Using Neurophysiological Data to Inform Feedback Timing: A Pilot Study

Jennifer Vogel-Walcutt and Julian Abich

University of Central Florida, 3100 Technology Parkway,
Orlando, FL 32826
jvogel@ist.ucf.edu

Abstract. In an effort to achieve a level of knowledge comparable to that which typically results from individual tutoring, innovative models of adaptive computer-based training are continually being tested and refined. Despite these efforts, adaptive computerized training programs still fall significantly short of the gold standard of one-on-one instruction. In response, this study used a previously developed model defining when to apply instructional feedback during learning in order to improve efficiency. Specifically, we compared the combination of performance and neuro-physiological indices to performance alone as indicators for when to adapt training. Contrary to our hypotheses, this study failed to demonstrate positive impact on knowledge acquisition, knowledge application, perceived cognitive load, or training efficiency. However, based on observational data, it is suspected that participants in neither group possessed enough available working memory capacity to attend to the supporting material. Consequently, this may account for the lack of differential findings.

Keywords: Feedback, EEG, physiological measures, simulation based training, adaptive intelligent systems.

1 Introduction

In an effort to achieve a level of knowledge comparable to that which typically results from individual tutoring, innovative models of adaptive computer-based training are continually being tested and refined [1, 2, 3]. Despite these efforts, adaptive computerized training programs, though superior to traditional classroom-based settings [4], still fall significantly short of the gold standard of one-on-one instruction [5, 6, 7]. In response, previous research has investigated the use of electroencephalogram (EEG) inputs to better inform when to provide adaptive training interventions, finding that workload data, when combined with performance data, can significantly better predict future performance compared to using performance data alone [8]. Based on these data, this paper describes a preliminary investigation using workload measures from the EEG to inform adaptation choices within a Simulation Based Training (SBT) environment.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 265–274, 2011.

1.1 Adaptive Trainers

Current adaptive trainers predominately alter instructional content or strategies for providing trainee support based on performance data alone [9]. It is hypothesized that one of the major reasons human tutors are more effective than these training systems is because a tutor has a richer source of data on which to base adaptations in instructional strategy. Specifically, in addition to performance data, an individual tutor observes a learner's behaviors, reactions, and emotional responses in real-time. Such observational data is subsequently used to inform the prescription of more optimal interventions.

Given the effectiveness of individualized human instruction, significant research has been devoted to identifying how specific components of tutoring impact learning. For example, the use of emotional recognition to inform instructional adjustments results in an improvement of 55% over traditional classroom environments [10]. Another parameter of consideration, and the focus of this paper, is the amount of working memory being utilized, or cognitive load. Working memory has a finite capacity that when exceeded, results in information loss [11]. However, too little cognitive load suggests that the learner is not fully engaged in the activity. Therefore, it has been posited that optimizing cognitive load may result in improved knowledge acquisition, assimilation in long-term memory, and eventual retrieval and application [12] and [3]. Thus, it is expected that consideration of cognitive load during learning, in addition to performance data, may provide additional information about the learner, allowing for more targeted and appropriate instructional intervention.

1.2 Cognitive Load

Cognitive load refers to the amount of working memory capacity utilized to complete a task [13]. It is broken down into three categories: Intrinsic, extraneous, and germane [13] and [14]. Intrinsic load is that which is inherent in the learning material itself. Extraneous load is the amount of working memory capacity expended on information that does not pertain to the learning material and germane load is the mental effort devoted to acquiring and developing schemas [15]. Thus, the goal of instructional design is to maximize the amount of working memory capacity devoted to germane load, minimize the amount of load devoted to extraneous information, and optimize the amount of effort devoted to intrinsic load [14]. To accomplish this goal, two pieces of information are necessary. First, we must understand how the instructional intervention will affect the cognitive load of the learner so that it can be adjusted (provided/removed/altered) as needed. Second, it is necessary to measure the cognitive load of the learner in real-time so that these adjustments can be made appropriately for each individual. In this study, instructional feedback, is the strategy utilized to improve training.

1.3 Feedback and Timing

The extensive reviews of the feedback literature find only a small, positive overall learning impact [16], [17], and [18]. However, the more adaptive the feedback is to the learner's needs, the better impact it has on the learning process [19], [20], and [21]. Therefore, if it is possible to better predict when an individual requires instructional support, it is more likely that feedback will be provided only when needed. Consequently, cognitive overload or distractions during learning can be reduced. Unfortunately, despite their expected positive impact, strategies such as feedback during a lesson or an activity can have the unwanted side effect of overwhelming working memory capacity. As such, the most efficient application of this strategy is needed. One way to improve the efficiency is to utilize real-time indices that can help improve the predictions of when learners require intervention.

1.4 EEG

The EEG can provide real-time cognitive workload measurements during learning tasks. It can be synced with the task timing and provide further insight into the amount of cognitive workload used to complete the task. To date, the EEG has been supported in the literature for its balance of usability and accuracy for measuring cognitive workload [22], [23] and [24]. Several studies have been conducted to validate its use in this context. Berka, et al., [23] was able to use Advance Brain Monitoring's (ABM) wireless headset EEG to discern varying cognitive workload levels during cognitive and assessment tasks using a combination of alpha, beta, and theta wave outputs. After conducting a similar study [25], they utilized the data to develop a generalizable model of cognitive workload that reflects changes in working memory load during learning. Subsequently, using these models, Vogel-Walcutt, et al. [8] were able to combine the output with performance data and better predict future performance when compared to predicting using past performance data alone.

1.5 Current Study

Therefore, based on previous research [8], [23], and [25], this study used the predictive data to develop a model for when to apply instructional feedback during learning in order to improve efficiency. Specifically, we compared the combination of performance and neuro-physiological indices to performance alone as indicators for when to adapt training. Objective, real-time measures of cognitive load were assessed throughout performance using an EEG [Advanced Brain Monitoring (ABM)]. Responses were then categorized into one of four classifications: hits (correct decision, high workload), misses (incorrect decision, high workload), guesses (correct decision, low workload) or slips (incorrect decision, low workload).
or slips (incorrect decision, high workload). During the performance only group, all incorrect decisions received feedback prompts (see fig. 1). In the performance plus workload group, misses and guesses received feedback (see fig. 2).

Our hypotheses were as follows:
Providing instructional feedback prompts during training based on performance plus workload data will result in:

> *H1: more effective knowledge acquisition.*
> *H2: more effective knowledge application.*
> *H3: lower perceived cognitive load during training and assessment.*
> *H4: more efficient training.*

2 Method

2.1 Participants

29 undergraduate students (12 male, 17 female) participated in the current study with a mean age of 18.9. Due to the protected nature of the assessment material, individuals were required to be United States citizens to participate. Further, they were required to have no prior knowledge of the subject matter in order to examine the impact of this training cycle on novice learners. The participants were recruited through a web-based human subject pool management software and were compensated with class credit.

2.2 Materials

Apparatus

Advanced Brain Monitoring's (ABM) wireless EEG sensor. The ABM EEG [23] sensor consists of a wireless headset containing equally distributed sensors throughout the cap and fitting over the head like a small hat. The headset is combined with the B-Alert® Software which is used to extract the user's data in real-time from the EEG so that it can be compared to the current learning experience [23] and [25]. The software output yields the probability of workload levels (the likelihood a person is experiencing high or low workload). Before analysis, all values are standardized to account for individual baseline differences.

Simulation Tasks

Threat-Assessment Training System (ThreATS). ThreATS [26] is a training tutorial used to familiarize participants to the task context. This program consists of an introductory component and two additional levels of training focused on specific decisions participants must make while using the USMC's Deployable Virtual Training Environment (DVTE) simulator. Participants viewed a series of training videos depicting the job of a Fire Support Team (FiST), and specifically, the role of the Forward Observer – Artillery (during a Call For Fire (CFF) task.

Decision-Making Assessment Scenarios (DMAS). The DMAS [26] requires participants to engage in computer simulated "Call for Fire" (CFF) scenarios. Each scenario presents participants with a battlefield that contain friend and enemy targets that are either stationary or moving. Participants must determine the threat level of the targets and use that information to decide the correct warning order (using multiple shots for moving units with imprecise coordinates or use a single shot for static units with precise coordinates), the correct sequence in which to destroy the targets, and the correct method of engagement (determine the proper type of ammunition for each target). Scenario Reference Sheets were provided during the DMAS to help participants distinguish between friend and foe targets, assess the layout of the scenario, and provide information about how to complete the DVTE radio sheet during the simulation.

Measures

Biographical Questionnaire (BQ). This 14-item questionnaire addresses personal identifiers such as age, race, gender, military experience, and degree of comfort with and frequency of use of computers.

Prior Knowledge Questionnaire (PKQ). Developed by the current authors, the PKQ consists of 4 free-response questions assessing their prior knowledge of or experience with the elements of the simulated task.

Cognitive Load Questionnaire (CLQ). The CLQ [15] is a single-item measure of perceived cognitive load during a task or set of tasks. Participants rate subjective impressions of cognitive load on a 9-point Likert-type scale, with higher scores indicating greater perceived cognitive load.

Declarative Knowledge Test (DKT). Developed by the current authors, the DKT consists of 12 factually-based multiple choice items designed to assess the extent to which the participant understands the proper terms used during the CFF task.

Procedural Knowledge Test (PKT). Developed by the current authors, the PKT consists of 7 factually-based multiple choice items designed to assess the extent to which the participant understands the proper execution of the CFF task.

Conceptual Knowledge Test (CKT). Developed by the current authors, the CKT is comprised of 9 factually-based multiple-choice items designed to assess conceptual knowledge regarding the understanding of the components involved in the task.

Integrated Knowledge Test (IKT). Developed for use in the current study, the IKT is comprised of 9 free-response items designed to assess inferences about and deeper knowledge of the FiST.

2.3 Procedure

Participants were split into two groups: Performance Only (PO) and Performance plus Workload (PW). After providing informed consent, participants completed the BQ. For those participants in the PW group, a baseline was established to account for individual differences in resting workload levels.

Introductory Phase - All participants watched a ThreATS video and completed a CLQ. They then completed the DKT, PKT, CKT and a CLQ regarding the experience of filling out the three knowledge tests. Next, a practice scenario in the DMAS was completed to familiarize participants with the simulator. Participants then completed a CLQ regarding the practice mission.

Training Phase - The two training phases of the experiment followed the same pattern. Participants were first asked to watch a training video and then answer the CLQ, after which they conducted a simulated mission in the DMAS. Following the mission, they were again asked to complete the CLQ. After each decision in the

DMAS training missions, participants received a 15-second instructional prompt in the upper right corner of their screen. Prompts were provided based on the heuristics developed for their group (see figs. 1 & 2).

Assessment Phase - Participants completed the DKT, PKT, CKT, IKT, and the CLQ regarding the knowledge tests. Next, participants completed a simulated mission, similar in complexity to that of the second training phase, however, no feedback prompts were provided to either group during this phase. The total time required to complete the entire study was approximately 2.5 hours.

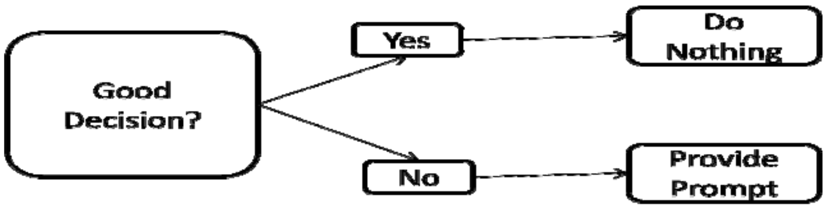

Fig. 1. Performance Group Flowchart

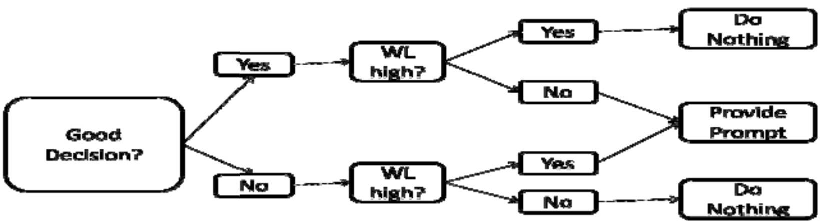

Fig. 2. Neuro-physiological + Performance Group Flowchart

3 Results

H1: Knowledge acquisition - A one-way between subjects analysis of variance (ANOVA) was used to test knowledge acquisition among both groups. Contrary to the hypothesis, the knowledge acquisition scores of the group that received feedback based on performance and neuro-physiological measures did not differ significantly from those of the performance alone group [Declarative $F(1, 27) = .194$, Procedural $F(1, 27) = 1.53$, $p = .227p = .663$, Conceptual, $F(1, 27) = .110$, $p = .742$, and Integrated $F(1, 27) = .083$, $p = .775$; see table 1]. These data suggest that the utilization of neuro-physiological data to indicate optimal intervention timing did not improve knowledge acquisition.

H2: Knowledge application - A one-way between subjects ANOVA was used to test decision making (DM) scores among both groups. Contrary to our hypothesis, DM scores did not differ significantly between groups [$F(1, 27) = .240$, $p = .628$; see table 1]. These data suggest that additional information provided by the neuro-physiological sensors to aid intervention timing did not improve DM skills.

H3: Cognitive load - A one-way between subjects ANOVA was used to compare subjective measures of cognitive load (CL) among both groups. Contrary to our hypothesis, CL scores did not differ significantly between groups during the training ($F(1,27) < .001$, $p = .992$) or assessment phases [$F(1,27) = 3.51$, $p = .072$ (after scenario); $F(1,27) = .134$, $p = .718$ (after tests); see table 2]. These data suggest that perceived cognitive load was not lowered when using neuro-physiological sensors to assist intervention timing.

H4: Efficient training - A one-way between subjects ANOVA was used to test training efficiency among both groups. Contrary to our hypothesis, no significant differences in training efficiency between groups were found [$F(1,27) = 2.52$, $p = .124$]. These data suggest that learning efficiency was not impacted by the additional specificity of the intervention timing.

Table 1. Means and Standard Deviations of Knowledge and Decision-making Scores

		Perf.		Perf. + Neuro	
Variables		M	SD	M	SD
Declarative	Pre	10.67	1.35	10.57	1.40
	Post	10.80	1.37	11.00	1.04
Procedural	Pre	15.73	2.09	14.79	3.97
	Post	17.80	3.36	15.71	5.54
Conceptual	Pre	4.93	1.67	3.93	1.49
	Post	7.53	1.36	7.36	1.50
Integrated	Post	6.70	1.59	6.92	2.42
Decision Making*	Post	-2.23	1.83	-2.57	1.86

Note. $N = 29$. Pre = Completed during introductory phase. Post = Completed during assessment phase. *DM was scored using penalty points; 0 was a perfect score

Table 2. Means and Standard Deviations of Cognitive Load Scores

	Perf.		Perf. + Neuro	
Variables	M	SD	M	SD
Training	5.93	1.34	5.93	1.14
Assessment	4.33	1.72	5.43	1.40
Knowledge Tests	5.60	1.24	5.43	1.28

Note. $N = 29$.

4 Discussion

The goal of the study was two-fold. First, we aimed to utilize neuro-physiological measures of cognitive load in combination with performance data to better classify learners' errors. Second we aimed to use these data to help inform an adaptive training system to provide tailored feedback at an optimal time to those individuals requiring intervention. Contrary to our hypotheses, however, this study failed to demonstrate positive impact on knowledge acquisition, knowledge application, perceived cognitive load, or training efficiency when incorporating these measures. The current results indicate that the proposed solution failed to positively impact training effectiveness and efficiency in its current form. However, based on observational data some inferences can be made regarding the lack of impact. For example, it was observed that few participants looked at the provided instructional feedback prompts. This may be due to the highly visual-strain of the task. It is possible that participants in neither group possessed enough available working memory capacity to attend to the supporting material. Consequently, this may account for the lack of differential findings.

4.1 Limitations of Current Study

Several limitations may have led to the lack of differences between groups including, task shedding, task overload, and small sample size. Task shedding may result as a consequence of task overload. Based on current Cognitive Load Theory (CLT) [11] and [13], it is believed that working memory has a limited capacity. It is possible that participants were already utilizing their full working memory capacity during the activity and were therefore unable to additionally attend to the feedback that was provided during this highly visual task. Since the feedback provided was visual, it may have overloaded the visual channel of working memory, leading to the inattention or inability to process the information in the feedback prompts.

Additionally, as in most studies, sample size may be another factor to account for the lack of significant differences. It is difficult to make inferential conclusions based on a weak sample size, since larger sampling error tends to be present in small samples. Accordingly, the data should be considered with caution.

4.2 Future Research

Future efforts may benefit from altering feedback placement sizing (i.e. full screen prompts). If it was the case that participants were already cognitively overloaded by the intrinsic load of the task, then it may be necessary to close the simulation in order to allow participants sufficient time to read and digest the instructional material. This could be accomplished by providing full screen prompts. In doing so, participants may experience a temporary reduction in task load which in turn may reduce task shedding. Otherwise, seeking an alternative modality all together, such as utilizing auditory prompting, may be more effective. Lastly, increasing the sample size based on a power analysis would improve the statistical power and ultimately lead to more confident conclusions.

Acknowledgement. This work is supported in part by the Office of Naval Research Grant N0000141010113. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ONR or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

1. Paquette, G.: Designing Virtual Learning Centers. In: Adelsberger, H.H., Pawlowski, J.M., Collis, B. (eds.) Handbook of Information Technologies for Education and Training, pp. 249–272 (2002)
2. McLaughlin, Luca, J.: A learner-centered approach to developing team skills through web-based learning and assessment. British Journal of Educational Technology 33(5), 571–582 (2002)
3. Kalyuga, S., Sweller, J.: Rapid Dynamic Assessment of Expertise to Improve the Efficiency of Adaptive E-learning. Educational Technology Research and Development 55(3), 83–93 (2005)
4. Bayraktar, S.: A meta-analysis of the effectiveness of computer-assisted instruction in science education. Journal of Research on Technology in Education 34(2), 173–188 (2001-2002)
5. Bloom, B.S.: The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13(6), 4–16 (1984)
6. Shute, V.J., Psotka, J.: Intelligent tutoring systems: Past, Present and Future. In: Jonassen, D. (ed.) Handbook of Research on Educational Communications and Technology, Scholastic Publications (1994)
7. Iglesias, A., Martinez, P., Aler, R., Fernandez, F.: Learning teaching strategies in an adaptive and intelligent educational system through reinforcement learning. Applied Intelligence 31, 89–106 (2009)
8. Vogel-Walcutt, J.J., Marino-Carper, T., Bowers, C., Nicholson, D.: Utilizing Learners' Internal States to Drive Feedback Decisions: A Preliminary Investigation (Manuscript under review) (2010)
9. Bolton, A., Campbell, G., Schmorrow, D.D.: Towards a closed-loop training system: Using a physiological-based diagnosis of the trainee's state to drive feedback delivery choices. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 409–414. Springer, Heidelberg (2007)
10. Porayska-Pomsta, K., Mavrikis, M., Pain, H.: Diagnosing and acting on student affect: the tutor's perspective. UMUAI 18(1-2), 125–173 (2008)
11. Sweller, J., Van Merrienboer, J.J.G., Paas, F.G.W.C.: Cognitive architecture and instructional between feedback timing, content and modality under high cognitive workload. Educ. Psych. Rev. 10, 251–296 (1998)
12. Mousavi, S.Y., Low, R., Sweller, J.: Reducing cognitive load by mixing auditory and visual presentation modes. J. Educ. Psych. 87(2), 319–334 (1995)
13. van Merrienboer, J.J.G., Sweller, J.: Cognitive load theory and complex learning: Recent developments and future directions. Educ. Psych. Rev. 17(2), 147–177 (2005)
14. Paas, F., Renkl, A., Sweller, J.: Cognitive Load Theory: instructional implications of the interaction between information structures and cognitive architecture. Instr. Sci. 32(1-2), 1–8 (2003)
15. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.M.: Cognitive load measurement as a means to advance cognitive load theory. Educ. Psych 38(1), 63–71 (2003)
16. Kluger, A.N., DeNisi, A.: Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psych. Bull. 119(2), 254–284 (1996)
17. Bangert-Drowns, R.L., Kulik, C.-L.C., Kulik, J.A., Morgan, M.T.: The instructional effect of feedback in test-like events. Rev. Educ. Res. 61(2), 213–238 (1991)

18. Mory, E.H.: Feedback research revisited. In: Jonassen, D.H. (ed.) Handbook of Research for Educational Communications and Technology, Simon & Schuster Macmillan, New York (2004)
19. Shute, V.J.: Focus on formative feedback. Rev.Educ. Res. 78(1), 153–189 (2008)
20. Dieterle, E., Murray, J.: Realizing adaptive instruction (Ad-In): The convergence of learning, instruction, and assessment. In: Presented at the 13th annual conference on HCII, San Diego, CA, July 19-24 (2009)
21. Wulfeck, W.: Adapting instruction. In: Presented at the 13th annual conference on human computer interaction international, San Diego, CA, July 19-24 (2009)
22. Shute, V.J., Zapata-Rivera, D.: Using an evidence-based approach to assess mental models. In: Ifenthaler, D., Pirnay-Dummer, P., Spector, J.M. (eds.) Understanding models for learning and instruction: Essays in Honor of Norbert M. Seel, pp. 23–41. Springer, Heidelberg (2008)
23. Luu, P., Poulson, C., Tucker, D.: Neurophysiological measures of brain activity: Going from the scalp to the brain. In: Presented at the 13th annual conference on HCII, San Diego, CA, July 19-24 (2009)
24. Berka, C., et al.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat Space Environ. Med. 78(5), 231–244 (2004)
25. St. John, M., Kobus, D.A., Morrison, J.G., Schmorrow, D.: Overview of the DARPA augmented cognition technical integration experiment. IJHCI 17(2), 131–149 (2004)
26. Berka, C., et al.: Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. IJHCI 17(2), 151–170 (2007)
27. Vogel-Walcutt, J.J., Nicholson, D.: Applied Learning Science Team Update. In: Paper presented at the ONR HPT&E: AITE program review, San Diego, CA, January 22 (2009)