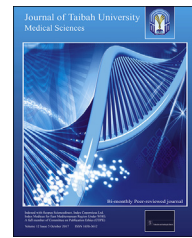Taibah University

Journal of Taibah University Medical Sciences

Educational Article

Cognitive load theory: Practical implications and an important challenge

CrossMark

Jimmie Leppink, Ph.D.

Department of Educational Development and Research, School of Health Professions Education, Maastricht University, Maastricht, The Netherlands

الملخص

اعتمد مجال التعليم الطبي مجموعة واسعة من النظريات من عدة مجالات أخرى. ومن الأمثلة الحديثة إلى حد ما هو نظرية التحمل الإدراكي، التي نشأت من علم النفس التربوي. العديد من الدراسات التجريبية المستوحاة من نظرية التحمل الإدراكي واستعراض الآثار العملية لنظرية التحمل الإدراكي أسهمت في إعداد القواعد الإرشادية لتصميم التعليم الطبي. في الوقت ذاته، وضعت عدة مجموعات بحثية أدوات لقياس التحمل الإدراكي في سياق التعليم الطبي. وعلى الرغم من هذه التطورات، يبقى الحصول على أدلة لأنواع مختلفة من التحمل الإدراكي تحديا هاما. ومن أجل ذلك، تهدف هذه المقالة إلى أمرين: لتزويد معلمي الطب بثلاث قواعد إرشادية لتصميم التعليمات والتقييم ولشرح بعض القضايا الأساسية في التحدي الباقي حول أنواع مختلفة من التحمل الإدراكي. تدور القواعد الإرشادية حول التقليل من النشاط الإدراكي الذي لا يسهم في التعلم، ووضع أهداف محددة للتعلم في عين الاعتبار، وتقدير العلاقة متعددة الأوجه بين التعلم والتقييم. وتتضمن القضايا الرئيسة حول التحمل الإدراكي السياق الذي يحدث في التعلم، والاستمرار في استخدام عنصر واحد لتقديرات الجهد العقلي، وتوقيت التحمل الإدراكي ونتائج التعلم.

الكلمات المفتاحية: نظرية التحمل الإدراكي؛ التعلم؛ التعليم؛ التصميم؛ القياس

Abstract

The field of medical education has adopted a wide variety of theories from other fields. A fairly recent example is cognitive load theory, which originated in educational psychology. Several empirical studies inspired by cognitive load theory and reviews of practical implications of cognitive load theory have contributed to guidelines for the design of medical education. Simultaneously, several research groups have developed instruments for the measurement of cognitive load in a medical education context. These developments notwithstanding, obtaining evidence for different types of cognitive load remains an important challenge. Therefore, the aim of this article is twofold: to provide medical educators with three key guidelines for the design of instruction and assessment and to discuss several fundamental issues in the remaining challenges presented by different types of cognitive load. The guidelines revolve around minimizing cognitive activity that does not contribute to learning, working with specific learning goals in mind, and appreciating the multifaceted relation between learning and assessment. Key issues around the types of cognitive load include the context in which learning occurs, the continued use of single-item mental effort ratings, and the timing of cognitive load and learning outcome measurements.

Corresponding address: Department of Educational Development and Research, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.
E-mail: jimmie.leppink@maastrichtuniversity.nl

Production and hosting by Elsevier

Introduction

The field of medical education has adopted a wide variety of theories from other fields. A recent example is cognitive load theory (CLT),[1–11] which originated in educational psychology.[1,2,6–9] CLT defines learning as the development

and automation of cognitive schemas stored in long-term memory about content to be learnt (e.g., anatomy of the human body[12] or a particular type of systematic problem-solving procedure[13,14]). A vast body of empirical work has demonstrated the narrow limits of human working memory,[15–18] and CLT states that the design of education has to respect these limits.[4,5,9,11] Several empirical studies inspired by CLT[12,20–22,24,26–28] and reviews of practical implications of CLT[4,5,10,11,19,23,25,29,30] have contributed to guidelines for the design of medical education. Simultaneously, several research groups have developed instruments for the measurement of cognitive load in a medical education context.[22,24,26–28] These developments notwithstanding, obtaining evidence for different types of cognitive load remains an important challenge. Therefore, the aim of this article is twofold: to provide medical educators with three key guidelines for the design of instruction and assessment and to discuss several fundamental issues in the remaining challenge concerning different types of cognitive load.

Three core guidelines for the design of instruction and assessment

Following the aforementioned definition of learning in CLT as the development and automation of cognitive schemas regarding content to be learnt, three types of cognitive load have been distinguished in the literature: *intrinsic cognitive load* (ICL), *extraneous cognitive load* (ECL) and *germane cognitive load* (GCL).[4,5,9,11] When confronted with information about content to be learnt, the incompleteness and lack of development – or lack of automation – of a learner's cognitive schemas about that content imposes ICL. The more content elements that need to be processed by working memory at a given time and/or the more interaction between elements (i.e., element interactivity[5]), the more ICL for a learner. Next, ECL is cognitive load due to cognitive processes that as such do not contribute to learning.[31,32] Finally, GCL has been viewed as cognitive load due to the deliberate engagement in cognitive processes that are beneficial to learning, including asking the right questions, appropriate self-explanation of content, accurate metacognitive monitoring of learning and performance, and following up on that monitoring with adequate learning activity.[9–11]

In recent years, several researchers have suggested a modified dual model that includes only ICL and ECL and gives a broader interpretation to ICL, depending on the goals of learning and instruction.[1–5,7,8,14] It is important to note that this dual model does not deny the *existence* of GCL; rather, it is cognitive load due to working memory resources allocated to dealing with ICL, or the part of ICL that benefits learning.[1,4,5] If none of the ICL is dealt with successfully, GCL is 0; if all ICL is dealt with successfully, all ICL is GCL. In other words, while in the traditional three-factor ICL/ECL/GCL cognitive load model[9–11] GCL is a distinct third type of cognitive load, in the modified two-factor ICL/ECL cognitive load model, GCL is a proportion (i.e., somewhere on a scale from 0 to 100%) of ICL. Effectively, the two models support exactly the same guidelines for the design of education and training. Since a variety

of articles and book chapters have provided rather detailed reviews and overviews of recommendations for education and training,[3–5,8–11,19,23,25,30,33] using examples from recently published research, this article focuses on three core guidelines, two of which have been considered mainly more recently.

Guideline (1): minimize cognitive activity that does not contribute to learning

The first guideline revolves around minimizing ECL, meaning that instruction should be designed in such a way that only a minimum of working memory resources is needed for cognitive processes that do not contribute to learning as such.[4,5,8–11] Well-known examples of such cognitive processes among learners who are new to a certain topic are having to verbally process information that ought to be presented visually[5] and having to divide one's attention between information sources, in different spaces or times, that could be integrated into a single source.[3,4,10] These effects eventually disappear as learners become more proficient, and providing support where it is not needed may contribute to ECL.[5,8,10] For instance, when early stage learners have to learn a complex procedure, ECL due to ineffective problem-solving search can be reduced by having them study a worked example of a successful completion of a procedure first.[34] However, this beneficial effect of support among novice learners disappears and eventually reverses when applied to more advanced learners.[35,36]

When we ask learners to do an objective structured clinical examination (OSCE) with possible diagnoses in mind and to explicitly engage in forward (i.e., from symptom to diagnosis) and backward (i.e., from diagnosis to symptom) clinical reasoning,[37–39] we can expect a higher ICL than when we ask learners to focus primarily on the manoeuvres of the OSCE procedure.[4,22,23] Likewise, when we ask undergraduate students to practice with a simulated patient in an authentic simulated workplace environment (i.e., simulated clinical immersion), they will probably experience a higher ICL than when we let them practice with that simulated patient outside such an environment, since in the latter case there are no environmental stimuli to pay attention to.[26] Moreover, in most medical procedures, it is not sufficient to merely learn the steps of a procedure. Rather, these steps often have to be undertaken in a particular sequence to ensure a correct solution. The order matters, and that interactivity adds to ICL. In such an environment, having to address patient cases where there are many possible diagnoses and/or several comorbidities[23] may take the ICL for less experienced learners to the limits of their working memory. However, more advanced learners will probably experience a lower ICL in such a situation because they can activate more developed and perhaps already more automated cognitive schemas than their less experienced peers.

Careful reflection on this ICL factor is of paramount importance, because in the aforementioned case (i.e., OSCE and simulated clinical immersion) and other settings in medicine and healthcare, several sources not yet mentioned can contribute to ECL. First, having to address patient cases that are very complex for learners at a given stage without adequate instructional support from a supervisor or the

environment is likely to trigger ineffective problem-solving search activity that does not contribute to learning.[5,10] Second, confusing instructions from a supervisor or peer student could trigger cognitive processes that hinder learning.[22] Third, distractors from the working environment[26] or even from one's own thoughts or emotions (e.g., pondering about feedback that a mannequin died[20,21]) can consume working memory resources that could otherwise be used for learning.

Guideline (2): all the way work with specific learning goals in mind

As mentioned previously, the modified two-factor ICL/ECL model proposes a broader interpretation of ICL depending on the goals of learning and instruction. This interpretation has resulted in a suggestion to introduce specific instructional goals as a key aspect to consider in CLT.[2] These goals are not necessarily limited to learning specific content; they may refer to motivational, affective, and metacognitive activities as well.[40] Whatever these goals refer to, they can help educators and researchers define what is ICL and what is ECL in a given context. All working memory capacity that is needed for activities that, as such, do not contribute to achieving the specific goal(s) under consideration is ECL. For example, Tremblay and colleagues[26] have demonstrated that undergraduate pharmacy students who have little if any prior experience with a simulated authentic workplace environment tend to focus more on environmental stimuli and less on clinical reasoning when practising in such an environment compared with outside such an environment. If the goal of an exercise is to have students learn how to address specific stimuli (e.g., phone ringing, colleagues passing by, patient files) in a workplace environment, cognitive load due to dealing with these stimuli is ICL. However, if the focus in an exercise is on clinical reasoning, cognitive load due to environmental stimuli can be considered ECL because it takes away working memory resources that could otherwise be used for clinical reasoning. Starting with practice outside such a workplace environment may then help educators and trainers to have students focus on the development of clinical reasoning.

Task complexity can be influenced by increasing either the number of content elements to be processed at a given time or the extent to which elements that have to be processed interact with each other (e.g., a fixed sequence in procedural steps). In OSCE design, for instance, complexity can be increased by instructing learners to perform the procedure with competing hypotheses about possible diagnoses in mind.[22,37−39] Moreover, task complexity can be increased by having to address more symptoms, comorbidities, and acuity of a case.[23] Of course, elevating complexity is unlikely to benefit learners who have little experience with OSCEs and do not yet know the manoeuvres very well. If the goal of a training exercise is to help learners practice specific manoeuvres, all cognitive load related to learning these manoeuvres is ICL. However, if the goal of the exercise is to learn how to perform a physical examination and engage in clinical reasoning with specific hypotheses in mind, ICL is that cognitive load that arises from engaging − or learning to engage − in clinical reasoning while performing the examination. It is at this stage, when

learners are somewhat more advanced, meaning they know the manoeuvres, that increases in the number of symptoms, comorbidities, and acuity of a case can help to achieve the goal of learning how to perform a physical examination and engage in clinical reasoning with specific hypotheses in mind for a range of cases. Finally, at the next stage, schema development and automation can be stimulated further through contextual interference, variability, and imagination.[10,11] Although these factors are likely to contribute to ECL among less experienced learners, they can help advanced learners to address an increased element interactivity associated with ICL.

In short, learning tasks and training activities should be designed with specific learning goals in mind. These learning goals will help educators and researchers to determine what cognitive activity is essential for achieving a given goal (i.e., essential for learning: ICL) and what cognitive activity is not essential for and/or may hinder achieving that goal (i.e., not essential for learning: ECL).[1,2] While CLT has traditionally focused on the learner, the modified dual ICL/ECL approach proposes *learner activity* as the main unit of analysis,[2] and both learner-related and activity-by-learner-interaction related factors[33] may, depending on the goals of the activity, influence ICL as well as ECL. In this view, any kind of 'other' cognitive load that does not fit within either ICL or ECL − call it GCL or whatever − is redundant.[1,2,4]

Guideline (3): appreciate the multifaceted relation between learning and assessment

Apart from distinguishing between ICL and ECL, specific learning goals can help to design appropriate *assessment of learning*. For example, if the goal of a training exercise is to make students familiar with manoeuvres that are needed to perform an OSCE, OSCEs that focus on these manoeuvres can help educators to assess the extent of students' mastery of these manoeuvres. Simultaneously, when well designed, these OSCEs may serve as *assessment for learning*, meaning that they may drive subsequent practice and learning. For example, once learners master the manoeuvres that are needed in a particular type of OSCE station, the next stage is to have students practice with OSCEs in a more hypothesis-driven approach.[22,37−39] Especially when students are about to do their internships, they may have the motivation as well as the cognitive schema development and, given the nature of the internship, need for practice at this next stage in the learning process.[22] Likewise, although simulated clinical workplace environments may be experienced as somewhat stressful, especially by undergraduate students, it is, for patient safety and for creating a safe learning environment for the student, probably advisable to have students practice at this level before moving on to real patients. Appropriately designed assessments should inform educators and students when it is time to move to the next level (i.e., from outside to inside a simulated workplace environment, and from a simulated workplace environment to a real workplace environment). In other words, these assessments are not end points as in assessments of learning after some training period. Rather, they are carried out *while learning* occurs, and as such can constitute a practice of high-frequency low-stakes assessments in a longitudinal trajectory rather than low-frequency

(end point) high-stakes assessments.[41−43] Of course, this does not exclude having high-stakes 'end point' assessments, as well.

Apart from learning particular content, we may want to train our students in monitoring their own learning and making appropriate choices in what to study and practice next.[44,45] However, it is important to note that these self-regulated learning processes also require working memory capacity, and it is perhaps for that reason that learners are unlikely to spontaneously use their learning task performance or effort invested in a learning task to reflect on which task to select next.[45] Whether the working memory resources allocated to engaging in these self-regulated learning processes are to be considered ICL or ECL in a given context depends on whether the development of these processes or skills constitutes a learning goal (i.e., ICL) or not (i.e., ECL).

Novice learners in particular tend to be poor at monitoring their own learning,[46−48] and poor monitoring is unlikely to result in accurate learning task selection.[45] Hence, novices need support in the development of these skills. When properly designed, assessment activities can serve as *assessment as learning*: not only are students informed how well they are doing and what they might do next, the assessment itself presents them with assessment criteria they may start using to monitor their learning from that point forward. Of course, given the narrow limits of working memory, to be effective in the latter, careful reflection is needed with regard to how much room there is, given other learning goals in a particular course or unit (e.g., examination procedure or clinical reasoning), for a learning goal on self-regulated learning skills.

Challenge of obtaining evidence for different types of cognitive load

Although CLT has clearly had a positive impact on education in medicine and other areas, obtaining evidence for different types of cognitive load remains an important challenge. This section focuses on three key issues around this challenge: the context in which learning occurs, the continued use of single-item mental effort ratings, and the timing of cognitive load and learning outcome measurements.

Context in which learning occurs

Whether a certain activity contributes to ICL and ECL depends on the learning goals.[2] Moreover, given specific learning goals, while more experienced learners tend to experience a lower ICL than their less experienced peers when confronted with the same content, instructional support that can reduce ECL among the latter may contribute to ECL among the former. Add to this the fact that, in the context of a given learning goal, a complex task may both constitute a higher ICL and trigger an ineffective problem-solving search (i.e., ECL) among less experienced than among more experienced learners,[4,5,10] and we come to realize that the relation between ICL and ECL may be non-linear and heavily context-dependent. This finding runs counter to the traditional conception that ICL and ECL are independent.

Context-dependence might, to some extent, also explain some of the considerable heterogeneity in the conceptualization of GCL. Young and Sewell[29] summarize it well: GCL has been defined by different research groups as related to task learning (in contrast to ICL which is related to task performance), as the conscious application of learning strategies such as comparing and contrasting,[49] and as depending on motivation, metacognitive skills, and other learner-related features. Moreover, others (for example Sweller and colleagues[9]) may link GCL to transfer of the learning material to other situations. Finally, although the conceptualizations of GCL mentioned thus far treat GCL as a third type of cognitive load, several scholars have suggested that, given the definition of learning in CLT (i.e., the development and automation of cognitive schemas), GCL should be redefined as working memory resources allocated to dealing with ICL[1−5,7,8,14,53] (i.e., as part of ICL) and have used that modified dual model in empirical studies.

Continued use of single-item mental effort ratings

Since the introduction of GCL in 1998,[9] many have treated ICL, ECL, and GCL as three additive and independent types of cognitive load, meaning that these three types of cognitive load added together form the total working memory load or cognitive load and each of these types of cognitive load can vary independently. Mental effort invested by a learner in a task or problem has been assumed to reflect the total working memory load or cognitive load.[50] This model, with three types of cognitive load forming the total cognitive load or mental effort, has been criticized for the following reasons.

First, single-item measurements simply do not meet the purpose. Not only are they − compared to multi-item measurements − unreliable,[51] they can never distinguish between different types of cognitive load.[52] Although some researchers have attempted to keep a particular type of cognitive load 'constant' in experimental design, no empirical support for the success of such an attempt has ever been provided. Moreover, using randomized controlled experiments may at best create conditions that are *on average* similar in a particular type of cognitive load but can never guarantee that different learners experience exactly the same ICL, the same ECL or − for that matter − the same GCL. Finally, several studies which asked participants to rate their mental effort and sets of items presumably related to the three types of cognitive load have reported that mental effort ratings are mainly if not exclusively a reflection of ICL.[13,24,28]

Despite the arguments against single-item mental effort ratings, this approach has enjoyed immense popularity, partly because single items are so easy to administer whereas other measures of cognitive load such as functional magnetic resonance imaging (fMRI),[53,54] electroencephalography (EEG),[55,56] eye-tracking,[57−60] and measures of secondary task performance[5,8,61] require expensive equipment that in many situations is difficult to use. Moreover, while self-report measures are inherently limited,[4,24] the aforementioned objective measures might, under certain assumptions, provide a measure of overall cognitive load but have not yet resulted in measures of types of cognitive load. Although there is legitimate disagreement about the role of GCL, both the two-factor and the three-factor model state that the ICL-ECL distinction is crucial for the

design of education and training.[1,2,4,5,10,11] In both models, learning can only be expected if ECL is minimized, and there is an ICL that stimulates learners to engage in learning. Hence, we need instruments that enable us to distinguish between ICL and ECL.

Timing of cognitive load and learning outcome measures

With the suggested change in CLT in which the focus shifts from learner to learner activity and specific learning goals determine what is ICL and what is ECL,[2] perhaps the questionnaires developed and used for the measurement of different types of cognitive load in recent years[12–14,22,24,26–28,62,63] fall short in that they fail to capture different sources of ICL and ECL (and GCL) in different contexts. Moreover, the fact that a given questionnaire yields a three-factor solution[13,14,27] does not mean that the three factors correspond to ICL, ECL, and GCL, and since the modified dual model does not deny the existence of GCL, finding three factors does not discard the modified dual model even if that third factor captures GCL. If we really want to investigate which of the two models is more plausible, we will probably need to administer tests for working memory capacity along with cognitive load and learning outcome measures in experiments that allow for careful variation in different types of cognitive load and where participants are motivated to use their working memory resources as much as possible. After all, if the combination of ICL and ECL is relatively high, additional GCL could result in cognitive overload. However, when that state of cognitive overload is reached depends on the limits of the individual learner's working memory and, obviously, motivation to invest a certain effort in the first place. In the two-factor model, where GCL is part of ICL, the total cognitive load is lower than the total cognitive load in the three-factor model unless a learner experiences zero GCL. In other words, in the case of a relatively high combination of ICL and ECL, cognitive overload should be slightly less likely in the two-factor than in the three-factor model.

Apart from the distinction between models on the role of GCL, if there is such a thing as GCL that can be captured with a questionnaire — such as attempted in recent years[13,14,24,27,28,62,63] — we should be able to find and replicate meaningful correlations between the factors that supposedly capture GCL and learning outcome measures.[64,65] Unfortunately, no such correlations have been found thus far. In this context, Young and Sewell[29] have made an important point: studies that have included a measurement of GCL[13,14] have generally administered that measurement fairly soon after a learning activity, leaving very little time for schema development or automation to occur. This may have created a restriction of range in GCL, and the latter is known to influence correlations of interest (i.e., more often towards than away from zero) and could thus partly account for the weak correlations between the supposed 'GCL' factor[13,14] and learning outcomes.

When we define learning as the development and automation of cognitive schemas (cf. CLT), learning is by definition a longitudinal phenomenon, in which types of cognitive load can vary with time. Unfortunately, however, the vast majority of studies of cognitive load and learning outcome measures administer each of these measures once in time, with cognitive load measures being administered either when learning (i.e., before performance) or after test performance.[4] Just as single-item measurements cannot distinguish between types of cognitive load, with one-time measurements we cannot separate variation within learners from variation between learners. Some studies have demonstrated that asking students to rate their mental effort multiple times during an activity tends to yield a lower average mental effort rating than asking a single mental effort rating at the end.[66,67] Moreover, these repeated measurements should not be averaged into a single rating but treated as is in multilevel models[68] or path models[69] to avoid ecological fallacies (e.g., a negative relation between a type of cognitive load and learning outcome appearing positive).

Conclusions

The introduction of CLT in medical education has helped move both medical education and CLT forward. We have seen a boom in empirical and theoretical work on CLT and its implications for medical education, and the medical domain provides notable opportunities for new research. Although the question on the distinction between different types of cognitive load remains a major challenge, the different models do support the same recommendations for education and training. The questions that call for further research should not discourage us from applying CLT to medical education but rather contribute to the excitement and motivation to advance that is already a key trademark of the medical education community.

Sources of support in the form of grants

Conflict of interest

The author has no conflict of interest to declare.

References

1. Kalyuga S. Cognitive load theory: how many types of load does it really need? **Educ Psychol Rev 2011**; 23: 1–19.
2. Kalyuga S, Singh AM. Rethinking the boundaries of cognitive load theory in complex learning. **Educ Psychol Rev 2015**. http://dx.doi.org/10.1007/s10648-015-9352-0.
3. Leppink J. Helping medical students in their study of statistics: a flexible approach. **J Taibah Univ Med Sci 2017**; 12: 1–7. http://dx.doi.org/10.1016/j.jtumed.2016.08.007.
4. Leppink J, Van den Heuvel A. The evolution of cognitive load theory and its application to medical education. **Perspect Med Educ 2015**; 4: 119–127.
5. Leppink J, Van Gog T, Paas F, Sweller J. Cognitive load theory: researching and planning teaching to maximise learning. In: Cleland J, Durning SJ, editors. *Researching medical education*. Chichester: Wiley & Blackwell; 2015. pp. 207–218.
6. Sweller J. Cognitive load during problem solving: effects on learning. **Cogn Sci 1988**; 12: 257–285.
7. Sweller J. Element interactivity and intrinsic, extraneous, and germane cognitive load. **Educ Psychol Rev 2010**; 22: 123–138.

8. Sweller J, Ayres P, Kalyuga S. *Cognitive load theory*. New York: Springer; 2011.

9. Sweller J, Van Merriënboer JJG, Paas F. Cognitive architecture and instructional design. **Educ Psychol Rev 1998**; 10: 251−296.

10. Van Merriënboer JJG, Sweller J. Cognitive load theory in health professions education: design principles and strategies. **Med Educ 2010**; 44: 85−93.

11. Young JQ, Van Merriënboer JJG, Durning SJ, Ten Cate O. Cognitive load theory: implications for medical education: AMEE Guide No. 86. **Med Teach 2014**; 36: 371−384.

12. Bergman EM, De Bruin ABH, Vorstenbosch MATM, Kooloos JGM, Puts GCWM, Leppink J, Scherpbier AJJA, Van der Vleuten CPM. Effects of learning content in context on knowledge acquisition and recall: a pretest-posttest control group design. **BMC Med Educ 2015**; 15: 133.

13. Leppink J, Paas F, Van der Vleuten CPM, Van Gog T, Van Merriënboer JJG. Development of an instrument for measuring different types of cognitive load. **Behav Res Methods 2013**; 45: 1058−1072.

14. Leppink J, Paas F, Van Gog T, Van der Vleuten CPM, Van Merriënboer JJG. Effects of pairs of problems and examples on task performance and different types of cognitive load. **Learn Instr 2014**; 30: 32−42.

15. Barouillet P, Bernardin S, Portrat S, Vergauwe E, Camos V. Time and cognitive load in working memory. **J Exp Psychol Learn Mem Cogn 2007**; 33: 570−585.

16. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. **Behav Brain Sci 2001**; 24: 152−153.

17. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. **Psychol Rev 1956**; 63: 81−97.

18. Peterson L, Peterson MJ. Short-term retention of individual verbal items. **J Exp Psychol 1959**; 58: 193−198.

19. Colbert-Getz JM, Baumann S, Shaffer K, Lamb S, Lindsley JE, Rainey R, Randall K, Roussel D, Stevenson A, Cianciolo AT, Maines T, O'Brien B, Westerman M. What's in a transition? An integrative perspective on transitions in medical education. **Teach Learn Med 2016**; 28: 347−352.

20. Fraser K, Huffman J, Ma I, Sobczak M, McIlwrick J, Wright B, McLaughlin K. The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trail. **Chest J 2014**; 145: 958−963.

21. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. **Med Educ 2012**; 46: 1055−1062.

22. Lafleur A, Côté L, Leppink J. Influences of OSCE design on students' diagnostic reasoning. **Med Educ 2015**; 49: 203−214.

23. Leppink J, Duvivier R. Twelve tips for medical curriculum design from a cognitive load theory perspective. **Med Teach 2016**; 38: 669−674.

24. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. **Med Educ 2015**; 49: 805−814.

25. Reedy GB. Using cognitive load theory to inform simulation design and practice. **Clin Simul Nurs 2015**; 11: 355−360.

26. Tremblay ML, Lafleur A, Leppink J, Dolmans DHJM. The simulated clinical environment: cognitive and emotional impact among undergraduates. **Med Teach 2017**; 39: 181−187. http://dx.doi.org/10.1080/0142159X.2016.1246710.

27. Sewell JL, Boscardin CK, Young JQ, Ten Cate O, O'Sullivan PS. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. **Med Educ 2016**; 50: 682−692.

28. Young JQ, Irby DM, Barilla-LaBarca ML, Ten Cate O, O'Sullivan PS. Measuring cognitive load: mixed results from a handover simulation for medical students. **Perspect Med Educ 2016**; 5: 24−32.

29. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. **Perspect Med Educ 2015**; 4: 107−109.

30. Young JQ, Ten Cate O, O'Sullivan PS, Irby DM. Unpacking the complexity of patient handoffs through the lens of cognitive load theory. **Teach Learn Med 2016**; 28: 88−96.

31. Sweller J, Chandler P. Why some material is difficult to learn. **Cogn Instr 1994**; 12: 183−223.

32. Sweller J, Chandler P, Tierney P, Cooper M. Cognitive load as a factor in the structuring of technical material. **J Exp Psychol 1990**; 119: 176−192.

33. Choi HH, Van Merriënboer JJG, Paas F. Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. **Educ Psychol Rev 2014**; 26: 225−244.

34. Sweller J, Cooper M. The use of worked examples as a substitute for problem solving in learning algebra. **Cogn Instr 1985**; 1: 59−89.

35. Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. **Educ Psychol 2003**; 38: 23−31.

36. Kalyuga S, Chandler P, Tuovinen J, Sweller J. When problem solving is superior to studying worked examples. **J Educ Psychol 2001**; 93: 579−588.

37. Lafleur A, Leppink J, Côté L. Clinical examination in the OSCE era : are we maintaining the balance between OS and CE? **BMJ Postgrad Med J 2017**. http://dx.doi.org/10.1136/postgradmedj-2016-134776.

38. Lafleur A, Laflamme J, Leppink J, Côté L. Task demands in OSCEs influence learning strategies. **Teach Learn Med 2017**. http://dx.doi.org/10.1080/10401334.2017.1282863.

39. Yudkowsky R, Otaki J, Lowenstein T, Riddle J, Nishigori H, Bordage G. A hypothesis-driven physical examination learning and assessment procedure for medical students: initial validity evidence. **Med Educ 2009**; 43: 729−740.

40. Plass JL, Homer BD, Kinzer CK. Foundations of game-based learning. **Educ Psychol 2015**; 50: 258−283.

41. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. **Med Teach 2011**; 33: 478−485.

42. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment and Kane's validity perspective. **Med Educ 2012**; 46: 38−48.

43. Van der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. Twelve tips for programmatic assessment. **Med Teach 2015**; 37: 641−646.

44. Artino AR, Brydges R, Gruppen LD. Self-regulated learning in healthcare profession education: theoretical perspectives and research methods. In: Cleland J, Durning SJ, editors. *Researching medical education*. Chichester: Wiley & Blackwell; 2015. pp. 155−166.

45. Kostons D, Van Gog T, Paas F. Training self-assessment and task-selection skills: a cognitive approach to improving self-regulated learning. **Learn Instr 2012**; 22: 121−132.

46. Bjork RA, Dunlosky J, Kornell N. Self-regulated learning: beliefs, techniques, and illusions. **Annu Rev Psychol 2013**; 64: 417−444.

47. Dunning D, Heath C, Suls JM. Flawed self-assessment: implications for health, education, and the workplace. **Psychol Sci Public Interest 2004**; 5: 69−106.

48. Dunning D, Johnson K, Ehrlinger J, Kruger J. Why people fail to recognize their own incompetence. **Curr Dir Psychol Sci 2003**; 12: 83−87.

49. Schnotz W, Kürschner C. A reconsideration of cognitive load theory. **Educ Psychol Rev 2007**; 19: 469−508.

50. Paas F, Tuovinen J, Tabbers H, Van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. **Educ Psychol 2003**; 38: 63−71.

51. Picho K, Artino AR. 7 deadly sins in educational research. **J Grad Med Educ 2016**; 8: 483−487.

52. Leppink J, Pérez-Fuster P. We need more replication research — a case for test-retest reliability. **Perspect Med Educ 2017**; 6. http://dx.doi.org/10.1007/s40037-017-0347-z.

53. Paas F, Ayres P, Pachman M. Assessment of cognitive load in multimedia learning: theory, methods and applications. In: Robinson DH, Schraw G, editors. *Recent innovations in educational psychology that facilitate student learning*. Charlotte: Information Age Publishing; 2008. pp. 11−35.

54. Whelan RR. Neuroimaging of cognitive load in instructional multimedia. **Educ Psychol Rev 2007**; 2: 1−12.

55. Antonenko P, Niederhauser DS. The effects of leads on cognitive load and learning in a hypertext environment. **Comput Hum Behav 2010**; 26: 140−150.

56. Antonenko P, Paas F, Grabner R, Van Gog T. Using electro-encephalography to measure cognitive load. **Educ Psychol Rev 2010**; 22: 425−438.

57. Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J. *Eye-tracking: a comprehensive guide to methods and measures*. Oxford: Oxford University Press; 2011.

58. Underwood G, Jebbert L, Roberts K. Inspecting pictures for information to verify a sentence: eye movements in general encoding and in focused search. **Q J Exp Psychol Sect A Hum Exp Psychol 2004**; 57A: 165−182.

59. Van Gog T, Jarodzka H. Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In: Azevedo R, Aleven V, editors. *International handbook of metacognition and learning technologies*. New York: Springer; 2013.

60. Van Gog T, Scheiter K. Eye tracking as a tool to study and enhance multimedia learning. **Learn Instr 2010**; 20: 95−99.

61. Brünken R, Plass JL, Leutner D. Direct measures of cognitive load in multimedia learning. **Educ Psychol 2003**; 38: 53−61.

62. Hadie SNH, Yusoff MSB. Assessing the validity of the cognitive load scale in a problem-based learning setting. **J Taibah Univ Med Sci 2016**; 11: 194−202.

63. Zukic M, Dapo N, Husremovic D. Construct and predictive validity of an instrument for measuring intrinsic, extraneous and germane cognitive load. **Univers J Psychol 2016**; 4: 242−248.

64. Leppink J. Cognitive load measures mainly have meaning when they are combined with learning outcome measures. **Med Educ 2016**; 50: 979.

65. Naismith LM, Cavalcanti RB. Measuring germane load requires correlation with learning. **Med Educ 2016**. http://dx.doi.org/10.1111/medu.13134.

66. Schmeck A, Opfermann M, Van Gog T, Paas F, Leutner D. Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. **Instr Sci 2015**; 43: 93.

67. Van Gog T, Kirschner PA, Kester L, Paas F. Timing and frequency of mental effort measurement: evidence in favour of repeated measurements. **Appl Cogn Psychol 2012**; 26: 833−839.

68. Leppink J, Van Merriënboer JJG. The beast of aggregating cognitive load measures in technology-based learning. **Educ Technol Soc 2015**; 18: 230−245.

69. Leppink J. On causality and mechanisms in medical education research: an example of path analysis. **Perspect Med Educ 2015**; 4: 66−72.