Journal of Taibah University Medical Sciences (2016) 11(3), 194–202

Taibah University

Journal of Taibah University Medical Sciences

www.sciencedirect.com

Educational Article

Assessing the validity of the cognitive load scale in a problem-based learning setting

Siti N.H. Hadie, MSc [a] and Muhamad S.B. Yusoff, PhD [b,*]

[a] Anatomy Department, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, Kubang Kerian, Kota Bharu, Kelantan, Malaysia

[b] Department of Medical Education, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, Kubang Kerian, Kota Bharu, Kelantan, Malaysia

Received 20 February 2016; revised 27 March 2016; accepted 1 April 2016; Available online 27 April 2016

الملخص

أهداف البحث: تم التحقق من صحة مقياس الاستيعاب المعرفي فقط في أطر التعلم غير القائم على حل المشكلات. بالتالي، أجريت هذه الدراسة لتقييم صلاحية هذا المقياس في بيئة التعلم القائم على حل المشكلات عن طريق اختبار صلاحية بنائه وثباته الداخلي.

طرق البحث: أجرينا دراسة مقطعية على ١٢٥طالبا بالسنة الأولى من كلية طب بعد محاضرة للتعلم القائم على حل المشكلات. وتم تحليل العامل التأكيدي لاختبار صلاحية بنائه باستخدام تحليل برمجيات إنشائية مهمة. كما تم تحديد الثبات الداخلي لهذه القائمة من خلال تحليل الفاعلية باستخدام برمجيات الحزمة الإحصائية للعلوم الاجتماعية.

النتائج: أتم ٩٣ طالب طب هذه الدراسة. وأظهر التحليل أن مقياس الثلاثة عوامل حقق مستوى قبول جيد من المؤشرات الصالحة، تعكس جودة صلاحية بنائه. وكان مقياس كرونباخ ألفا أكثر من ٠.٧ مما يشير إلى مستوى عال من الثبات الداخلي. كما حققت جميع البنود عامل استيعاب موحد أكثر من ٠.٥ مما يشير الى إسهامات عالية من المقاييس. وكان معدل الاستيعاب الفعلي للطلبة والتعلم الذاتي عاليا بينما كان الاستيعاب الاستثنائي ضعيفا. أشارت هذه النتيجة إلى أن الطلبة تعلموا جيدا خلال هذه المحاضرة بالرغم من صعوبة الارشادات.

الاستنتاجات: نظرا لأن مستوى الاستيعاب المعرفي ناتج مهم لكفاءة التصميم التعليمي، نوصي بأن المقياس المذكور، يعد صالحا وموثوقا، ويجب أن يستخدم مستقبلا إما كمقياس للبحوث أو كأداة تشخيصية لتقييم نتائج التعلم القائم على حل المشكلات.

الكلمات المفتاحية: نظرية الاستيعاب المعرفي؛ مقياس الاستيعاب المعرفي؛ التعلم القائم على حل المشكلات؛ الصلاحية؛ الثقة

* Corresponding address. Department of Medical Education, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, Kubang Kerian, Kota Bharu, Kelantan, Malaysia.

E-mail: msaiful_bahri@usm.my (M.S.B. Yusoff)

Peer review under responsibility of Taibah University.

ELSEVIER Production and hosting by Elsevier

Abstract

Objectives: The cognitive load scale has only been validated in non-problem-based learning settings. Hence, this study was conducted to assess the validity of this scale in a problem-based learning environment by testing its construct validity and internal consistency.

Methods: We conducted a cross-sectional study on 125 first-year medical students after a problem-based learning session. Confirmatory factor analysis was performed to test its construct validity using the Analysis of Moment Structure software. The internal consistency of this inventory was determined through reliability analysis using the Statistical Package for Social Sciences software (SPSS).

Results: A total of 93 medical students completed the inventory. The analysis showed that the three-factor scale attained an acceptable level of goodness-of-fit indices, indicating good construct validity. The scale's Cronbach's alpha was more than 0.7, indicating a high level of internal consistency. All of the items attained a standardized factor loading of more than 0.5, which indicated high contributions to the respective scales. The mean levels of students' intrinsic load and self-perceived learning were high, and the mean level of students' extraneous load was low. These findings signalled that students learned well during the session despite difficult instruction.

Conclusion: Because the cognitive load level is an important outcome for the efficiency of instructional design, we suggest that the aforementioned scale, which is valid and reliable, should be used in the future either as a research measurement or diagnostic feedback tool for problem-based learning evaluation.

1658-3612 © 2016 The Authors.

 http://dx.doi.org/10.1016/j.jtumed.2016.04.001

Keywords: Cognitive load theory; Cognitive load scale; Problem-based learning; Reliability; Validity

Introduction

Cognitive Load Theory (CLT) is an instructional theory that describes the occurrence of optimal learning when instructional material is designed in a manner that fits the function of human cognition.[1] The elements in CLT have been based on a strong foundation of working memory research,[2–5] and its principles have been proven to result in an efficient learning environment by dozens of empirical studies.[6–10] The main focus of CLT is to envisage the outcome performance of learners by taking into account the capacity and restrictions of human cognition while designing instructional material.[11] Human cognition includes a combination of working memory, which has a limited capacity for and duration of information processing, and long-term memory, which has an unlimited capacity for storing cognitive schemas and high automaticity.[12] Human working memory is central to this theory because it receives and converts information sent by sensory memory into cognitive schema and transfers it to long-term memory for storage.[3] Because the processes of schema construction and automation indicate learning, it is vital for instructional material to not exceed the limited capacity of working memory.[5] Hence, it is imperative for instructors to understand the different types of cognitive loads (e.g., the amount of information) that could contribute to working memory overload before implementing the theory into their instructional design.

Within the first two decades of its introduction, CLT identified three types of cognitive load: intrinsic, extraneous and germane loads.[12–14] Intrinsic load refers to the difficulty level of the instructional content, resulting from the amount of inter-correlation between essential elements in the instructional material.[15] However, one could have learned some of these elements prior to instruction; hence, this load could be affected by the learners' prior knowledge.[16] Extraneous load is the unnecessary load imposed by poorly designed instruction. In other words, this load occurs when distracting elements are introduced that are not related to the content of instruction.[17] Extraneous load refers to the unnecessary load that is commonly associated with how a teacher prepares and delivers the instruction.[15] These two types of loads strongly contribute to the theoretical framework of CLT and are commonly manipulated to produce an efficient learning environment.

By contrast, germane load, which was introduced at a later stage of the theory, refers to the mental effort that is consciously invested by the learners while processing elements of the intrinsic load.[12] Hence, it is the 'good load' that needs to be maximized during learning. However, there is a lack of concordance between the theoretical explanation and empirical findings on the existence of germane load in previous research, thus leading to the reconceptualization of this theory.[17–21] The new dimension of CLT excludes germane load from the framework due to its vague predictive and explanatory roles in schema construction and automation, as reported by many studies,[22–28] and this dimension was further strengthened by the fact that germane load could not be distinguished from the intrinsic load and measured independently.[19]

Measuring the cognitive loads after an instructional process provides valuable information on the mental activities of the learners. The information obtained from this measurement could shed some light on the issue of variability in learning outcomes that exists between different learners despite receiving similar instruction or different formats of instruction.[18] In general, measurement of the cognitive load could be conducted either subjectively, by measuring the learners' self-rating on perceived mental effort, or objectively, by investigating outcome variables, such as task performance; input variables, such as task difficulty; and process-related behavioural variables, such as psychophysiological measures.[11,29] Due to its objectivity and the validity of the measurement, the latter approach is more favoured among cognitive researchers.[13,30–36] However, the objective measure of cognitive load has some drawbacks. Apart from being a time-consuming measure, its usability is limited to research settings because it requires the recruitment of 'suitable' subjects who can complete the tasks to be measured. Furthermore, the objective measure of cognitive load does not make a distinction between the different types of loads, of which the information is important for future teaching and learning improvement.[14] These shortcomings are often addressed by the use of cognitive load subjective rating scales.

There are several versions of subjective rating scales for cognitive load measurements. The early versions of these scales are unidimensional tools that measure either the total cognitive load or individual types of cognitive loads through the use of semantic differential scales.[16,33,37,38] Despite the fact that they are easy to implement and do not interfere with other task performance, the validity of these ratings was only based on two assumptions[1]: the learners are capable of cogitating their mental load during learning and[2] there is direct correlation between the students' self-rating and the actual level of cognitive load.[16,33,37] Therefore, the validity and reliability of these measures are uncertain and have been addressed in several studies.[21,39–41] In addition, research at a later stage revealed that the unidimensional tools, when combined to simultaneously measure different types of cognitive load, failed to distinguish each load as a separate entity.[32,33] This led to a new dimension of research by Leppink at el.,[18] who introduced a three-factor with ten-item instrument termed the cognitive load scale that was developed through principal component and confirmatory factor analyses. A subsequent study by Leppink at el.[19] revealed that this scale was able to differentiate between intrinsic and extraneous loads, but not germane load. Hence, their findings support the reconceptualization of the CLT with regard to the existence of germane load.

Subsequently, the cognitive load scale has been used in several studies to measure the two different types of cognitive load in different academic contexts, which include linguistics, statistics, psychology, computer science, social and health sciences, and medicine.[42–44] The third previously unnamed factor was termed the 'self-perceived learning domain' because the four items of this domain reflect students' perceptions of how much the instruction enhanced their understanding of the topic.[44] Morrison et al.,[42] who studied the level of cognitive loads in a computer science context, conducted a confirmatory factor analysis on the best-fit model proposed by Leppink et al.[19] and reported the same model fit. This indicated that the cognitive load scale has good construct validity across different academic contexts. Apart from that, Leppink et al.[19] reported the high internal consistency of each factor, with Cronbach's alpha ranging from 0.785 to 0.947, reflecting the high reliability of the measure.

Nonetheless, the construct validity of this scale across different modes of instructional delivery has yet to be explored. Since its development, this scale has been commonly used in a lecture-based instructional setting, although its use is not limited to this setting. One exception is a study by Bergman et al.,[44] who measured the cognitive loads of the learners using this scale after an anatomy dissection class. The study revealed interesting findings that might have implications for the future role of clinically applied contexts in learning.[44] The modes of instructional design in medical education are moving from conventional, didactic teaching methods (e.g., lecture and tutorial) towards problem-based learning (PBL).[45] In PBL, group members focus on the process of discovery via clinical scenarios as triggers for learners to identify their own learning issues, brainstorm on key issues and perform independent learning through research. Afterwards, the results are presented to the group and refined in group discussions.[46] In comparison to non-PBL (i.e., conventional, didactic teaching methods), PBL learners acquire knowledge in an active and self-directed way, through integration of basic and clinical science subjects with minimal guidance from PBL tutors.[45,46] We anticipated that similar problems as those mentioned by Bergman et al.[44] will be encountered in PBL instruction, as this type of learning requires students to learn within a clinical context despite being novices. As suggested by Qiao et al.[47] in their review, cognitive load theory should be explored and its principles should be adopted during PBL sessions to lessen the mental burden of medical students during learning. As they highlighted, PBL might seem to be an interesting self-learning strategy and is accepted widely by medical educators, but these learning sessions might impose a high cognitive load on students who lack experience in clinical reasoning.[47] To verify their concern, we attempted to investigate the construct validity (i.e., convergent and discriminant) and internal consistency of the CLS among medical students in a PBL setting. In addition, as preliminary data, we attempted to measure the intrinsic load, extraneous load and self-perceived learning after a PBL session.

Materials and Methods

Study design and ethical clearance

We conducted a cross-sectional study at a public Malaysian medical school after receiving permission from the institution's Human Research Ethics Committee. Participation was on a voluntarily basis.

Participants' background

This study involved 125 first-year medical students during the 2015/2016 academic session who underwent SPICES (i.e., student-centred, problem-based, integrated, community-based, elective, systematic and spiral) medical curriculum in the School of Medical Sciences, Universiti Sains Malaysia. The medical school has implemented the SPICES curriculum since its inception in 1979 and was the first to offer PBL in Asia.[48] In the first year, students were divided into 15 PBL groups (i.e., 8–9 students per group) and went through a total of 25 PBL sessions. Each PBL session consisted of 2 small group meetings of 2 h duration, usually scheduled for the beginning and end of each week.

Sample size and sampling method

The sample size was calculated based on the recommended ratio of 5–10 subjects per item[49]; consequently, 50 to 100 participants is an adequate sample size for testing the construct validity of a 10-item scale. Purposive sampling was employed to select participants, and verbal consent was sought from the participants prior to the study. All of the first-year medical students were invited to participate in this study after a PBL session. None of the participants' personal profiles (e.g., gender, religion, ethnic groups) were obtained to ensure their anonymity. Medical students who agreed to participate were asked to respond to the scale immediately after the PBL session ended.

The cognitive load scale

The CLS is a ten-item inventory that was developed and validated by Leppink et al.[18,19] Written permission was obtained from its developer through e-mail prior to the study. The CLS uses a ten-point semantic rating scale ranging from 'not all of the case' to 'completely the case', which measures participant's subjective ratings of the listed items: items 1, 2 and 3 measure the intrinsic load; items 4, 5 and 6 measure the extraneous load; and the remaining items measure the students' self-perceived learning. The inventory is provided in Appendix 1.

The authors provided a short briefing (less than 5 min) on the CLS and informed the participants to complete the CLS, which was expected to take less than 5 min, immediately after the PBL session. Participants were asked to hand over the completed CLS form to the authors immediately after the PBL session.

Data analysis

The psychometric properties of the CLS were evaluated by confirmatory factor analysis (CFA) using the Analysis of Moment Structure (AMOS) software. Goodness-of-fit indices were assessed to support the latent constructs of the CLS. These indices are summarized in Table 1. The latent construct was considered to fit the model if it achieved the acceptance level.

Table 1: Goodness of fit indices that were used to signify model fit.

Name of category	Name of index	Level of acceptance
Absolute fit[a]	Root Mean Square of Error Approximation (RMSEA)	less than 0.08[72]
	Goodness of Fit Index (GFI)	more than 0.9[73]
Incremental Fit[b]	Comparative Fit Index (CFI)	more than 0.9[74]
	Tucker–Lewis Index (TLI)	more than 0.9[75]
	Normed Fit Index (NFI)	more than 0.9[76]
Parsimonious fit[c]	Chi Square/Degree of Freedom (Chisq/df)	less than 5[77]

[a] Absolute Fit: Measures the overall goodness-of-fit for both the structural and measurement models collectively. This type of measure does not make any comparison to a specified null model (incremental fit measure) or adjust for the number of parameters in the estimated model (parsimonious fit measure).

[b] Incremental Fit: Measures goodness-of-fit that compares the current model to a specified "null" (independence) model to determine the degree of improvement over the null model.

[c] Parsimonious Fit: Measures goodness-of-fit representing the degree of model fit per estimated coefficient. This measure attempts to correct for any "overfitting" of the model and evaluates the parsimony of the model compared to the goodness-of-fit.

The contributions of the observed variables (i.e., the items of CLS) to the latent variables (i.e., constructs) were estimated by standardized factor loadings – high factor loadings indicate a high contribution of the item to the construct.[50] Modification Indices (MI) were created that estimated the correlations between variables, and a reduction of chi-square values was seen if these correlations added to the model fit.[50] Any observed variables should have standardized residual covariance (SRC) values of less than 2 to signify that the model is correct.[50,51] Therefore, observed variables were retained in the model if they met the acceptable values of MI, SRC and standardized factor loading.[50] Observed variables should only be removed based on a theoretical basis or with a literature review.[52–54]

Cronbach's alpha coefficient measured reliability with Statistical Package for Social Sciences (SPSS) software to signify the internal consistency. Cronbach's alpha values of more than 0.7 were considered to be high internal consistency, and values between 0.6 and 0.7 were considered to be satisfactory internal consistency.[55]

Construct validity was assessed through the assessment of convergent validity and discriminant validity, which reflected the internal structure of any constructs.[56] Convergent validity was checked with the size of factor loading, average variance extracted (AVE) and composite reliability (CR). The item factor loading values should be reasonably high (0.5 or more) on respective constructs to signify convergent validity.[57] The authors calculated AVE and CR manually based on the recommendations of previous studies.[57,58] Convergent validity was achieved if the AVE values were more than 0.5 and the CR values were more than 0.6.[54,57] Discriminant validity of a construct was tested by comparing its shared variance (SV) and AVE values. SV is given as the square of the correlation between two constructs.[59] AVE values higher than SV values signified an acceptable level of discriminant validity.[58] A correlation between constructs of less than 0.85 was considered to be good discriminant validity.[50]

Results

Table 2 summarizes the results of the CFA. The analysis revealed that one-factor model of CLS failed to achieve a model fit, indicating that CLS has multiple constructs. The results showed that the original three-factor model with 10 items achieved acceptable values on the goodness-of-fit indices, suggesting a good model fit. In addition, the correlation values between the constructs were less than 0.85, indicating good discriminant validity, as illustrated in Figure 1. The correlation between the intrinsic load and self-perceived learning was 0.17, between the intrinsic and extraneous loads was 0.13, and between the extraneous load and self-perceived learning was −0.31.

The reliability analysis confirmed that the final model showed a high level of internal consistency, as the Cronbach's alpha was greater than 0.7 (Table 3). The composite reliability values of the CLS constructs ranged between 0.83 and 0.95, as shown in Table 4, indicating a high level of convergent validity. All of the standardized factor loadings were more than 0.5, suggesting an adequate level of convergent validity.[57] In addition, the AVE value of each construct was more than its SV value, indicating a good level of discriminant validity.

Table 5 shows the mean and standard deviation of each construct. The intrinsic load score was more than average, indicating that a high level of PBL content complexity was perceived by the students. The extraneous load score was

Table 2: The results of confirmatory factor analysis of CLS.

Variable	X^2 – statistic (df)	p-value	Goodness of fit indices					
			ChiSq/df	RMSEA	GFI	CFI	NFI	TLI
One-factor model[a]	295.003 (35)	<0.001	8.429	0.284	0.619	0.650	0.625	0.549
3-factor model[a]	36.885 (32)	0.253	1.153	0.041	0.929	0.993	0.953	0.991

[a] The original construct of the CLS was supported for a model fit.

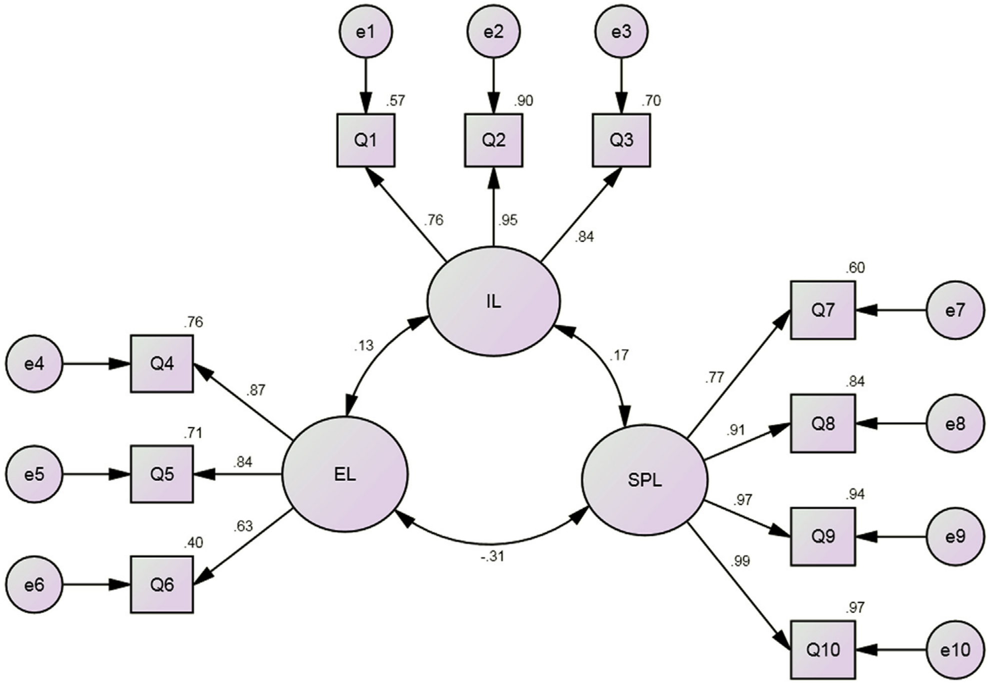

Figure 1: Standardized factor loading of the CLS constructs based on the final model.

Table 3: The reliability analysis of the 10 items of the CLS based on the final model.

No	Item	Standardized factor loading	[b]Domain	[a]Cronbach's Alpha	[c]AVE	[d]CR
1	The topics covered in the PBL were very complex.	0.76	Intrinsic load	0.88	0.73	0.89
2	The PBL covered terminologies that I perceived as very complex.	0.95				
3	The PBL covered concepts and definitions that I perceived as very complex.	0.84				
4	The instructions and explanations during the PBL were very unclear.	0.87	Extraneous load	0.82	0.62	0.83
5	The instructions and explanations during the PBL were full of unclear language.	0.84				
6	The instructions and explanations during the PBL were, in terms of learning, very ineffective.	0.63				
7	The PBL really enhanced my understanding of the topics covered.	0.77	Self-perceived learning	0.95	0.84	0.95
8	The PBL really enhanced my understanding of the terminologies covered.	0.91				
9	The PBL really enhanced my knowledge of concepts and definitions.	0.97				
10	The PBL really enhanced my knowledge and understanding of the subject.	0.99				

[a] Reliability analysis; Cronbach's alpha coefficient.
[b] Domains were predetermined based on a previous study.
[c] AVE (Average Variance Extracted) was calculated manually based on formula given by Fornell & Larcker.[51]

$$AVE = \frac{\sum_{i=1}^{n} \lambda_i^2}{n} \quad \lambda = \text{standardized factor loading} \quad n = \text{number of item}$$

[d] CR (Composite Reliability) was calculated based on formula given by Fornell & Larcker.[51]

$$CR = \frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2}{\left(\sum_{i=1}^{n} \lambda_i\right)^2 + \left(\sum_{i=1}^{n} \delta_i\right)} \quad \lambda = \text{standardized factor loading}, \quad \delta = \text{error variance}$$

Table 4: AVE and SV of the CLS based on the final model.

Factors	AVE	SV by factor		
		IL	EL	SPL
Intrinsic load	0.73	1	0.017	0.029
Extraneous load	0.62	0.017	1	0.096
Self-perceived learning	0.84	0.029	0.096	1

Table 5: The mean and standard deviation of the CLS scores.

CLS domain	N	Mean	Std. Deviation
Intrinsic load	93	6.33	1.83
Extraneous load	93	3.72	2.06
Self-perceived learning	93	7.05	1.89

below average, indicating minimal exposure to distracting elements during the PBL session. The self-perceived learning score was more than average, indicating that the students perceived the PBL as an interesting session, which motivated them to learn.

In summary, our results showed that the three-factor CLS achieved an acceptable level of construct (i.e., convergent and discriminant) validity with a high level of internal consistency. The psychometric credentials and usability of this scale as well as the future implications of the CLT on PBL instruction are discussed in the next section.

Discussion

Since its establishment as a means to measure cognitive loads, the psychometric properties of CLS in lecture-based instruction have been evaluated in several studies.[18,19,42] These studies reported that CLS has good construct validity with a high level of internal consistency.[18,19,42] Our study, although conducted in a different context of instruction, revealed similar findings; thus, our study strengthened the psychometric credentials of this inventory across different modes of instructional delivery.

The three-factor model of CLS has an appropriate latent construct to measure cognitive loads as the goodness-of-fit indices were attained – indicating an acceptable level of construct validity.[55] The reliability analysis showed that there was high internal consistency of CLS components, as the Cronbach's alpha values were greater than 0.7.[55] Similar to Leppink et al.,[18,19] we found that the correlations between CLS constructs were independent, as the correlation coefficients were less than 0.85, suggesting good discriminant validity.[50] This finding demonstrates that CLS successfully measured the different types of mental loads and thus supports the multidimensional concept of CLT.

Our preliminary data showed that the PBL session imposed a high intrinsic load and a low extraneous load to the students. In addition, the students' self-perceived learning scores were noted to be high. These results suggest that the students invested more mental effort processing the intrinsic load than the extraneous load. Aligned with the reconceptualization of germane load, the invested mental effort might be consciously allocated by students during the PBL session, as evidenced by their high self-perceived learning score.[17] These findings fit the concept of PBL in medical school because students were exposed to difficult clinical-based triggers despite being novices,[60,61] students have autonomy in learning because they were able to decide the content of their discussion and deliver information in a manner that could be easily understood by their peers,[62] and students could learn at their own pace with the appropriate use of available resources.[63]

Despite the fact that the PBL method is constructed based on a solid framework consisting of several learning theories and approaches,[60,64–66] there were some concerns raised by educators with regard to the outcome of PBL, particularly regarding the inadequate self-perceived knowledge acquisition demonstrated by students who underwent PBL curriculum.[63,67] In addition, some authors highlighted the lack of empirical evidence to support the effectiveness of the PBL approach on improving the knowledge base, clinical competency and clinical consultation.[68,69] There are issues addressed by cognitive researchers with regards to the implications of minimal guidance during PBL instruction because the information searching strategies conducted by students imposes heavy mental loads on their working memory.[70] Due to the use of fully occupied working memory resources for information searching, schema acquisition would be hampered; hence, less information is stored in long-term memory.[70] Although their assumption was supported by empirical evidence in cognitive research, it has been disputed by constructivists who favour PBL.[71] These arguments could only be resolved by conducting a well-designed experimental study that compares all of the outcome variables measures, including the cognitive load levels, between PBL and other types of instruction, as suggested by Berkson (1993) in her review:

> *"Because no one has yet been able to characterize, and, therefore measure, the cognitive components that make up problem solving, a direct answer to the question of whether PBL teaches problem solving better that traditional school is unavailable."*

Hence, this study supports the need for more research to explore students' mental workload during PBL session. Understanding students' cognitive load level would provide better insight for PBL tutors on how to manage the cognitive loads of their students without jeopardizing the concept of PBL.

Future research should contribute towards a better understanding of how CLT could be adopted in a constructivist framework of PBL. Measuring the cognitive loads of students during PBL sessions using the CLS could be a good start toward implementing this work. Measuring students' cognitive loads should not be limited to one or several PBL sessions, but should be measured longitudinally across different durations, and the CLS should be used as a diagnostic feedback measure before making any improvisations to PBL instruction. The fact that medical students needed less than five minutes to complete the CLS indicates the feasibility of administering this measure after a PBL session.

We acknowledge several limitations of our study. First, this study was conducted at a medical school in Malaysia, so the findings might not be generalizable to other institutions. A multi-centre study is recommended to verify the present findings. Second, this study was conducted after one PBL session, which might not completely reflect the respondents' judgments regarding "the PBL". "The PBL" could refer to a session as well as to respondents' experiences with PBL in a broader context. Moreover, the fact that respondents were nested within learning groups may have influenced their responses to some extent; it is known that students from the same group tend to yield more similar responses than random students from different groups. Therefore, future research should measure cognitive loads longitudinally across different PBL sessions to verify the psychometric credentials of the CLS constructs. Third, these data assessed the construct validity of the questionnaire in terms of its internal structure with regard to the convergent and discriminant validity of the measured constructs. Further study would benefit from an additional randomized controlled experiment that uses the current questionnaire and manipulates an element in the PBL session (experimental vs. control condition) that is known to make a difference in terms of either the intrinsic or extraneous cognitive load. This will help to establish empirical support for the validity of the questionnaire that might help the field and advance the theory. Finally, the sample size was relatively small, so the findings should be interpreted with caution. Future research should involve a larger study cohort.

Conclusions

This study supports the construct validity, reliability, feasibility and significance of the CLS as a tool to measure cognitive loads of medical students in a PBL setting. We believe that the CLS could be used as a PBL evaluation feedback tool. By measuring medical students' cognitive loads together with other outcome variables of PBL, educators could provide high-quality evidence on the strengths and weaknesses of PBL as well as a strong justification for improving PBL instruction, if required, for future implementation.

Practice points

- The cognitive load level is an important outcome for the efficiency of instruction.
- Cognitive load measurements during PBL provide insight into how much students learned during the session.
- The strengths and weaknesses of PBL can be determined through the cognitive load measure.
- Cognitive loads can be measured using the cognitive load scale.
- The cognitive load scale is a construct-valid and reliable tool.

Conflict of interest

The authors have no conflict of interest to declare.

Authors' contributions

SNHH conceived and designed the study, provided research materials, wrote the initial and final drafts of the article, and provided logistic support. SNHH has critically reviewed and approved the final draft and is responsible for the content and similarity index of the manuscript. MSBY administered the research materials, collected and organized data, analysed and interpreted data, wrote the initial and final drafts of the article, and provided logistic support. MSBY has critically reviewed and approved the final draft and is responsible for the content and similarity index of the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jtumed.2016.04.001

References

1. Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. **Educ Psychol 2003**; 38(1): 1–4.
2. Baddeley AD, Hitch G. Working memory. In: Gordon HB, editor. *Psychology of learning and motivation*. Academic Press; 1974. pp. 47–89.
3. Baddeley A. The episodic buffer: a new component of working memory? **Trends Cogn Sci 2000 11/1**; 4(11): 417–423.
4. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. **Psychol Rev 1956**; 63(2): 81.
5. Cowan N. The magical mystery four: how is working memory capacity limited, and why? **Curr Dir Psychol Sci 2010**; 19(1): 51–57.
6. Clark R, Nguyen F, Sweller J. *Efficiency in learning: evidence-based guidelines to manage cognitive load*. San Francisco: John Wiley & Sons; 2006.
7. Agostinho S, Tindall-Ford S, Roodenrys K. Adaptive diagrams: handing control over to the learner to manage split-attention online. **Comput Educ 2013 5**; 64(0): 52–62.
8. Leahy W, Sweller J. Cognitive load theory, modality of presentation and the transient information effect. **Appl Cogn Psychol 2011**; 25(6): 943–951.
9. Mousavi SY, Low R, Sweller J. Reducing cognitive load by mixing auditory and visual presentation modes. **J Educ Psychol 1995**; 87(2): 319.
10. Tabbers HK, Martens RI, Merriënboer JJ. *Multimedia learning and cognitive load theory: effects of modality and cueing symposium on cognitive load theory*; 2000. pp. 1–11.
11. Plass JL, Moreno R, BrÜnken R. Cognitive load theory. In: Plass JL, Moreno R, BrÜnken R, editors. New York: Cambridge University Press; 2010. pp. 1–5.
12. Sweller J, Van Merrienboer JJ, Paas FG. Cognitive architecture and instructional design. **Educ Psychol Rev 1998**; 10(3): 251–296.
13. Sweller J. Cognitive load during problem solving: effects on learning. **Cogn Sci 1988**; 12(2): 257–285.
14. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. **Educ Psychol 2003/03/01**; 38(1): 63–71.
15. Sweller J, Chandler P. Why some material is difficult to learn. **Cogn Instr 1994**; 12(3): 185–233.
16. Kalyuga S, Chandler P, Sweller J. Levels of expertise and instructional design. **Hum Factors J Hum Factors Ergon Soc March 1, 1998**; 40(1): 1–17.
17. Kalyuga S. Cognitive load theory: how many types of load does it really need? **Educ Psychol Rev 2011/03/01**; 23(1): 1–19.
18. Leppink J, Paas F, Van der Vleuten CM, Van Gog T, Van Merriënboer JG. Development of an instrument for measuring

different types of cognitive load. **Behav Res 2013/12/01**; 45(4): 1058−1072.
19. Leppink J, Paas F, Van Gog T, van Der Vleuten CP, Van Merrienboer JJ. Effects of pairs of problems and examples on task performance and different types of cognitive load. **Learn Instr 2014 4**; 30: 32−42.
20. de Jong T. Cognitive load theory, educational research, and instructional design: some food for thought. **Instr Sci 2010**; 38(2): 105−134.
21. Schnotz W, Kürschner C. A reconsideration of cognitive load theory. **Educ Psychol Rev 2007/12/01**; 19(4): 469−508.
22. Van Gog T, Paas F, Van Merriënboer JJ. Process-oriented worked examples: improving transfer performance through enhanced understanding. **Instr Sci 2004**; 32(1−2): 83−98.
23. Große CS, Renkl A. Effects of multiple solution methods in mathematics learning. **Learn Instr 2006**; 16(2): 122−138.
24. Van Gog T, Paas F, van Merriënboer JJ. Effects of process-oriented worked examples on troubleshooting transfer performance. **Learn Instr 2006**; 16(2): 154−164.
25. Bannert M, Hildebrand M, Mengelkamp C. Effects of a metacognitive support device in learning environments. **Comput Hum Behav 2009**; 25(4): 829−835.
26. Van Gog T, Jarodzka H, Scheiter K, Gerjets P, Paas F. Attention guidance during example study via the model's eye movements. **Comput Hum Behav 2009**; 25(3): 785−791.
27. de Croock MB, van Merriënboer JJ, Paas FG. High versus low contextual interference in simulation-based training of troubleshooting skills: effects on transfer performance and invested mental effort. **Comput Hum Behav 1998**; 14(2): 249−267.
28. van Merriënboer JJG, Schuurman JG, de Croock MBM, Paas FGWC. Redirecting learners' attention during training: effects on cognitive load, transfer test performance and training efficiency. **Learn Instr 2002 2**; 12(1): 11−37.
29. Galy E, Cariou M, Mélan C. What is the relationship between mental workload factors and cognitive load types? **Int J Psychophysiol 2012 3**; 83(3): 269−275.
30. Chevalier A, Kicka M. Web designers and web users: influence of the ergonomic quality of the web site on the information search. **Int J Human-Computer Stud 2006**; 64(10): 1031−1048.
31. Cegarra J, Chevalier A. The use of Tholos software for combining measures of mental workload: toward theoretical and methodological improvements. **Behav Res 2008**; 40(4): 988−1000.
32. DeLeeuw KE, Mayer RE. A comparison of three measures of cognitive load: evidence for separable measures of intrinsic, extraneous, and germane load. **J Educ Psychol 2008**; 100(1): 223.
33. Cierniak G, Scheiter K, Gerjets P. Explaining the split-attention effect: is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? **Comput Hum Behav 2009**; 25(2): 315−324.
34. Brunken R, Plass JL, Leutner D. Direct measurement of cognitive load in multimedia learning. **Educ Psychol 2003**; 38(1): 53−61.
35. Chandler P, Sweller J. Cognitive load while learning to use a computer program. **Appl Cogn Psychol 1996**; 10(2): 151−170.
36. Marcus N, Cooper M, Sweller J. Understanding instructions. **J Educ Psychol 1996**; 88(1): 49.
37. Ayres P. Using subjective measures to detect variations of intrinsic cognitive load within problems. **Learn Instr 2006 10**; 16(5): 389−400.
38. Paas FG. Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. **J Educ Psychol 1992**; 84(4): 429.
39. van Gog T, Paas F. Instructional efficiency: revisiting the original construct in educational research. **Educ Psychol 2008/01/28**; 43(1): 16−26.
40. Debue N, van de Leemput C. What does germane load mean? An empirical contribution to the cognitive load theory. **Front Psychol 2014-October-1**; 5 [Original Research].
41. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. **Med Educ 2015**; 49(8): 805−814.
42. Morrison BB, Dorn B, Guzdial M. Measuring cognitive load in introductory CS: adaptation of an instrument. In: *Proceedings of the tenth annual conference on International computing education research; Glasgow, Scotland, United Kingdom*; 2014. pp. 131−138. 2632348: ACM.
43. Leppink J, van deb Heuvel A. The evolution of cognitive load theory and its application to medical education. **Perspect Med Educ 2015**; 4(192).
44. Bergman E, de Bruin AB, Vorstenbosch MA, Kooloos JG, Puts GC, Leppink J, et al. Effects of learning content in context on knowledge acquisition and recall: a pretest-posttest control group design. **BMC Med Educ 2015**; 15(133): 2−11.
45. Lewis AD, Menezes DAB, McDermott HE, Hibbert LJ, Brennan S-L, Ross EE, et al. A comparison of course-related stressors in undergraduate problem-based learning (PBL) versus non-PBL medical programmes. **BMC Med Educ 2009**; 9(1): 1−8.
46. Srinivasan M, Wilkes M, Stevenson F, Nguyen T, Slavin S. Comparing problem-based learning with case-based learning: effects of a major curricular shift at two institutions. **Academic Medicine 2007**; 82(1): 74−82.
47. Qiao YQ, Shen J, Liang X, Ding S, Chen FY, Shao L, et al. Using cognitive theory to facilitate medical education. **BMC Med Educ 2014**; 14: 79.
48. Servant V. Some observations on a hybrid PBL curriculum in medical education at Universiti Sains Malaysia. **Educ Med J 2013**; 5(4): e67−e71.
49. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical assessment. **Res Eval 2005**; 10(7): 1−9.
50. Brown TA. *Confirmatory factor analysis for applied sciences*. New York: The Guilford Press; 2006.
51. Arbuckle JL. *IBM SPSS Amos 19 user's guide*. Chicago, US: IBM; 1995. Available from: http://www.amosdevelopment.com/download/amos.pdf.
52. Kline RB. *Principles and practice of structural equation modeling*. 3rd ed. New York: Guilford Publications; 2010.
53. Piaw CY. *Statistik Penyelidikan Lanjutan*. Malaysia: McGraw Hill; 2009.
54. Zainudin A. *Structural equation modeling using AMOS graphic*. Shah Alam: UiTM Press; 2012.
55. Streiner LD, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th ed. New York: Oxford University Press; 2008.
56. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. **Am J Med 2006**; 119(2). 166. e7−166. e16.
57. Hair JJ, Black W, Babin B, Anderson R. *Multivariate data analysis*. Upper Saddle River, NJ: Pearson Prentice-Hall; 2009.
58. Fornell C, Larcker D. Evaluating structural equation models with unobservable variables and measurement error. **J Mark Res 1981**; 18(1): 39−50.
59. Arifin WN, Yusoff MSB, Naing NN. Confirmatory factor analysis (CFA) of USM Emotional Quotient Inventory (USMEQ-i) among medical degree program applicants in Universiti Sains Malaysia (USM). **Educ Med J 2012**; 4(2).
60. Savery JR, Duffy TM. Problem based learning: an instructional model and its constructivist framework. **Educ Technol 1995**; 35(5): 31−38.
61. Collins A, Brown JS, Newman SE. *Cognitive apprenticeship: teaching the crafts of reading, writing, and mathematics*In **Knowing, learning, and instruction: essays in honor of Robert Glaser**, 18; 1989. pp. 32−42.
62. Dolmans D, Schmidt H. The advantages of problem-based curricula. **Postgrad Med J September 1, 1996**; 72(851): 535−538.

63. Albanese MA, Mitchell S. Problem-based learning: a review of literature on its outcomes and implementation issues. **Acad Med 1993**; 68(1): 52−81.
64. Schmidt HG. Problem-based learning: rationale and description. **Med Educ 1983**; 17(1): 11−16.
65. Coles C. *Evaluating the effects curricula have on student learning: toward a more competent theory for medical education. Innovation in medical education: an evaluation of its present status*; 1990. pp. 76−87.
66. Barrows HS. A taxonomy of problem-based learning methods. **Med Educ 1986**; 20(6): 481−486.
67. Prince KJAH, van de Wiel M, Scherpbier AJJA, can der Vleuten CPM, Boshuizen HPA. A qualitative analysis of the transition from theory to practice in undergraduate training in a PBL-medical school. **Adv Health Sci Educ 2000**; 5(2): 105−116 [Article].
68. Kilroy D. Problem based learning. **Emerg Med J 2004**; 21(4): 411−413.
69. Colliver JA. Effectiveness of problem-based learning curricula: research and theory. **Acad Med 2000**; 75(3): 259−266.
70. Kirschner PA, Sweller J, Clark RE. Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. **Educ Psychol 2006/06/01**; 41(2): 75−86.
71. Hmelo-Silver CE, Duncan RG, Chinn CA. Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). **Educ Psychol 2007**; 42(2): 99−107.
72. Browne MW, Cudeck R. Alternative ways of assessing model fit. **Sociol Methods Res 1992**; 21(2): 230−258.
73. Jöreskog K, Sörbom D. *LISREL VI users guide*. 3rd ed. Moorsville, IN: Scientific Software; 1984.
74. Bentler PM. Comparative fit indexes in structural models. **Psychol Bull 1990**; 107(2): 238.
75. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. **Psychol Bull 1980**; 88(3): 588.
76. Bollen KA. A new incremental fit index for general structural equation models. **Sociol Methods Res 1989**; 17(3): 303−316.
77. Marsh HW, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: first-and higher order factor models and their invariance across groups. **Psychol Bull 1985**; 97(3): 562.

How to cite this article: Hadie SNH, Yusoff MSB. Assessing the validity of the cognitive load scale in a problem-based learning setting. J Taibah Univ Med Sc 2016;11(3):194−202.