Developing a Methodological Approach to Measure Cognitive Load during Complex Problem Solving.

Clodagh Reid, Conor Keighrey, Rónán Dunbar, Niall Murray, Jeffrey Buckley.





Clodagh Reid
Athlone Institute of Technology

Ireland
c.reid@research.ait.ie

Conor Keighrey
Athlone Institute of Technology

Ireland
c.keighrey@research.ait.ie

Rónán Dunbar
Athlone Institute of Technology

Ireland
rdunbar@ait.ie

Niall Murray
Athlone Institute of Technology

Ireland
nmurray@research.ait.ie

Jeffrey Buckley
Athlone Institute of Technology, KTH Royal Institute of Technology
Ireland, Sweden
jbuckley@ait.ie

ABSTRACT

Problem solving is an important element of engineering and technology disciplines and spatial ability contributes to learners' success in problem solving in these areas (Wai, et al., 2009). As excessive cognitive load can impede an individual's capacity to process information (Kirschner, Paas & Kirschner, 2009) it is posited that higher levels of spatial ability may reduce the cognitive load experienced when problem solving and thus support increased learner performance and capacity to learn from problem solving episodes. Based on this hypothesis, there is a need to establish appropriate methods to measure cognitive load in educational contexts. Using Cognitive Load Theory (Sweller, 1988) as a theoretical framework, this paper presents a pilot study of a methodological approach to measure cognitive load experienced in real-time during complex problem solving activities through the use of physiological sensors. Postgraduate students (n=26) were administered the Tower of Hanoi, a complex problem solving task (Eielts et al, 2018). While completing the task, physiological sensors were worn by participants on their non-dominant hand capturing details of electrodermal activity, which is an indicator of cognitive load experienced in real time (Setz, et al., 2010). Subjective data on the levels of cognitive load experienced was also captured following the task, where participants completed a 9-point Likert-type item. The analysis of the data for this study illustrated that through the use of a physiological sensor and application of novel time monitoring software, the electrodermal activity of an individual can provide an insight into their experience whilst problem solving. This approach may present a valid way to capture cognitive data of students throughout authentic problem solving scenarios which would support the determination of variables underpinning success.

Keywords: Cognitive load, Spatial ability, Problem solving, Physiological sensors.

Introduction

Cognitive load theory describes how the mental effort experienced by an individual when processing information can impact their capacity to process that information effectively (Kirschner, et al., 2018; Sweller, 2011, 2010). Through impacting processing capacity, cognitive load can affect the individual's performance on a task (Kirschner, et al, 2018; Paas, et al., 2004; Van Gog, et al., 2010) thus impeding their learning. This is of particular importance to teaching and learning as educators strive to support learners in understanding new concepts. Within engineering and technology education, learners experience a variety of problems and problem types to aid in acquiring and contextualising technical and transversal competencies (Buckley, et al., 2018; Kirschner, et al., 2009; Schoenfeld, 1983). The problems learners experience can vary in how well or ill-defined they are, and also in terms of being open or closed-ended. Therefore, these problems require learners to adopt a variety of approaches to solve them (Reid, et al., 2018).

Throughout problem solving, cognitive load may be experienced for a number of reasons, such as a lack of knowledge or the way through which the problem is communicated (Kirschner, et al., 2018; Paas, et al., 2004, 2003). The effects of cognitive load on problem solving in education have been explored on both an individual and collaborative level (Kirschner, et al., 2018; Kirschner, et al., 2009; Paas, et al., 2004) as problem solving can take place in either format. Subjective measures of cognitive load such as the mental effort Likert-type scale developed by Pass (1992) have become common measures of cognitive load in studies as they are unobtrusive and offer valid and reliable indications of overall cognitive load experienced during a task (Sweller, 2011). Following technological advancements, objective measures such as electroencephalography (EEG), pupilometry, eye-tracking and heart-rate (HR) measurement have been explored for their capacity to measure the cognitive functions of an individual during a task (Antonenko, 2010; Palinko, 2010; Sweller, 2011). This paper presents a methodological approach to measure cognitive load throughout complex problem solving in an unobtrusive manner through the use of a medical grade physiological sensor, the Empatica E4 sensor. Specific emphasis was placed on determining the capacity of the E4 sensor to capture precise data in relation to the cognitive load experienced by individuals in real time. As this study is based in technology and engineering education, the nature of problem solving, spatial ability, and cognitive load, and the interaction of these three variables are critical factors.

Problem solving

Tasks or exercises become problems when an individual does not know how to go about solving them (Buckley, et al., 2018) and therefore problems occur in a variety of ways in engineering and technology education. In solving problems, learners can use various methods and approaches to develop appropriate solutions (Jonassen, et al., 2009; Reid, et al., 2018) and due to the prevailing treatment of knowledge in technology and engineering education, they may utilise heuristics to reason about the problem (Buckley, et al., 2018). While working to understand the information presented in a problem and to identify an appropriate means to solve it, learners may experience substantial cognitive load due to their limited processing capacity (Marois, & Ivanoff, 2005). This is a significant consideration for educators and highlights the importance of supporting the learner with appropriate instruction throughout the problem solving process to limit the effects of cognitive load (Paas, et al., 2004, 2003). From a foundational research perspective, there is a need to consider the nature of problems that occur in context, but to work with problems that are a valid indicator of an important construct. Therefore to study problem solving in context with a novel instrument, it can be necessary to use abstract problems first prior to working with contextually validated authentic problems.

Spatial Ability

Spatial ability has been attributed to success in STEM disciplines (Uttal & Cohen, 2012) and is noted for its impact on student retention within these subject areas (Sorby, 2001). It has also been demonstrated to correlate with performance in technology and engineering education specifically (Buckley, et al., 2019; Lin, 2016; Sorby, et al., 2013). Based on these investigations and the role of spatial ability in problem solving in engineering and technology, the investigators postulated that spatial ability may influence the cognitive load experienced by students when problem solving throughout engineering and technology education. In considering the circumvention-of-limits hypothesis (Hambrick, et al., 2012) which describes the interaction between domain knowledge and generic cognitive abilities of disciplinary importance in terms of problem solving performance, there is a clear need to establish levels of spatial ability in participants in problem solving research in the context of technology and engineering education.

Cognitive load

As discussed, cognitive load theory proposes that the load experienced during information processing can impact the successful processing of information and thus influence the performance of the individual on the task (Kirschner, et al., 2018; Sweller, 1988). Cognitive load can be further classified in terms of extraneous, intrinsic and germane cognitive load (Paas, et al., 2003; Sweller, 2010).

Germane cognitive load is outlined as the resources of working memory which are used to deal with intrinsic cognitive load (Kirschner, et al., 2018). Intrinsic cognitive load relates to the essential complexity of the information which must be processed (Sweller, 2010). Cognitive load that is unnecessary and interferes with the processing of information is called extraneous cognitive load (Paas, et al., 2003). Extraneous cognitive load can be caused by the way the information is presented to an individual (Kirschner, et al., 2018; Paas, et al., 2003; Sweller, 2011). In considering a methodological approach to measure cognitive load, there is a need to consider if there is a specific type of cognitive load which is the target of measurement, or if the agenda is to elicit the degree of experienced cognitive load as a whole.

Methodology Development

Measuring Cognitive Load

Subjective Measures

There are various measures which can be used to investigate the cognitive load experienced by individuals when carrying out a task, such as the NASA-TLX and a number of Likert scale approaches (Leppink & Van Merriënboer, 2015; Van Merriënboer & Sweller, 2005; Plas & Kalyunga, 2018). The NASA-TLX is outlined as a direct subjective measure of cognitive load (Plas & Kalyunga, 2018), making it a potentially suitable mechanism to measure self-reported cognitive load of participants in the study. Mental effort scales and task difficulty scales, also included in the NASA-TLX, are outlined throughout cognitive load literature as measures of overall cognitive load (Leppink & van Merriënboer, 2015; Plas & Kalyunga, 2018), which is the focus of this paper.

Paas (1992) developed and validated a 9-point Likert-type item to evaluate the overall cognitive load experienced by an individual (Leppink & van Merriënboer, 2015; van Merriënboer & Sweller, 2005). Additionally, recent investigations have explored the capacity of similar Likert-type items to accurately measure extraneous, intrinsic and germane cognitive load (Leppink, et al, 2014, 2013). In developing the 9-point Likert-type items for this study the investigators sought to measure overall cognitive load, difficulty experienced with the task, concentration and stress to determine factors that may influence participant performance. Stress can also be measured through objective measures of cognitive load.

Objective Measures

Various sensors and software have been used to investigate the cognitive load experienced by individuals throughout activities. These include pupilometry, eye-tracking, electroencephalography (EEG), heart rate (HR) measurements (Sweller, 2011) and electrodermal activity (EDA) measurements (Keighrey, et al., 2017; Setz, et al., 2010). Recent works have demonstrated that measures of EDA can be used to monitor cognitive load experienced by individuals during arithmetic and reading tasks (Nourbakhsh, et al., 2012) and of individuals using a driving simulator (Son & Park, 2011). EDA has also been evaluated for the capacity to discriminate between stress and cognitive load experienced with an activity (Setz, et al., 2010).

The use of physiological sensors can often impact natural interaction with a problem solving task. To reduce such influence, a non-intrusive device was selected. More specifically an Empatica E4 wristband. The Empatica is a medical grade device which captures physiological measures of EDA, HR, blood volume pulse (BVP), skin temperature, and movement data (accelerometer). In measuring physiological responses, it is pivotal that baseline measurements are obtained so that they can be compared to the readings when the individual is exposed to the stimulus (Setz, et al., 2010). Typically, baseline measurements range between 2 and 20 minutes (Alvarsson, et al., 2010; McDuff, et al., 2014; Setz, et al., 2010). However, recent works have highlighted 5 minutes as an appropriate duration as it offers sufficient time for the individual to relax following their arrival at the session and for baseline measures to be gathered (Alvarsson, et al., 2010).

Problem Solving Task

The investigators sought to examine the cognitive load experienced by individuals during a complex problem solving activity. As this was a pilot study and participants would come from various backgrounds with different technical knowledge bases, the investigators sought a general problem with no/minimal discipline-specific knowledge required to solve it.

Having explored various styles of problems the Tower of Hanoi (TOH) was identified as the problem solving task that would be used for this study as it is representative of a complex problem solving task (Eielts et al, 2018). The difficulty of the task can also be increased through the addition of discs which would allow the investigators to monitor the cognitive load experienced between an easier and more difficult problem solving activity.

The TOH can be solved through an optimal number of moves (Eielts, et al., 2018; Stewart & Eliasmith, 2011) making it necessary for the number of moves made by participants to be monitored throughout the task. In order to monitor the moves made by participants a software programme was developed whereby the participants' engagement with the task could be tracked. A start point for the activity was incorporated into the software so that the time could be observed between participants being told to begin the task and the first move being made. Each move made by participants could be inputted into the software. Through incorporating this feature the time taken between moves could be time-mapped onto the sensor data to monitor the physiological responses of participants between moves. This would allow for a robust analysis of the data following the completion of the study.

Methodology Implementation

Participants

Postgraduate students (n=26) were invited to take part in the study. The participants consisted of 13 males and 13 females with ages ranging between 21 and 48. Participation in the study was voluntary and individuals were entered into a draw for a €20 voucher as an incentive for participation. For the purposes of this paper, only 2 of the participants are discussed. The selection criteria are discussed below.

Spatial Ability Testing

The Purdue Spatial Visualisation Test: Rotations (PSVT:R) (Bodner & Guay, 1997), Surface Development test (SDT) and Paper Folding test (PFT) (Ekstrom, et al., 1976) were used to measures participant spatial ability. A composite z-score of each of the spatial tests were derived for each of the participants. Z-scores were used as the purpose of this approach is to examine participants with varying levels of spatial ability within the group, rather than to compare the results to international datasets. Participants' scores on each of the three spatial tasks were converted to standardised z-scores relative to the entire cohort and then averaged to determine a single z-score per participant.

Sensor Setup
As per the device recommendations, participants were fitted with the Empatica E4 on the non-dominant. The sensors for EDA were positioned in-line with the joint of the second and third finger at the wrist (Figure 1) to allow for an accurate and precise reading of the physiological response. Participants were instructed to relax for a period of 5 minutes to obtain baseline measurements of EDA.

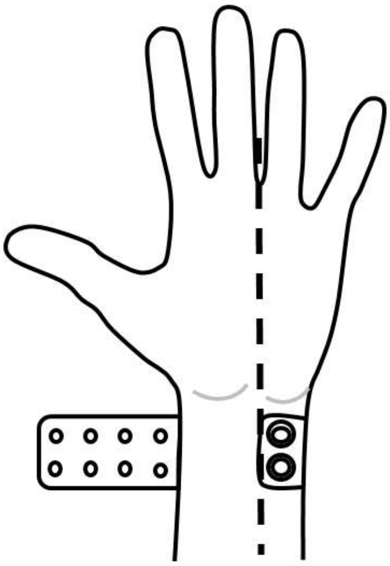

Figure 21. Empatica E4 sensor set-up

Problem Introduction
Following the baseline period of 5 minutes, participants were presented with the 3-disc TOH task. They were asked to indicate whether they had seen the problem before and if they subsequently knew how to solve it. These indications did not act as an element of selection criteria. Once this was complete, the instructions for the task were explained to participants as follows:

'The goal of the task is to get the arrangement of 3 discs on the left-most peg to the right-most peg. The discs must be arranged in the same order i.e. largest on the bottom to smallest on the top. There are two conditions:

1. *Only one disc can be moved at a time from one peg to another*
2. *A larger disc cannot be placed on a peg that already contains a smaller disc'*

Participants were provided with the opportunity to ask questions to ensure that they understood the task. They were then instructed to begin the problem.

Problem Solving Task
Throughout the problem solving task participants were recorded. A video camera was setup and focused specifically on participant hand gestures. This ensured a level of anonymity was achieved, in addition to providing opportunity to capture the interaction with the activity.

In parallel to this, a novel software was developed to objectively capture user interaction. The software solution automated the capture of task completion (time), interaction time, and number of moves taken to complete the task. The objective of this was to create an accurate measure of user performance.

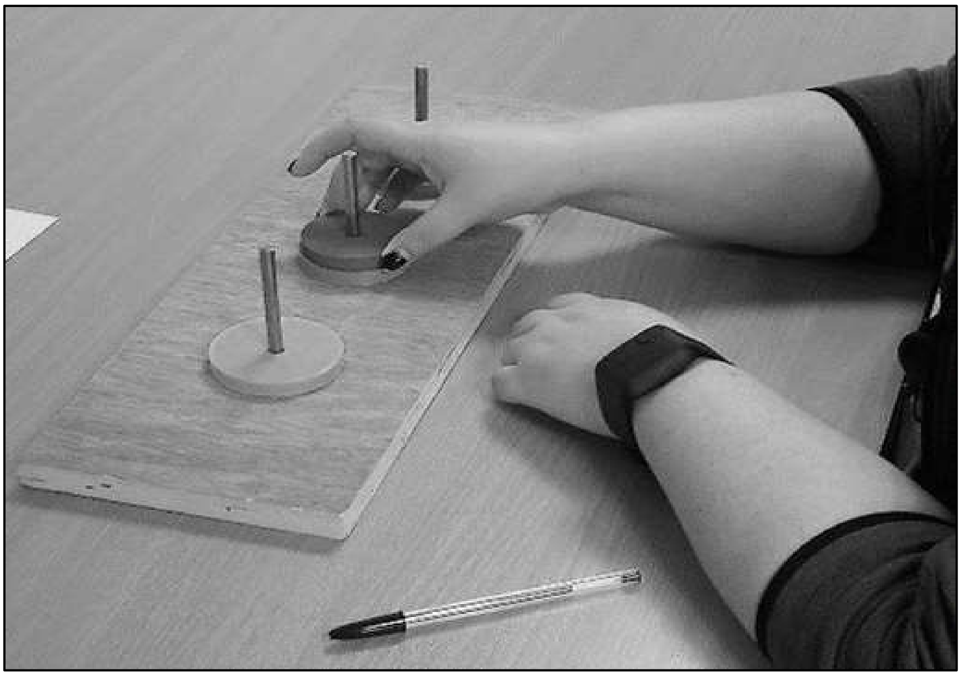

Figure 22. Task setup.

Each move was marked by the investigator electronically with the final move concluding the time tracking on this session. A move was counted when the disc was placed on a peg and released by the participant.

Having completed the task participants were asked to indicate on a 9-point Likert-type item the amount of mental effort, difficulty, stress and concentration they experienced throughout the task. When the item was completed, participants were presented with the second problem, the 4-disc TOH. The same process was repeated for the 4-disc problem.

Results

The data presented in Table 1 represents the mean (M) and standard deviation (SD) of self-reported mental effort, difficulty, stress and concentration experienced by all participants (n = 26) with the 3- and 4-disc TOH tasks. The table also includes the means and standard deviations of the number of moves made by participants during the 3- and 4-disc TOH.

Table 1. Mean and standard deviation of participant responses and performance.

	3 Disc Problem		4 Disc Problem	
	M	SD	M	SD
Mental Effort	3.46	1.68	5.27	1.61
Difficulty	2.96	1.37	5.08	1.55
Stress	2.19	1.33	3.69	1.83
Concentration	4.81	1.67	5.73	1.56
Moves	8.12	2.29	28.12	12.09

Participants were categorised in terms of spatial scores. For each participant a spatial z-score was determined based on their percentage score on the PSVT:R, SDT and PFT. Participant 16 had the highest overall spatial score (z = 1.33), while participant 20 had the lowest overall spatial score (z = -1.89) of the 26 participants. The remaining data documented will focus on participant 16 and 20. Table 2 presents the z-scores of these two participants on the self-reported 9-point Likert-type item. The z-scores outline where each of these participants ratings sit within the overall cohort with respect to the mean and standard deviation.

Table 2. Self-report data

Participant	Mental Effort	Difficulty	Stress	Concentration	No. of Moves
			3 Disc Tower of Hanoi		
16	0.92	0.03	-0.90	0.11	-0.05
20	-0.87	-0.70	-0.14	-0.48	-0.49
			4 Disc Tower of Hanoi		
16	-0.17	-0.05	-1.47	0.17	-0.75
20	0.45	0.60	0.17	0.17	1.56

The EDA data of participants was evaluated relative to their problem solving performance. In order to identify EDA events of importance within the data of each participant, a time threshold for moves on the 3- and 4-disc TOH was determined. Any time that was below or equal to this threshold was considered as a normal time spent on a move and anything above this threshold was considered as a large time spent on a move. Thresholds were determined for each participant by calculating the mean time spent on a move, standard deviation of time for a move and adding the standard deviation to the mean. Table 3 presents the mean, standard deviation and threshold of time for participants 16 and 20.

Table 3. Time threshold for participant moves

Participant	Move Time (M)	Move Time (SD)	Move Time Threshold
		3 Disc Tower of Hanoi	
16	8.205	15.750	23.955
20	3.507	1.327	4.834
		4 Disc Tower of Hanoi	
16	4.735	4.120	8.855
20	3.267	1.838	5.104

The EDA data collected during the 3- and 4-disc TOH for participants 16 and 20 is presented in Figures 3, 4, 5 and 6. The values displayed for EDA are z-scores.

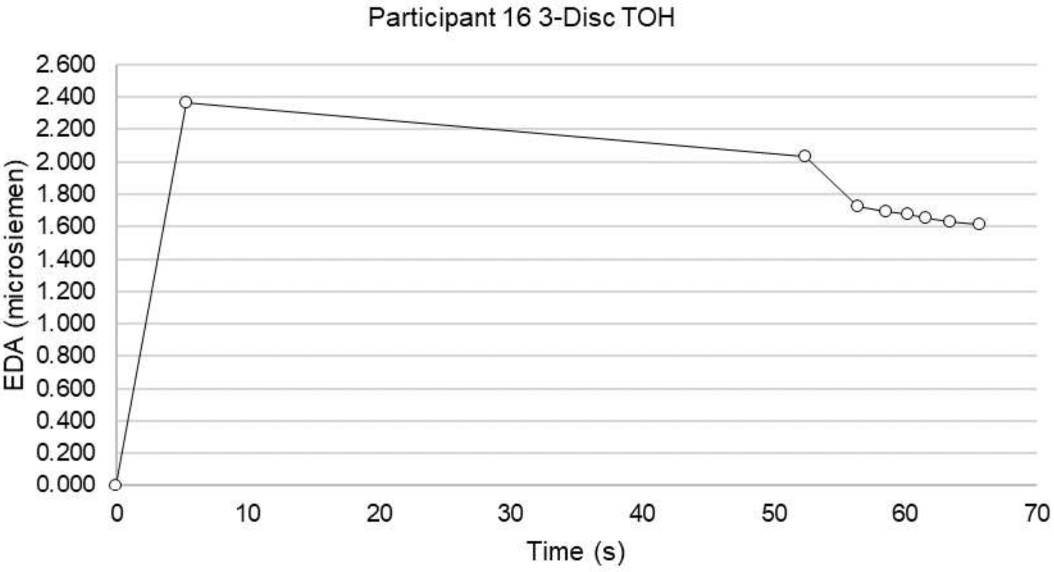

Figure 23. Participant 16 EDA Data during 3-disc TOH

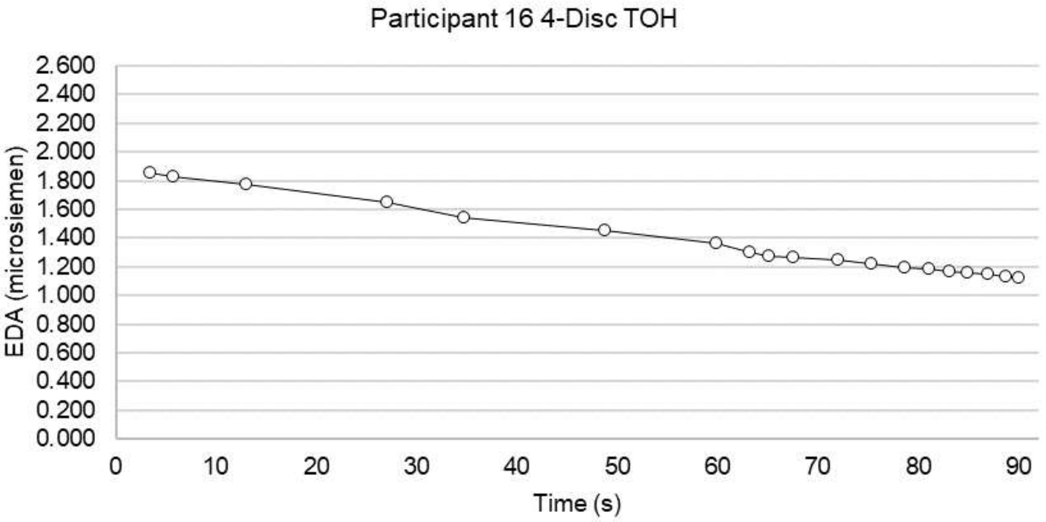

Figure 24. Participant 16 EDA during 4-disc TOH

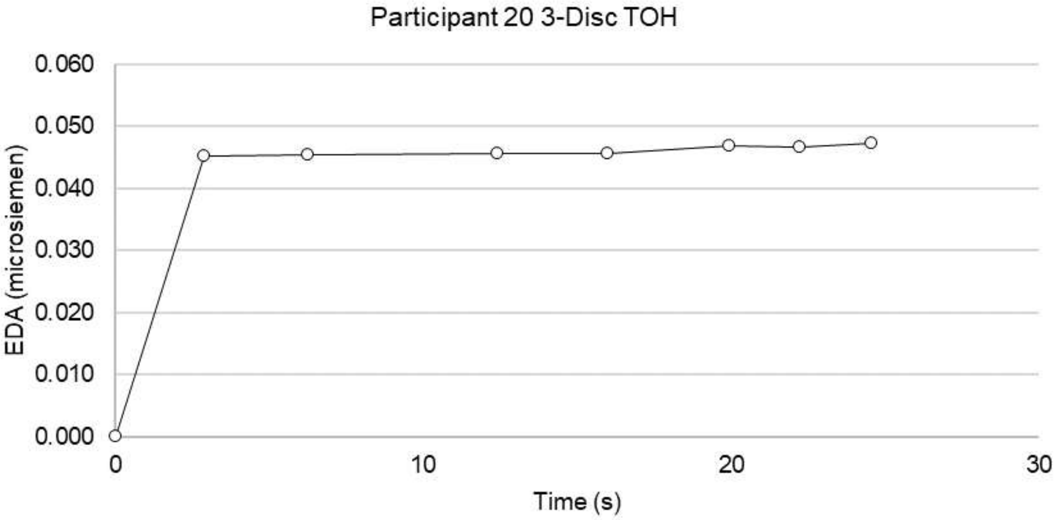

Figure 25. Participant 20 EDA during 3-disc TOH

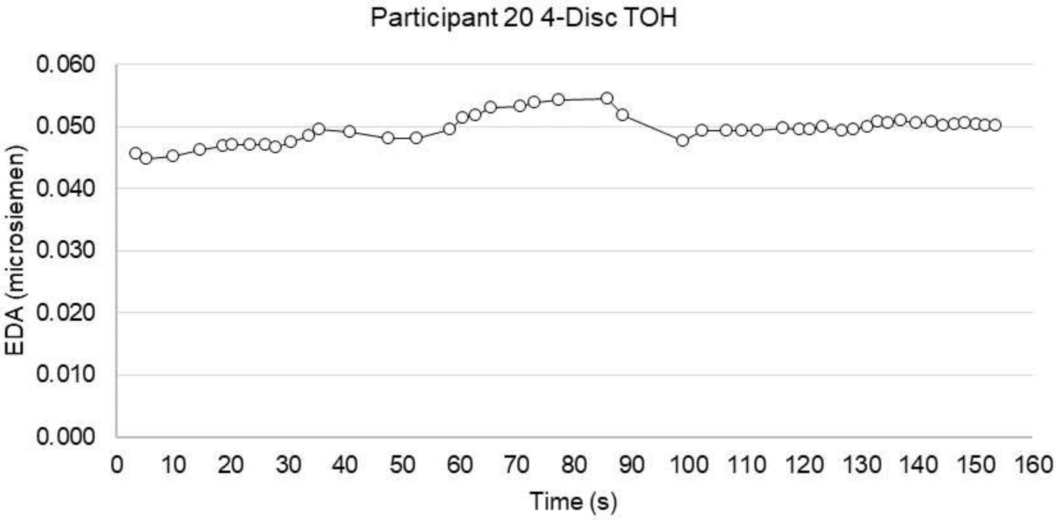

Figure 6. Participant 20 EDA during 4-disc TOH

Discussion

The purpose of this investigation was to develop a methodological approach to objective measure the cognitive load experienced by individuals throughout a problem solving task. The Empatica E4 physiological sensor was used to measure the EDA of participants during the task as increased EDA can be indicative of increased mental effort (Setz, et al., 2010). Overall percentage spatial scores of the participants were used to organise the data. Participant 16 and 20 were identified to have the highest and lowest spatial score respectively of the 26 participants. The EDA data for these participants was then analysed.

Through evaluating the EDA data of participants it was critical that a baseline measurement was obtained so that the EDA of the individual in a relaxed state could be compared to the data collected during the problem solving episode. The first point on the graphs in figure 3 and 5 represents the baseline EDA of participants 16 and 20. The second point on these graphs represents the time taken by participants between being instructed to begin the task and making their first move.

Through capturing the time taken between moves using a software programme, investigators were able to map these times to the EDA data. The time thresholds, presented in Table 3, were then used to determine points of interest within the EDA data where there was a large delay between moves relative for each person. With participant 16, there was an initial increase in EDA with the 3-disc TOH. Following a large delay, EDA continued to decrease throughout both problem solving tasks. Participant 20 did not have much variation in EDA with the first problem, however, there was an increase and variation during the second problem.

Investigating the data in this manner provides an insight into the experience of a student during problem solving. This approach also affords the capacity to isolate a moment within the problem to examine the reason it may have had an effect on the individual. The use of a physiological sensor to monitor EDA during problem solving may inform educational practice by supporting educators in examining pedagogical strategies relative to individual learners' cognitive abilities, exploring the effects of group dynamics in design activities, and identifying potentially unknown sources of cognitive load and stress in students and teachers. The future of this work will focus on further analysis of the data and monitoring the cognitive load experienced by students throughout authentic engineering and technology problems.

References

Antonenko, P., Paas, F., Grabner, R., & Van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*. 22(4), 425-438.

Bodner, G. M., & Guay, R. B. (1997). The purdue visualization of rotation test. *The Chemical Educator*. 2(4).

Buckley, J., Seery, N., & Canty, D. (2018). Investigating the use of spatial reasoning strategies in geometric problem solving. *International Journal of Technology and Design Education*. 341-362.

Eielts, C., Pouw, W., Ouwehand, K., Van Gog, T., Zwaan, R. A., & Paas, F. (2018). Co-thought gesturing supports more complex problem solving in subjects with lower visual working-memory capacity. *Psychological research*. 1-12.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Hambrick, D.Z., Libarkin, J.C., Petcovic, H.L., Baker, K.M., Elkins, J., Callahan, C.N., Turner, S.P., Rench, T.A. & LaDue, N.D. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology: General*. 141(3). 397-403.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*. 50(9), 904-908.

Keighrey, C., Flynn, R., Murray, S., & Murray, N. (2017). A QoE evaluation of immersive augmented and virtual reality speech & language assessment applications. In 2017

Ninth International Conference on Quality of Multimedia Experience (QoMEX). IEEE. 1-6.

Kirschner, F., Paas, F. & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational psychology review*. 21, 31-42.

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano, J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning.* 13(2), 213-233.

Leppink, J., & van Merriënboer, J. J. (2015). The beast of aggregating cognitive load measures in technology-based learning. *Educational Technology & Society*. 18(4), 230-245.

Lin, H. (2016). Influence of design training and spatial solution strategies on spatial ability performance. *International Journal of Technology and Design Education*. 26(1), 123-131.

Marois, R. & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*. 9 (6), 296-305.

Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In Proceedings of the *24th Australian Computer-Human Interaction Conference* (420-423). ACM.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist*. 38(1), 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional science*. 32(1), 1-8.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications*. 141-144.

Reid, C., Dunbar, R., & Buckley, J. (2018). A preliminary model of problem categorisation to explore the cognitive abilities required for problem solving in engineering education. In N. Seery, J. Buckley, D. Canty, & J. Phelan (Eds.), *PATT2018: Research and Practice in Technology Education: Perspectives on Human Capacity and Development*. Westmeath, Ireland: Athlone Institute of Technology.

Schoenfeld, A. (2009). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. *Colección Digital Eudoxus*. (7).

Schoenfeld, A. (1983). The wild, wild, wild, wild, wild world of problem solving (A review of sorts). *For the Learning of Mathematics*, 3(3), 40–47.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine*. 14(2), 410-417.

Son, J., & Park, M. (2011). Estimating cognitive load complexity using performance and physiological data in a driving simulator. In *Adjunct Proceedings of the Automotive User Interfaces and Interactive Vehicular Applications Conference*.

Sorby, S. A. (2001). Improving the Spatial Skills of Engineering Students: Impact on Graphics Performance and Retention. *Engineering Design Graphics Journal*. 65(3), 31-36.

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*. 26, 20-29.

Stewart, T., & Eliasmith, C. (2011). Neural cognitive modelling: A biologically constrained spiking neuron model of the Tower of Hanoi task. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. 33(33).

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*. 12(2), 257-285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*. 22(2), 123-138.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation*. 55, 37-76.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how?. In *Psychology of learning and motivation*. 57. 147-181.

Van Gog, T., Paas, F., & Sweller, J. (2010). Cognitive load theory: Advances in research on worked examples, animations, and cognitive load measurement. *Educational Psychology Review*. 22(4), 375-378.

Van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review*. 17(2), 147-177.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*. 101(4), 817.