Health Professions Education

Volume 6 | Issue 1 Article 1

2020-03-01

Use of Concept Retrieval Technique as an Assessment Tool of Long-term Knowledge of Medical Students

Moeber Mahzari
College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia, Ministry of the National Guard, Health Affairs, Riyadh, Saudi Arabia

Marwa Kambal
College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

Tarig Mohammed
College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

Awad Alshahrani
College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia, Ministry of the National Guard, Health Affairs, Riyadh, Saudi Arabia

Fahad Al wadi
College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

Recommended Citation

Mahzari, Moeber; Kambal, Marwa; Mohammed, Tarig; Alshahrani, Awad; and Al wadi, Fahad (2020) "Use of Concept Retrieval Technique as an Assessment Tool of Long-term Knowledge of Medical Students," *Health Professions Education*: Vol. 6: Iss. 1, Article 1.
DOI: 10.1016/j.hpe.2019.06.001
Available at: https://hpe.researchcommons.org/journal/vol6/iss1/1

Available online at www.sciencedirect.com

ScienceDirect

Health Professions Education 6 (2020) 85–91

www.elsevier.com/locate/hpe

Use of Concept Retrieval Technique as an Assessment Tool of Long-term Knowledge of Medical Students

Moeber Mahzari [a,b,c,*], Marwa Kambal [a], Tarig Mohammed [a], Awad Alshahrani [c,a,b], Fahad Al wadi [a]

[a] *College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia*
[b] *King Abdullah International Medical Research Center, Riyadh, Saudi Arabia*
[c] *Ministry of the National Guard, Health Affairs, Riyadh, Saudi Arabia*

Received 10 February 2019; revised 1 June 2019; accepted 13 June 2019
Available online 19 June 2019

Abstract

Purpose: Medical students' knowledge assessment is a challenge. Multiple-choice questions (MCQs) despite their limitations are commonly used for this purpose. The Concept Retrieval Technique (CRT) is a relatively new knowledge assessment method. In this study we investigated the use of CRT to assess a particular long-term knowledge of medical students from different study levels.
Method: One hundred forty-four medical students (50% females) participated in an Endocrinology CRT exam. The students were either in fourth, fifth-or sixth-year of the medical program. Fourth-year students were tested before they received an endocrinology course whereas students from the fifth- and sixth-year respectively took the course ten and twenty-four months before being tested. The CRT was rated by two raters independently.
Results: There was strong effect of year of study on the CRT score. Fourth-year students scored significantly lower than fifth- and sixth-year students. No significant differences in CRT score emerged between fifth- and sixth-year students. The inter-rater agreement was acceptable indicating that the CRT is a reliable assessment method. The CRT showed construct validity.
Conclusion: The CRT is an assessment tool that could assess long-term knowledge retention properly, and has the ability to differentiate students according to their study level.

Keywords: Concept retrieval technique; Semantic network; Concept maps

1. Introduction

Medical curricula have developed over the last decade significantly under the influence of the profession and the community needs. Multiple related domains need to be taught and assessed in medical schools including knowledge, clinical skills, professional attributes, communications skills and others. Among all the domains, the knowledge acquired

* *Corresponding author.* King Saud Bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City, National Guard Health Affairs, Mail Code 3130, P.O. Box 3660, Riyadh, 11481, Saudi Arabia.

E-mail addresses: mahzarim@ksau-hs.edu.sa, Moeber@hotmail.com (M. Mahzari), kambalm@ksau-hs.edu.sa (M. Kambal), mohamedt@ksau-hs.edu.sa (T. Mohammed), awadshahrani@gmail.com (A. Alshahrani), Alwadi089@ksau-hs.edu.sa (F. Al wadi).

Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region.

https://doi.org/10.1016/j.hpe.2019.06.001

during medical school forms the scaffold where the students build on. Moreover, sufficient knowledge is fundamental to clinical reasoning, a competency that each medical student should develop.[1] Therefore, it is important to assess the knowledge properly however, effective and simple assessment of medical knowledge is a challenging task. Effective assessment method should be reliable, valid, have a positive impact on students' future learning and be cost effective.[2]

Multiple knowledge-directed assessment tools are available but, all of them have limitations and deficient in a way or another.[2,3] The most commonly used method in this regard is Multiple-Choice Questions (MCQs). A prominent advantage of MCQ exams is the ability to encompass a large extent of knowledge in a single exam. Moreover, administering and correcting MCQ exams are relatively easy. However, a major limitation of MCQ exams is the significant variability in its quality.[4,5] Indeed, Poorly written MCQs are quite common and usually assess recall of isolated facts as writing a good quality MCQ directed to higher levels of knowledge processing is a challenging task and time consuming.[5,6] Therefore the search for alternative knowledge-focused assessment methods will always be an area of interest to the medical education community.

It is therefore self-evident (but not common practice) that an understanding of the process of knowledge storage and retrieval from the mind should be used in the development of assessment tools. A widely accepted theory of this process is semantic network theory.[7–11] A semantic network represents semantic relations between concepts in the mind. Accordingly, the knowledge of medical students regarding any topic is represented in the mind as a set of interrelated concepts. Students acquire and use new concepts to reconstruct and embellish their semantic networks as they progress in their learning. The retrieval of the knowledge later on, for instance to answer questions on a test or solve a medical problem, requires activation of these networks.[7,8] The better the semantic network is organized, that is: the stronger the relationships between the concepts are, and the more concepts are held together, the easier activation of that knowledge is. An assessment tool focusing directly on the contents of these semantic networks makes sense to use if feasible and easy to administer.

A well-known instrument that does just that is the concept map.[12] A concept map is a diagram produced by a student that includes concepts and their interrelations on a given topic as he or she remembers them. It is a visual representation of knowledge organization in the mind.[12,13] However, using concepts maps as an assessment tool is a challenge. Concept maps vary to a very large extent from one student to another based on each student knowledge organization in his mind.[14,15] Therefore in order to score concept maps, all possible concepts and all conceivable relationships between all these concepts need to be part of an answer key. This task is extremely difficult. For this reason concept maps are primarily used as a teaching tool rather than an assessment one.[16,17]

A relatively recently developed assessment tool is the concept retrieval technique (CRT). Like the concept map, the CRT is rooted in the theory of semantic networks of medical knowledge. In this test, the students are asked to simply list all concepts he or she knows about a particular topic as a measure of how much he or she learnt about it. In order for the student to list all important concepts, he or she will have to activate the relevant semantic network from his mind. The better his knowledge is organized; the more concepts he will be able to retrieve.[18–20] The CRT has been used successfully as an assessment tool in problem-based learning (PBL) to assess immediate knowledge gain.[18–21] Moreover, CRT score was found to be correlated positively with short essay scores, suggesting that the CRT is a valid instrument for assessing knowledge.[18,19]

Until now, up to our knowledge the tool has only been used to measure knowledge acquisition over the course of a single day.[18–21] But knowledge acquisition in the context of medicine is a process that takes place over much longer time frames. In this study, we were particularly interested whether the CRT was able to represent knowledge of endocrinology in 4–6th year medical students, one year before students attended a course specifically focused on advanced knowledge of this topic, and one year after this course. Our predictions were that the effects of learning, both short-term (directly after the course) and long-term (after at least 10 months) would be reliably represented in the CRT-scores of these students.

2. Method

The study is a cross sectional study about the use of CRT to assess medical students' knowledge about three topics in an endocrinology course.

2.1. Participants

The participants in this study were all from one governmental Saudi medical school. This medical school accept around 300 Saudi national students per year (around third of them female). All the students

from the same year study the same integrated courses at the same time but male and female students are separated. One hundred forty-four medical students conveniently and voluntarily recruited to participate in the study (50% of them were female). One-third of the participants were students in fourth-year, a second third were in their fifth-year, and the rest were in the sixth-year of the medical program. Students from fifth and sixth year must have taken the endocrinology course in order to participate in the study.

To recruit the students to the study, the study team attended with the students the last 10 min of a regular lecture, after the lecture finished the research team introduced the available students to the study. The students who remained after the study introductory session and signed consent then took the CRT exam. This procedure was done for male and female students from each study level separately.

2.2. Materials

The CRT consisted of three questions on three different endocrinology topics, the questions asked about both biomedical and clinical sciences knowledge. The test was developed by two endocrinologists who are involved in teaching of the endocrinology course in the curriculum. The questions were as follows: 1. List all relevant concepts about pathogenesis of graves' hyperthyroidism. 2. List all relevant concepts about primary Adrenal insufficiency (including pathogenesis, diagnosis and management of primary adrenal insufficiency). 2. List all relevant concepts about diagnosis of type 2 diabetes. A list of all possible concepts that can be considered correct answers for each question constructed by the same two endocrinologists who developed the CRT questions. The list of all admissible concepts used in the scoring of the CRTs is presented in Appendix 2.

2.3. Procedure

The medical curriculum of the Saudi medical school where the study was conducted is an integrated curriculum where biomedical and clinical sciences are integrated throughout all the courses. However, the students spend two preparatory years studying discipline based basic sciences before they start the integrated courses of the medical curriculum in third year of their study. The endocrinology course is a four-week course delivered during the last month of the first semester in year four of the medical program. The course is delivered mainly through problem-based learning (PBL) where each week is about one problem. Relevant lectures, laboratory and clinical teaching sessions are also conducted in each week. The first week of the course is about thyroid disorders including hyperthyroidism and hypothyroidism; the second week is about pituitary disorders while the third week is about adrenal disorders including adrenal insufficiency. The last week of the course is about different types of diabetes including type 2.

The study participants who were from fourth year did the CRT before they took the endocrinology course while fifth-and sixth-year students took the course respectively around ten and twenty-four months prior to the study. The CRT was administered to the students in a supervised classroom setting. To help the students understand how to respond to the CRT, they were given a one-page example of a non-related CRT question with its standard answer (appendix 1).

2.4. Statistical analysis

Two content experts scored the exam independently using the answer key. Each correct concept was given one mark. Inter-rater agreement was assessed by Cohen's Kappa and was found to be 0.67 for Question 1, 0.71 for Question 2 and 0.59 for Question 3 which was considered acceptable. Raters then sat together and resolved differences of opinion. Then total scores for each student were calculated.

CRT scores were computed by summing the number of concepts written down for each of the three questions and their total. The data was analyzed through one-way analysis of variance (ANOVA) using the Statistical Package for the Social Sciences (SPSS v20). Bonferroni correction was applied to counteract the effect of multiple comparisons on the alpha level of 0.05.

3. Results

One hundred forty-four medical students participated in the study. All the participants were included in the final analysis. Table 1 shows means, standard deviations, per question and total CRT for students from each year. The mean total CRT score was 2.9 (minimum score 0 and maximum score 10.0) for 4th year student, 10.5 (minimum score 2.0 and maximum score 20.0) for 5th year students and 8.6 (minimum score 2.0 and maximum score 20.0) for 6th year students.

A one-way ANOVA was conducted using the Bonferroni correction. The results of the analysis showed a strong significant overall effect of year on CRT-performance: F (2, 143) = 51.23, $p < .0001$, partial

Table 1
Sample sizes (N), means and standard deviations for the three questions and total CRT-scores by curriculum year.

N	Year 4		Year 5		Year 6	
	48		48		48	
	Mean	SD	Mean	SD	Mean	SD
Question 1 CRT	.9	.9	2.4	1.2	2.3	1.2
Question 2 CRT	.4	.9	3.4	2.2	2.8	2.1
Question 3 CRT	1.6	1.3	4.6	2.4	3.6	2.0
Total CRT	2.9	2.3	10.5	4.6	8.6	4.1

eta-squared = 0.42. (The latter indicates that differences in year-of-study explain 42% of the variation in CRT-scores.) This effect was also visible for each of the three questions separately. Pairwise comparisons revealed that—overall–the increase in number of concepts retrieved between almost all years was statistically significant (Year 4 versus Year 5: $p < .0001$, Year 4 versus Year 6: $p < .0001$) with the exception of Year 5 versus Year 6: $p = .054$). The actual difference between Year 5 and 6 indicates that sixth-year students retrieved *fewer* concepts than fifth-year students but this effect is only marginally significant. Fig. 1 shows the differences for each CRT question and the total CRT score between the three years. No significant differences were observed between male and female students.

4. Discussion

The CRT has served as a sensitive formative assessment tool immediately after learning activity.[18–20] However, there have been no published studies on its use as a summative assessment of long-term knowledge. In this study, we wanted to examine the use of CRT to assess medical students' knowledge in endocrinology in relation to students' study level and time of exposure to the tested knowledge. The total CRT score was significantly different between students from different studying levels. The fourth-year students significantly recalled fewer concepts compared to either fifth-or sixth-year students. This finding supports that CRT is sensitive to reflect the difference in knowledge level in relation to knowledge exposure even if knowledge exposure was some time ago. The reason of why fourth-year students scored lower on the test could be explained on the basis of semantic network theory. As it is known, semantic networks are established in the mind by first learning activity and mature to expert level in relation to advancement in learning about a given topic. In this study, Fourth-year students acquired some endocrinology knowledge in the context of other courses but were not yet systematically exposed to the topics tested in the exam. Therefore, their semantic network of the topics tested by the CRT was poor in terms of the number of concepts and poorly structured in terms of the relationships between these concepts, if any at all. The result was that they, on average, retrieved few concepts. On the other hand, student from year 5 and 6 already studied the endocrinology topics tested by the CRT. Therefore, their semantic networks of the tested knowledge were better developed which was reflected by more concepts being retrieved by the students. Having said that, fifth-year students recalled more concepts

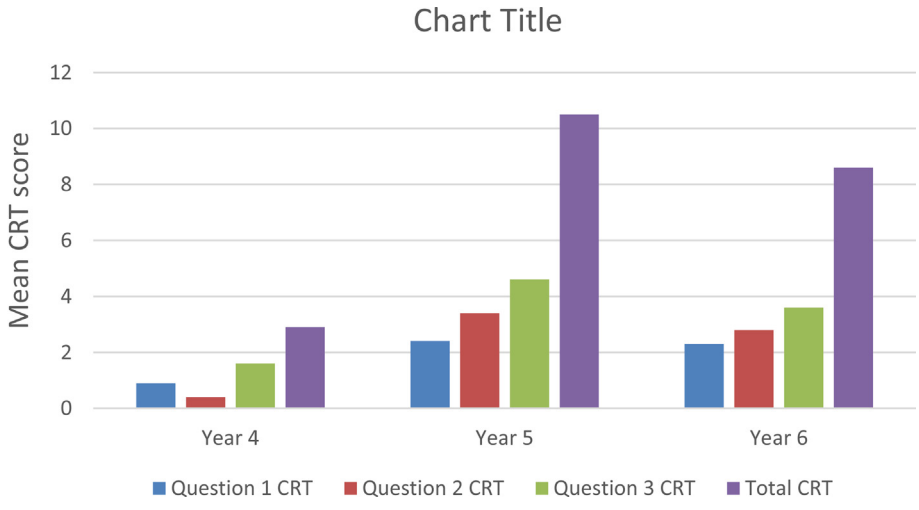

Fig. 1. The Mean CRT scores by year of study.

when compared to sixth-year students, however the difference between the two groups did not reach statistical significance. Both groups; fifth-and sixth-year students; were exposed to the endocrinology course, however the time from the exposure to the course when the study conducted was more than doubled for sixth-year students compared to students from the fifth year. As we have said before, the semantic networks of knowledge mature to an expert level as we learn more about a given topic. However, if the learning is interrupted for a significant period of time then these networks are expected to not improve and actually some of the concepts may decay or become less easily accessible.[22] The sixth-year students were more senior than fifth-year students but their exposure to the specific knowledge that was tested in the CRT was acquired longer ago. The exposure to the tested knowledge was at one specific time during the curriculum with probably insufficient formal consolidation of this specific knowledge afterwards. Therefore, somewhat lower number of concepts retrieved by the sixth-as compared to fifth-year students is likely due to decay of knowledge in the absence of sufficient knowledge consolidation activity. This decay of knowledge due to insufficient consolidation is a phenomenon from which medical education as a whole suffers.[23–25]

The reliability and validity of the CRT as an instrument for measuring long-term retention was tested in this study. The extent of agreement between independent raters is considered the most meaningful indicator of reliability for free recall-type of questions.[26] The reasonable inter-rater agreement of the CRT in our study indicates that it is a reliable assessment tool when used to assess long-term knowledge. Other studies have demonstrated somewhat higher inter-rater agreements when CRT was used for immediate/short term knowledge acquisition.[27] It seems that assessing long-term retention of concepts is slightly harder to do, because students may come up more often with concepts not foreseen in the answer key but nevertheless sometimes meaningful.

One way of assessing the validity of an instrument is looking at its construct validity, which is the degree to which a test is measuring what it sets out to test.[28] Borsboom et al. explains that if one is concerned with establishing the validity of a new measure, what needs to be demonstrated is that manipulation of the (non-observable) attribute being measured results in changes in the scores of the measure.[29] The Construct validity of CRT is assessed in this study indirectly by testing the change in CRT score in relation to learning over time. The difference in CRT scores clearly linked to learning suggests that the CRT is indeed valid.

Using the CRT as an assessment tool seems feasible. Its advantage lies in its direct assessment of the semantic networks of specific knowledge in a holistic and comprehensive way. Second, it is easy to develop as a single question is easy to write and could test a large span of knowledge. However, the rating of the answers is time-consuming and more prone to subjectivity compared to MCQs. Having said that, Development of a comprehensive answer key to be used by the test raters is a necessary step to overcome this limitation to some degree. In addition, attempts are underway to score CRTs automatically using artificial intelligence of a computer program.[27]

5. Conclusion

Our study is the first to examine the use of CRT to assess long-term knowledge of medical students. CRT is a valid and reliable assessment tool for long term knowledge assessment and has the ability to differentiate students according to their study level.

CRT shows promise and hope that a new knowledge-directed assessment tool can be put in the hands of the medical teacher, in the tradition of the concept-mapping technique but without its disadvantages. However, the study was conducted in one medical school and with a limited number of students. Direct comparison of the CRT with MCQs and short-essay questions to assess its usefulness as an alternative are next research steps.

Appendix 1.

Example of a concept retrieval test (This example used in our CRT as a leading example on how to answer the questions).

Write down everything you remember about the topic of osmosis.

Do not use full sentences, use powerpoint-type response. Initially this may look somewhat artificial. But try to remember how the process goes and what factors are involved. For instance: “Mmm, …osmosis takes place in cells. Water molecules go through the semipermeable membrane of the cell. But why? There are concentration differences between the cell itself and its environment. This leads water molecules from an area with high concentration to the lower concentration in the cell. Movement of water molecules is through Brownian movement. This diffuses the water.”

This retrieval of everything you remember about osmosis leads to the following concepts:

Cells.
Water molecules.
Differences in concentration.
Semipermeable membrane.
Diffusion.
Brownian movement.

Appendix 2.

First question: List all relevant concepts about pathogenesis of graves' hyperthyroidism.

- Autoimmune disorder
- Thyroid
- T cells
- B cells
- Thyroid stimulating antibodies (TSI,TRAb,TBI)
- TSH receptors
- Thyroid hormones synthesis
- High thyroid hormones in tissues
- Hyperthyroidism

Second question: List all relevant concepts about primary Adrenal insufficiency (including athogenesis, diagnosis and management of primary adrenal insufficiency).

- Adrenal disease
- Autoimmune adrenalitis
- Infiltrative diseases
- Nausea and vomiting
- Low blood pressure
- Low cortisol
- Low mineralocorticoides
- High ACTH
- Low sodium
- High potassium
- ACTH stimulation test
- Hydrocortisone
- Fludrocortisone
- Stress management

Third question: List all relevant concepts about diagnosis of type 2 diabetes.

- Sedentary life style
- Family history of diabetes
- Metabolic syndrome
- Obesity
- Family history
- Acanthosis Nigerians
- Insulin resistance
- Hyperglycemia
- Polyuria
- Polydipsia
- Weight loss
- Hyperglycemic hyperosmolar state
- Fasting glucose
- Random glucose
- Hemoglobin A1c
- Oral glucose tolerance test

References

1. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39(4):418–427.
2. Van Der Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract*. 1996;1(1):41–67.
3. Cameron T, McCoy AL, Ballard A, et al. *Curriculum inventory standardized instructional and assessment methods and resource types*. Washington, DC: Association of American Medical Colleges; 2012. MedBiquitous curriculum inventory working group standardized vocabulary subcommittee.
4. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007 Jan 25;2007(356):387–396.
5. Vanderbilt AA, Feldman M, Wood IK. Assessment in undergraduate medical education: a review of course exams. *Med Educ Online*. 2013;18.
6. Bridge PD, Musial J, Frank R, Roe T, Sawilowsky S. Measurement practices: methods for developing content-valid student examinations. *Med Teach*. 2003;25(4):414–421.
7. Schmidt HG. Foundations of problem-based learning: some explanatory notes. *Med Educ*. 1993;27(5):422–432.
8. Steward D, Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. *Acad Med*. 1991;66(9): S70–S72.
9. Chiarello C, Burgess C, Richards L, Pollock A. Semantic and associative priming in the cerebral hemispheres: some words do, some words don't ... sometimes, some places. *Brain Lang*. Jan 1990;38(1):75–104.
10. Landauer TK, Dumais ST. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev*. 1997;104(2):211.
11. Plaut DC. Semantic and associative priming in a distributed attractor network. In: *Paper presented at: Proceedings of the 17th annual conference of the cognitive science society*. 1995.
12. Novak JD. Concept maps and Vee diagrams: two metacognitive tools to facilitate meaningful learning. *Instr Sci*. 1990;19(1):29–52.
13. Novak JD, Gowin DB. *Learning how to learn*. Cambridge, UK: Cambridge University Press; 1984.
14. Ruiz-Primo MA, Shavelson RJ. Problems and issues in the use of concept maps in science assessment. *J Res Sci Teach*. 1996; 33(6):569–600.
15. Ho V, Kumar RK, Velan G. Online testable concept maps: benefits for learning about the pathogenesis of disease. *Med Educ*. 2014;48(7):687–697.

16. Daley BJ, Torre DM. Concept maps in medical education: an analytical literature review. *Med Educ*. 2010;44(5):440–448.
17. Edmondson KM. *Assessing science understanding through concept maps*. London, UK: Elsevier; 2000.
18. Chng E, Yew EH, Schmidt HG. Effects of tutor-related behaviours on the process of problem-based learning. *Adv Health Sci Educ Theory Pract*. 2011;16(4):491–503.
19. Chng E, Yew EH, Schmidt HG. To what extent do tutor-related behaviours influence student learning in PBL? *Adv Health Sci Educ Theory Pract*. 2015;20(1):5–21.
20. Yew EH, Schmidt HG. What students learn in problem-based learning: a process analysis. *Instr Sci*. 2012; 40(2):371–395.
21. Rotgans JI, Schmidt HG. Situational interest and learning: thirst for knowledge. *Learn Instr.* 2014;32:37–50.
22. Berry DC, Berry D, Dienes ZP, Dienes Z. *Implicit learning: theoretical and empirical issues*. Psychology Press; 1993.
23. Dudai Y, Karni A, Born J. The consolidation and transformation of memory. *Neuron*. 2015;88(1):20–32.
24. Amaral F, Troncon LEdA. Retention of knowledge and clinical skills by medical students: a pro-spective, longitudinal, one-year study using basic pediatric cardiology as a model. *Open Med Educ J*. 2013;6(1).
25. Arthur Jr W, Bennett Jr W, Stanush PL, McNelly TL. Factors that influence skill decay and retention: a quantitative review and analysis. *Hum Perform*. 1998;11(1):57–101.
26. Leiva FM, Ríos FJM, Martínez TL. Assessment of interjudge reliability in the open-ended questions coding process. *Qual Quantity.* 2006;40(4):519–537.
27. Hays GJ. *Developing a new measure of conceptual knowledge: the Concept Retrieval Test*. Ph.D.-thesis. Rotterdam, the Netherlands: Erasmus University; 2018.
28. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281.
29. Borsboom D, Mellenbergh GJ, van Heerden J. The concept of validity. *Psychol Rev.* 2004;111(4):1061–1071.