Adapting differentiated cognitive load measurement in physics classroom

Xian Wu and Diego Valente
Department of Physics, University of Connecticut, Storrs, Connecticut 06269, USA

Cognitive Load Theory is a very influential framework in educational psychology. There have been numerous efforts recently to develop and implement a differentiated way to measure the three major independent aspects of cognitive load, namely intrinsic, extraneous, and germane cognitive load to gain insight on how students learn from educational interventions. This study adapts a validated self-reported questionnaire to measure the three aspects of cognitive load on students completing administered in-class tutorials. It verifies the feasibility of such a measurement strategy in an authentic educational context. Collected data shows the adapted questionnaire measures intrinsic and extraneous cognitive load adequately. Further polishing is needed for the items addressing germane cognitive load.

I. INTRODUCTION

Cognitive Load Theory (CLT) has recently drawn attention in the educational research and instructional design literature. This framework provides a theoretical foundation for the allocation of working memory resources while engaging with problem solving and learning tasks [1]. Three different aspects of cognitive load are distinguished by the theoretical model: the intrinsic, extraneous, and germane cognitive loads. Intrinsic cognitive load (ICL) is regarded as the perceived difficulty level of a particular task. It is determined by the number of interactive elements of the task, the task type and the knowledge level of the learner. Extraneous cognitive load (ECL), on the other hand, is related to the imperfect nature of a task's representation, which embodies elements such as the design of the task, its wording, and its graphical or audio information, when available. The ECL is the part of the cognitive load which is irrelevant to the objective of the task. Therefore, ECL can be varied by modifying the design of the task without changing the topic of the task (unlike the ICL). Lastly, the Germane cognitive load (GCL) reflects the working memory resources devoted to task objectives. In many previous studies, the GCL has been found to be directly related to learning outcomes or problem-solving performance [2]. Recently, Sweller [3] suggested a new formulation to relate GCL to ICL and ECL. According to the newly proposed formulation, the GCL in learners emerges from the identification and processing of the portions of information available that are useful to a problem and that achieve learning. This process takes place after working memory resources are initially devoted to processing the representation of learning tasks associated with ECL to extract useful information, which in turn determines ICL. Therefore, if assuming learners with high motivation levels, GCL can be considered as a function of ICL and ECL.

Many studies use CLT as a lens to analyze the way students interact with assigned problems and learning interventions [4]. Practical instructional design principles have been generated to avoid hampering learning. With recent developments and implementation of CLT in education, the question of how to measure cognitive load has emerged [5, 6]. Self-reporting, physiological parameter measurements, and dual-task estimation are three major methods of cognitive load measurement.

Self-report scale in measuring cognitive load was first introduced by Paas [7]. The questionnaire usually contains items for learners to rate their perceived mental effort and difficulty level of the learning intervention. Unfortunately, the format and the content of a such questionnaire varies from study to study as there is no single accepted standard for these .

Physiological measures can also be used to indicate cognitive load and are overall a promising approach. Pupillary reactions [8], heart rate [9] and blood rate variability [10] have been adopted to measure the perceived workload, time pressure, and emotional strain, but the question on how to precisely relate the physiological parameters to cognitive load still needs to be answered.

Some studies [11, 12] have used a dual-task approach to estimate the cognitive load. Such approach runs a secondary task concurrently with the primary task and records reaction time and performance in both tasks. This approach can reveal cognitive load in real time, but it is challenging to differentiate ICL, ECL, and GCL from these measurements.

Earlier studies focused on assessing mental effort, pressure, or performance to reflect cognitive load as a unidimensional quantity. It is now known that it was difficult to distinguish between different types of cognitive load in these measurements, therefore limiting the applications of CLT in education. Furthermore, due to the difficulty of measuring ICL, ECL, and GCL separately, certain critical elements of the cognitive load architecture, such as the relationships between the different aspects of cognitive load, cannot be easily analyzed. This difficulty therefore hinders further refinements of the existing theoretical model.

To address this issue, Klepsch et al. [13] developed and validated a compact self-report questionnaire measuring ICL, ECL, and GCL in a differentiated way. Their goal was to "develop a reliable domain-unspecific questionnaire that could be validated and used in various learning situations" (page 5). They avoided using domain-related or learning objective-related terms in the statements of the items in their questionnaire with the purpose of improving compatibility with various learning contexts. Their work showed that a questionnaire with no more than ten 7-point Likert scale items could be a reliable tool in measuring ICL, ECL, and GCL separately without any pre-training required on the part of students. In their study, they designed 24 learning tasks with varying ICL, ECL and GCL. The authors chose to present the learning tasks and questionnaire online to 95 undergraduate students from a German university that were recruited to participate in their study. 65 out of the 95 participants completed both the learning task and the questionnaire. The questionnaire they designed showed good reliability: a high Cronbach's alpha value (>= .80) was reported for items addressing each of the aspects of cognitive load. However, it can be argued that their study has a few drawbacks. Firstly, the learning tasks they designed and assigned to participants were not part of participants' coursework, weakening their measurement on the GCL compared to the other two aspects of cognitive load due to a large variance on students' motivation. The reported medium-low GCL rating suggests participants were generally not very interested in the topics of their learning tasks. Secondly, it is evident that a computer-based, self-paced online learning environment is essentially different from an authentic learning environment. When a student is placed in front of a computer to participate in an interview, she must process in her mind all the information presented

on the computer screen, leaving her more susceptible to cognitive overload, particularly when redundant information is available. On the other hand, when the student is sitting in a classroom, she may take notes or discuss the activity with peers to avoid cognitive overload. The change in the learning environment causes behaviors and cognition in the classroom to be essentially different than when the student is in an interview situation. It is therefore interesting to adapt and validate cognitive load measuring methods to a real learning environment where students are engaged with their coursework. In the present study, we have used the questionnaire developed by Klepsch et al. to make students self-report ICL, ECL, and GCL imposed by the in-class group-work tutorials of a calculus-based introductory physics class. We hope our adapted questionnaire will help grant us insights on the cognitive load architecture of our tutorial activities, by extension also helping shed light on improved instructional design.

II. SIGNIFICANCE

There are two major goals for our present study. First, we want to validate the adaptation of a self-reported, differentiated cognitive load measure in an authentic classroom learning context. The do so, the reliability of the items designated to each cognitive load aspect will be assessed. Additionally, we would like to use the discriminated cognitive load measurements to reveal the cognitive load architecture. This study will not address the theoretical foundations of the cognitive load model. We will, however, focus on the relationships between ICL, ECL, and GCL as measured by the questionnaire while aiming to provide insight for further refinements to both the method of differentiated cognitive load measurements and the formulation of cognitive load.

III. METHODS AND MATERIALS

Participants were recruited from a calculus-based introductory physics course for life sciences students taught by one of the authors at a large public university in the Spring 2019 semester. 69 students were enrolled in the course. More than 90% of students attended classes regularly, according to the instructor's observation and the recording of in-class clicker questions. On some class days, students were asked to work in groups on assigned tutorials. These tutorials were designed by the course instructor based on the identified learning objectives of listed class topics. The structure of the tutorials provided students scaffolding support to help students practice problem-solving skills. The content and representation of these tutorials have been improved iteratively every semester according to student feedback and suggestions from other instructors. The course instructor was active during group discussions, pausing the activity when necessary to address prevailing questions. On class days with tutorials, these group activities typically took most of the class time. When time permitted, cognitive load

questionnaires were handed out to students for the rating of the group-based tutorial they had just finished. Students submitted tutorials and questionnaires at the end of the class period before they left. During the semester, four tutorials were administered accompanied by a questionnaire. They were, in chronological order, "Center of Mass", "Torque and Rotation", "Angular Momentum", and "Temperature and Heat". This study included the data collected on all four tutorials. Since the effect of the content on cognitive load is out of the scope of the present study, we refer to these four tutorials as tutorials 1-4 in the following discussion.

TABLE I. Items of the cognitive load questionnaire by Klepsch et al. [13]

Type of load	Item
ICL	For this task, many things needed to be kept in mind simultaneously.
ICL	This task was very complex.
GCL	I made an effort not only to understand several details, but to understand the overall context.
GCL	My point while dealing with the task was to understand everything correct.
GCL	The learning task consisted of elements supporting my comprehension of the task.
ECL	During this task, it was exhausting to find the important information.
ECL	The design of this task was very inconvenient for learning.
ECL	During this task, it was difficult to recognize and link the crucial information

The items we used in our study to assess cognitive load were from "the second version of the naïve rating question" (page 10) designed by Klepsch et al. [13]. The items are listed in Table I. The ICL and ECL portions of the questionnaire were reported to have good reliability, while the three items addressing GCL showed relatively low reliability with a Cronbach's α-value of 0.67. It is worth noting that item #5 on the questionnaire, an item designed to assess GCL was found to be much weaker than the other two GCL items. As a result, the authors of the original study removed this item from their data analysis, leading to a much higher α-value. In our study, we decided to include all eight items in the questionnaire assigned to students for two reasons: first, all our tutorials used in class are designed to be problem-solving worksheets for the purpose of enhancing students' conceptual understanding and problem-solving skill, which is inherently aligned with the GCL item Klepsch et al. suggested be removed from the list; second, we wanted to use our study to verify their findings on the reliability of the GCL items.

IV. RESULTS AND DISCUSSION

A. The reliability of the questionnaire

The reliability of the questionnaire was first analyzed separately for each tutorial and then aggregated to give an overall view. For each aspect of cognitive load, we calculated Cronbach's α-value for student ratings on the items. To obtain Cronbach's α-value for the overall reliability (including the data collected on all four tutorials), we adopted the formulas presented by Rodriguez and Madeda [14] to calculate a weighted mean of α-values for each aspect of cognitive load, where the weight is a function of the number of participants and the number of the items. Moreover, the Spearman-Brown Prophecy formula was used to adjust the aggregated α-value of ICL to account for the fewer number of items since it is a well-known fact that less items will yield a lower Cronbach's α-value.

TABLE II. Cronbach's α-values of the items for ICL, ECL, and GCL on the four tutorials

	Tutorial 1	Tutorial 2	Tutorial 3	Tutorial 4
ICL	0.702	0.487	0.731	0.809
ECL	0.806 (0.813 excluding item#6)	0.753	0.834	0.876
GCL	0.703 (0.715 excluding item#3)	0.609 (0.688 excluding item#5)	0.798 (0.827 excluding item#5)	0.815

We found that the items for ECL had a good reliability with aggregated $\alpha = 0.826$. This is consistent with the results reported by Klepsch et al. [13]. Changing the measuring context from a clinical interview setup to a real classroom environment did not affect the reliability of the items influencing ECL. The items for ICL also showed acceptable reliability with adjusted aggregated $\alpha = 0.790$. It is worth noting that the questionnaire only contained two items for ICL, which might be the reason of the alpha value's fluctuation. However, the part of the questionnaire testing GCL showed a much lower reliability with aggregated $\alpha = 0.752$. The data collected on two out of the four tutorials suggested deleting the fifth item would increase the reliability [see Table II]. This finding matches the weak item reported by the study of Klepsch et al. We therefore excluded the fifth item from our Cronbach's alpha calculation and obtained a raw aggregated α-value of 0.747 and an adjusted aggregated α-value of 0.816 (including the Spearman-Brown adjustment). We observed that including the fifth item could not improve the quality of the questionnaire, so we only included the third and the fourth items in the following discussion on GCL rating.

As a next step, we averaged students' ratings on the items designated for each aspect of cognitive load (ICL, ECL, and GCL) to be the measured in the group-based tutorials. Our goal was to determine whether student ICL, ECL, and GCL ratings on each of the four tutorials were different. ICL ratings on the four tutorials were observed to be significantly different ($F(3, 232) = 27.997$, $p < .001$). The ECL ratings on the four tutorials were also significantly different ($F(3, 232) = 15.621$, $p < .001$), while in contrast the GCL ratings on the four tutorials were not significantly different ($F(3, 232) = 2.498$, $p = .060$). Since the present study only included four tutorials and there were no specific reasons for choosing these four tutorials, we are currently not well-equipped to discuss potential effects this particular choice of problems and content topics had on the three aspects of cognitive load. Nevertheless, the ICL and ECL ratings found in our present study suggest students can distinguish between the inherent difficulty and the representation of the tutorials and report the differences on the ICL and ECL items as expected. Table III includes the p values of all *post hoc* pairwise comparisons with Bonferroni corrections. The ICL rating showed that students felt Tutorials 1 (4.54 ± 0.14) and 4 (4.77 ± 0.17) were on the same difficulty level and were significantly different from Tutorials 2 (6.08 ± 0.09) and 3 (5.53 ± 0.14). The ECL rating also showed that students felt the wording and the design of Tutorials 1 (3.20 ± 0.17) and 4 (2.96 ± 0.17) were easier to interpret than Tutorials 2 (4.32 ± 0.15) and 3 (3.93 ± 0.16). On the other hand, students rated the GCL items on all tutorials at about the same high level (i.e., Tutorial 1: 5.47 ± 0.12, Tutorial 2: 5.62 ± 0.11, Tutorial 3: 5.77 ± 0.13, Tutorial 4: 5.92 ± 0.13). Comparing with the GCL rating reported by Klepsch et al. [13], the much higher GCL ratings in our study suggested that students in our study had high motivation to learn from the in-class tutorials.

TABLE III p-values of pairwise comparisons with Bonferroni corrections on ICL, ECL, and GCL between tutorials

	Tutorial 1	Tutorial 2	Tutorial 3	Tutorial 4
ICL				
Tutorial 1	N/A	< .001	< .001	1.000
Tutorial 2	< .001	N/A	0.019	< .001
Tutorial 3	< .001	0.019	N/A	0.001
Tutorial 4	1.000	< .001	0.001	N/A
ECL				
Tutorial 1	N/A	< .001	0.007	1.000
Tutorial 2	< .001	N/A	0.474	< .001
Tutorial 3	0.007	0.474	N/A	< .001
Tutorial 4	1.000	< .001	< .001	N/A
GCL				
Tutorial 1	N/A	1.000	0.441	0.061
Tutorial 2	1.000	N/A	1.000	0.524
Tutorial 3	0.441	1.000	N/A	1.000
Tutorial 4	0.061	0.524	1.000	N/A

B. The relationships between the measured ICL, ECL, and GCL

With the self-reported ICL, ECL, and GCL ratings on the questionnaire, the relationships between these three aspects of cognitive load can be probed through the calculation of the correlations between the measured ICL, ECL, and GCL. The measured values for the ICL and ECL showed a statistically significant, positive correlation ($r = .586$, $n = 236$, $p < .001$), which was unexpected since the separation between ICL and ECL is theoretically dichotomous according to CLT: ICL is defined to be related to the inherent difficulty of the task while ECL is evoked by the part of the task that does not relate to the objectives of the task (i.e., the representation, irrelevant content, etc.). Such a partition is a conceptual necessity but also critical for practical applications. Many influential authors have suggested [2, 15] that in theory, ICL and ECL are additive and that their sum reflects the total cognitive load imposed by the task. Researchers have therefore generated instructional design principles to reduce ICL without changing the learning objectives based on their experimental results, with those design principles showing positive effects on increasing student learning. However, the transition of the experimental context from clinical interview rooms to real classrooms is not trivial. It is not ethical, for example, to assign a learning task to a group of students while knowing its imperfect design may impose higher ICL and therefore hinder learning in classroom. On the other hand, this is not as important an issue in the design of a controlled experiment. A feasible method therefore needs to be developed to allow the comparison between task designs across different task topics. We believe this exacerbates the limitation of only using one self-reported questionnaire to distinguish between the three aspects of cognitive load.
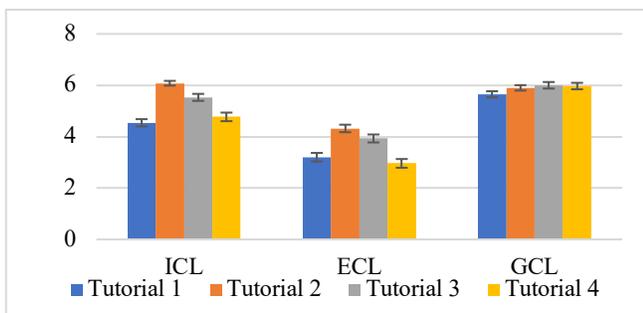

FIG 1. The ICL, ECL, and GCL ratings for the four tutorials. Error bars depict the standard error.

Our study has shown students were able to rate the ICL and ECL items differently [see Figure 1], which suggests students understood the ICL and the ECL imposed by the tutorials emerged from different aspects of the activities completed. The strong positive correlation between the ICL and ECL ratings suggested the inherent difficulty of the tutorials may have affected students' perception on tutorial

designs or vice-versa. To students, the objective of working on an in-class tutorial is to understand the activity and then complete it. As long as tutorials were generally well designed, it was unrealistic to expect students to be able to self-report the feeling of "the problem is hard" on two separate layers: the difficulty due to the content and the difficulty due to the activity's imperfect representation. The ratings they gave to the ICL items and ECL items were therefore all associated with their perceived difficulty of the task.

A statistically significant, negative correlation between the measured ECL and GCL ($r = -.186$, $n = 236$, $p = .004$), was also found, suggesting a task imposing higher ECL may have resulted in a lower GCL. This is consistent with CLT. With limited working memory resources, more processing capacity devoted to the part of the task irrelevant to the objectives will result in less processing capacity devoted to completing the task.

Lastly, there was no statistically significant correlation between the measured ICL and GCL ($r = .004$, $n = 236$, $p = .956$). There was no relationship between the difficulty of the tutorials and students' motivation and effort devoted to them. Since CLT states cognitive overload hinders student learning, tasks with extremely high ICL and ECL will typically result in low GCL. This result shows both the content and the representation of the tutorials administered were effective in avoiding cognitive overload in students.

V. CONCLUSIONS AND FURTHER IMPLICATIONS

In the present study, we have verified the reliability of the items designated to measuring ICL, ECL, and GCL separately. The questionnaire designed by Klepsch et al. was found to provide reliable measurements on students' perception of the difficulty level of the task, its design, and their overall devoted effort. We also found a strong relationship between the two theoretically independent cognitive load components: ICL and ECL. We felt this was due to the nature of the self-report measuring method. As a future plan for studying such a relationship, we recommend adding a secondary measuring method in parallel to the self-reporting questionnaire to measure the imposed ICL and ECL using another perspective.

We have also used the measurements on ECL and GCL to reveal the negative impact of ECL on learning. This finding was consistent with CLT and demonstrated the value of measuring cognitive load in a differentiated way.

As an educational implication of measuring cognitive load, our study has shown that it is feasible to adopt such a measurement regularly in the classroom for assessing the imposed cognitive load of learning interventions. This tool can also indicate to instructors whether the interventions might be too challenging or too trivial for students and whether the implemented interventions are actually hindering student learning through cognitive overload.

[1] J. Sweller, *Cognitive science*, **12**, 2 (1988).
[2] F. G. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven, *Educational psychologist*, **38**, 1 (2003).
[3] J. Sweller, *Educational psychology review*, **22***, 2 (2010).
[4] R. E. Mayer, *The Cambridge handbook of multimedia learning* (Cambridge university press, 2005).
[5] F. G. Paas, A. Renkl, and J. Sweller, *Educational psychologist*, **38**, 1 (2003).
[6] T. De Jong, *Instructional science*, **38**, 2 (2010).
[7] F. G. Paas, *Journal of Educational Psychology*, **84**, 429 (1992).
[8] T. Van Gog, L. Kester, F. Nievelstein, B. Giesbers, and F. G. Paas, *Computers in Human Behavior*, **25**, 325 (2009).
[9] P. Nickel and F. Nachreiner, Human Factors, **45(57)**, 5 (2003).
[10] T. K. Fredericks, S. D. Choi, J. Hart, S. E. Butt, and A. Mital, *International Journal of Industrial Ergonomics*, **35**, 1097 (2005).
[11] R. Brünken, J. L. Plass, and D. Leutner, *Educational Psychologist*, **38**, 53 (2003).
[12] K. E. DeLeeuw and R. E. Mayer, *Journal of Educational Psychology*, **100**, 223 (2008).
[13] M. Klepsch, F. Schmitz, and T. Seufert, *Frontiers in psychology*, **8**, 1997 (2017).
[14] M. C. Rodriguez and Y. Maeda, *Psychological methods*, **11**,3 (2006).
[15] J. Sweller, J. J. Van Merrienboer, and F. G. Paas, *Educational psychology review*, **10**,3 (1998).