that are extraneous to the theoretical frameworks used, student understandings of the questions asked and their psychological context. First, students may differ on how they answer to short open-ended questions (longer and developed vs. shorter and concise answers, with or without the use of medical terminology); this may represent a 'verbosity' construct that could confound inferences on student knowledge organization. Second, student understandings of the questions may influence their answers: this might have happened in the probing intervention where the interpretation of the 'hand-over' concept depends on students' previous clinical experiences. Third, learned OSCE behaviors and anxiety associated with the exam might have influenced student discourses. Fourth, student verbalizations were categorized after consensus on a small assessor panel, which facilitated an immersive approach to the research topic but is unable to analyze inter-rater reliability of semantic categorization.

Conclusions

Y6 medical students' semantic competence and prototypical verbalizations were associated with higher OSCE and global MD scores. Semantic competence and clinical verbalizations matching disease prototypes may identify medical students with better knowledge organization. This work provides a foundation for future research based on the analysis of medical student verbalizations to infer about their knowledge organization, support assessment decisions and study medical knowledge development.

Acknowledgments: The authors acknowledge the collaboration of all participating students and standardized patients, as well as the contribution made by the assessor panel.

Research funding: None declared.

Author contribution: Conceptualization: PGD; Methodology: PGD, VHP, NS; Formal analysis and investigation: PGD, VHP, FP; Writing (original draft preparation): PGD; Writing (review and editing): PGD, VHP, CVV, SJD, NS, FP; Supervision: CVV, NS. All authors have accepted responsibility for the entire content of this manuscript and approved its submission. The views expressed here are those of the authors and not necessarily those of the Department of Defense or other federal agencies.

Competing interests: Authors state no conflict of interest.

Informed consent: Informed consent was obtained from all individuals included in this study.

Ethical approval: This research complied with relevant national regulations and institutional policies and has been approved by the Ethics Committee of the School of Medicine – University of Minho in January 2017.

References

1. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving. Cambridge, MA: Harvard University Press; 1978.
2. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. Acad Med 1990;65: 611–21.
3. Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. Med Educ 2007; 41:1133–9.
4. Kalyuga S. Rapid cognitive assessment of learners' knowledge structures. Learn Instr 2006;16:1–11.
5. Bordage G. Prototypes and semantic qualifiers: from past to present. Med Educ 2007;41:1117–21.
6. Chi MT, Feltovich PJ, Glaser R. Categorization and representation of physics problems by experts and novices. Cognit Sci 1981;5:121–52.
7. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev 1993;100: 363–406.
8. Pellegrino J, Chudowsky N, Glaser R, editors. Knowing what students know: the science and design of educational assessment. Washington, DC: National Academies Press; 2001.
9. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. Med Educ 2007;41: 1124–30.
10. Cruse A. Meaning in language: an introduction to semantics and pragmatics. UK: Oxford University Press; 2011.
11. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. Acad Med 1991;66:S70–2.
12. Lemieux M, Bordage G. Propositional versus structural semantic analyses of medical diagnostic thinking. Cognit Sci 1992;16:185–204.
13. Bordage G, Connell KJ, Chang RW, Gecht MR, Sinacore JM. Assessing the semantic content of clinical case presentations: studies of reliability and concurrent validity. Acad Med 1997;72:S37–9.
14. Baker EA, Connell KJ, Bordage G, Sinacore J. Can diagnostic semantic competence be assessed from the medical record? Acad Med 1999;74: S13–5.
15. Durning SJ, Artino A, Boulet J, La Rochelle J, Van Der Vleuten C, Arze B, et al. The feasibility, reliability, and validity of a post-encounter form for evaluating clinical reasoning. Med Teach 2012;34:30–7.
16. Dory V, Gagnon R, Charlin B, Vanpee D, Leconte S, Duyver C, et al. In brief: validity of case summaries in written examinations of clinical reasoning. Teach Learn Med 2016;28:1–10.
17. Smith S, Kogan JR, Berman NB, Dell MS, Brock DM, Robins LS. The development and preliminary validation of a rubric to assess medical students' written summary statements in virtual patient cases. Acad Med 2016;91:94–100.
18. Norman G. Dual processing and diagnostic errors. Adv Health Sci Educ Theory Pract 2009;14:37–49.

19. Custers EJ. Medical education and cognitive continuum theory: an alternative perspective on medical problem solving and clinical reasoning. Acad Med 2013;88:1074–80.
20. Papa FJ. A dual processing theory based approach to instruction and assessment of diagnostic competencies. Med Sci Educ. 2016;26:787–95.
21. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. ESC Guidelines for the diagnosis and management of chronic coronary syndromes: the task force for the diagnosis and management of chronic coronary syndromes of the European society of cardiology (ESC). Eur Heart J 2019;41:407–47.
22. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and nominal group in medical education research. Med Teach 2017;39:14–9.
23. LaRochelle J, Durning SJ, Boulet JR, van der Vleuten C, van Merrienboer J, Donkers J. Beyond standard checklist assessment: question sequence may impact student performance. Perspect Med Educ 2016;5:95–102.
24. Schuwirth LW, van der Vleuten CP. Programmatic assessment and Kane's validity perspective. Med Educ 2012;46:38–48.
25. Bissessur SW, Geijteman EC, Al-Dulaimy M, Teunissen PW, Richir MC, Arnold AE, et al. Therapeutic reasoning: from hiatus to hypothetical model. J Eval Clin Pract 2009;15:985–9.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/dx-2021-0048).