EEG-Related Changes in Cognitive Workload, Engagement and Distraction as Students Acquire Problem Solving Skills

Ronald H. Stevens[1], Trysha Galloway[1], and Chris Berka[2]

[1] UCLA IMMEX Project, 5601 W. Slauson Ave. #255, Culver City, CA 90230
immex_ron@hotmail.com,
tryshag@gmail.com
[2] Advanced Brain Monitoring, Inc, Carlsbad, CA 90045
chris@b-alert.com

Abstract. We have begun to model changes in electroencephalography (EEG)-derived measures of cognitive workload, engagement and distraction as individuals developed and refined their problem solving skills in science. For the same problem solving scenario(s) there were significant differences in the levels and dynamics of these three metrics. As expected, workload increased when students were presented with problem sets of greater difficulty. Less expected, however, was the finding that as skills increased, the levels of workload did not decrease accordingly. When these indices were measured across the navigation, decision, and display events within the simulations significant differences in workload and engagement were often observed. Similarly, event-related differences in these categories across a series of the tasks were also often observed, but were highly variable across individuals.

1 Introduction

Skill development has been described as occurring in stages that are characterized by distinctive amounts of time and mental effort required to exercise the skill [1] [10]. Given the complexities of skill acquisition it is not surprising that a variety of approaches have been used to model the process. For instance, some researchers have explored the improved powers of computation in combination with machine learning tools to refine models of skill acquisition and learning behaviors in science and mathematics. Such systems rely on learner models that include continually updated estimates of students' knowledge and misconceptions based on actions such as choosing an incorrect answer or requesting a multimedia hint. Although such learner models are capable of forecasting student difficulties, [12] or identifying when students may require an educational intervention, they still rely on relatively impoverished input due to the limited range of learner actions that can be detected by the tutoring system (e.g., menu choices, mouse clicks) and latency.

Application of neurophysiologic approaches, including the quantification of EEG correlates of workload, attention and task engagement have also been used to provide objective evidence of the progression from stage 2 to stage 3 [2] [3]. There is a large

C. Conati, K. McCoy, and G. Paliouras (Eds.): UM 2007, LNAI 4511, pp. 197–206, 2007.

and growing literature on the EEG correlates of attention, memory, and perception [5], although there is a relative dearth of EEG investigations of the process skill acquisition and learning. EEG researchers have generally elected to employ study protocols that utilize training-to-criterion to minimize variability across subjects and to ensure stable EEG parameters could be characterized. In most studies, the EEG data is not even acquired during the training process leaving a potentially rich data source untapped.

Thus, while advanced EEG monitoring is becoming more common in high workload / high stress professions (such as tactical command, air traffic controllers) the ideas have not been comprehensively applied to real-world educational settings, due in part to some obvious challenges. First, the acquisition of problem solving skills is a gradual process and not all novices solve problems in the same way, nor do they follow the same path at the same pace as they develop domain understanding. Next, given the diversity of the student population it is difficult to assess what their relative levels of competence are when performing a task making it difficult to accurately relate EEG measures to other measures of skill. This is further complicated as strategic variability makes analyzing the patterns of students' problem solving record too complicated, costly, and time consuming to be performed routinely by instructors. Nevertheless, there are many aspects of science education that could benefit from deriving data from advanced monitoring devices and combining them with real-time computational models of the tasks and associated outcomes conditions.

This manuscript describes a beginning synthesis of 1) a probabilistic modeling approach where detailed neural network modeling of problem solving at the population level provides estimates of current and future competence, and, 2) a neurophysiologic approach to skill acquisition where real-time measures of attention, engagement and cognitive work load dynamically contribute estimates of allocation of attention resources and working memory demands as skills are acquired and refined.

2 Methods

2.1 The IMMEX™ Problem Solving Environment

The software system used for these studies is termed IMMEX™ whose program structure is based on an extensive literature of how students select and use strategies during scientific problem solving [6] [15].

To illustrate the system, a sample biology task called *Phyto Phyasco* provides evidence of a student's ability to identify why the local potato plants are dying. The problem uses a multimedia presentation to explain the scenario and the student's challenge is to identify the cause. The problem space contains 5 Main Menu items which are used for navigating the problem space, and 38 Sub Menu items describing local weather conditions, soil nutrients, plant appearance, etc. These are decision points, as when the student selects them, s/he confirms that the test was requested and is then presented the data. When students feel they have gathered the information needed to identify the cause they attempt to solve the problem.

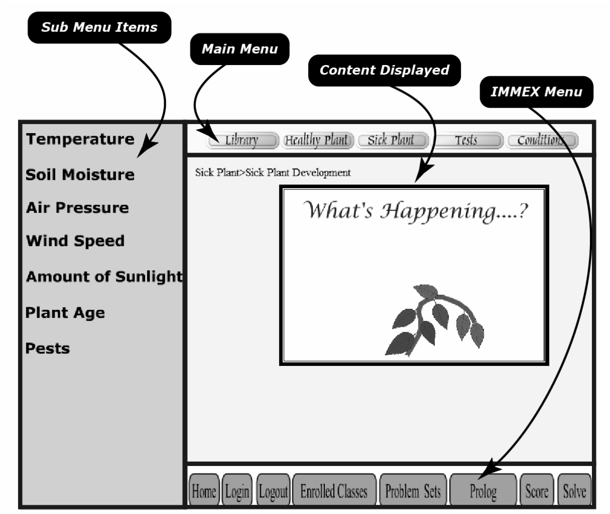


Fig. 1. Sample IMMEX™ simulation. In the Phyto Phyasco simulation, the farmer's potato plants are dying and the challenge for the student is to identify the cause by examining local weather conditions, nutrients, etc. Students navigate throughout the problem space using the Main Menu items and select data resources and make decisions using the Sub Menu Items. The resulting data is shown in the Display.

The IMMEX database serializes timestamps of how students use these items, which are then used to train competitive, self-organizing ANN [11]. As IMMEX problem sets contain many parallel cases learning trajectories can then be developed through Hidden Markov Modeling (HMM) that not only reflect and model students' strategy shifts as they attempt series of cases, but also predict future problem solving performance.

Students often begin by selecting many test items, and consistent with models of skill acquisition [4], refine their strategies with time and select fewer tests, eventually stabilizing with an approach that will be used on subsequent problems. As expected, with practice solve rates increase and time on task decreases. The rate of stabilization, and the strategies stabilized with are influenced by gender, experience [13], and individual or group collaboration. Students often continue to use these stabilized strategies for prolonged periods of time (3-4 months) when serially re-tested [11].

IMMEX problem solving therefore represents a task where it is possible to construct probabilistic models of many different aspects of problem solving skill acquisition. The constraints of working memory are likely to be relevant during such skill acquisition where working memory capacity can become exceeded, and the ability to combine probabilistic performance models with EEG workload metrics could shed light on how different working memory capacities are needed as students gain experience and begin to stabilize their strategies?

2.2 The B-Alert®System

A recently developed commercial wireless EEG sensor headset has combined a battery-powered hardware and sensor placement system to provide a lightweight, easy-to-apply method to acquire and analyze six channels of high-quality EEG (Advanced Brain Monitoring, Inc. Carlsberg, CA). This headset requires no scalp preparation and provides a comfortable and secure sensor-scalp interface for 12 to 24 hours of continuous use. Standardized sensor placements include locations over frontal, central, parietal and occipital regions (sensor sites: F3-F4, C3-C4, Cz-PO, F3-Cz, Fz-C3, Fz-PO). Data are sampled at 256 samples/second with a bandpass from 0.5 Hz and 65Hz (at 3dB attenuation). Quantification of the EEG in real-time, referred to as the B-Alert® system, is achieved using signal analysis techniques to identify and decontaminate fast and slow eye blinks, and identify and reject data points contaminated with excessive muscle activity, amplifier saturation, and/or excursions due to movement artifacts. Wavelet analyses are applied to detect excessive muscle activity (EMG) and to identify and decontaminate eye blinks.

2.3 Subjects and Study

Subjects (n=12) first performed a single 30-minute baseline EEG test session to adjust the software to accommodate individual differences in the EEG (Berka, 2004). They then performed multiple IMMEX problem sets targeted for 8^{th}-10^{th} grade students.

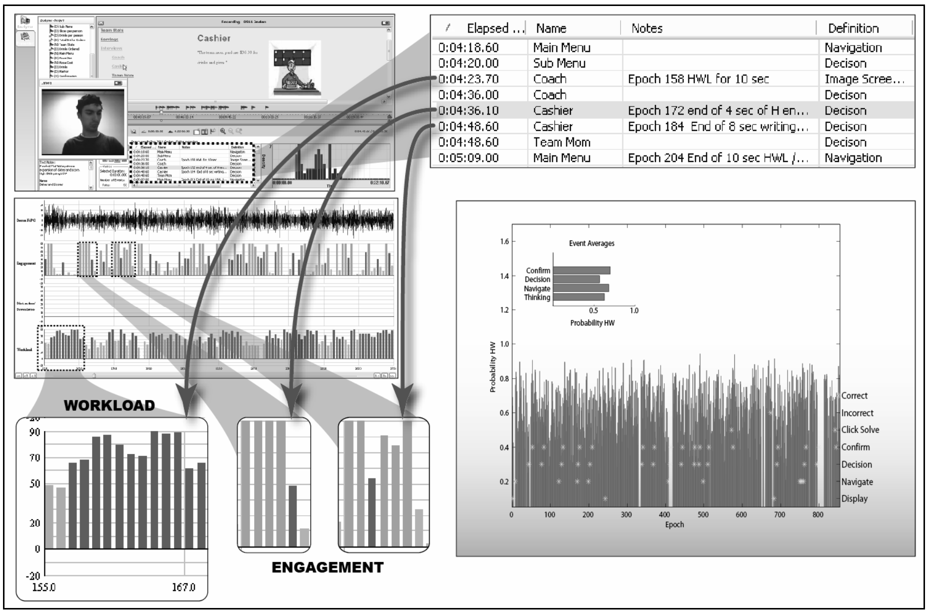


Fig. 2. Relating EEG Workload and Engagement Indexes with Problem Solving Events. The user (not described in the text) is shown engaged in IMMEX problem solving while keyboard and mouse events are simultaneously recorded. Below shows the real-time output of the B-Alert cognitive indexes where workload and engagement data streams were linked with events in the log. In the lower right corner, the timestamps of IMMEX data requests and displays are integrated with the EEG workload indices and then plotted against the one-second epochs of the task. The upper left histograms average the workload indices for each of the IMMEX events including the one second prior to and after the event.

These included *Phyto Pyiasco*, the biology problem described above, *Get Organized* where the goal is to diagnose disorders of organ systems, and a mathematics problem called *Paul's Pepperoni Pizza Palace*.

Subjects generally performed at least 3 cases of each problem set allowing the tracking of strategies and cognitive changes across problem sets as well as cases while students gained experience. Then we aligned the EEG output metrics on a second-by-second basis with the problem solving actions to explore the within-task EEG metric changes. For this alignment, we used software (Morea, Techsmith, Inc.) that captures output from the screen, mouse click and keyboard events as well as video and audio output from the users (Figure 2).

The B-Alert software output includes EEG metrics (from 0.1-1.0) for distraction (DT), engagement (E), and workload (WL) calculated for each 1-second epoch using quadratic and linear discriminant function analyses of model-selected EEG variables derived from power spectral analysis of the 1-Hz bins from 1-40Hz.

These metrics have proven utility in tracking both phasic and tonic changes in cognitive states, and in predicting errors resulting from either fatigue or overload [3]. The cognitive indices are expressed as histograms for each 1-second epoch of the problem solving session and show the probability of WL, E, or DT. By integrating B-Alert and IMMEX data request time stamps, the navigation, decision, and display-related events are then overlaid onto the cognitive indices.

3 Results

3.1 Distributions of Engagement, Distraction and Workload During IMMEX Problem Solving

Figure 3 illustrates the dynamics of the B-Alert EEG measures during IMMEX problem solving for six students over a ten-minute period. In each window, the top display is E, the middle is DT and the bottom is E. Each bar in the histograms represents averaged metrics at 1-second epochs.

Panels A, C and to a lesser extend F most closely represents students who were productively engaged in problem solving; workload levels were moderate and the levels were alternating with cycles of high engagement. Many cycles were associated with navigation and interpretation events (data not shown). Panel B illustrates a student who may be experiencing difficulties and might not be prepared to learn. The workload and engagement levels were low and distraction was consistently high.

The student in Panel D encountered a segment of the simulation that induced 10-15 seconds of distraction (middle row) and decreased workload and engagement. Through the data interleaving process the data that the student was looking at was retrieved, which in this case was an animation of a growing plant. Panel E shows a student who, while not distracted, appeared to be working at beyond optimal capacity with workload levels consistently near 100%. Probabilistic performance models for this student [11] [13] suggested a difficulty in developing efficient strategies on his own.

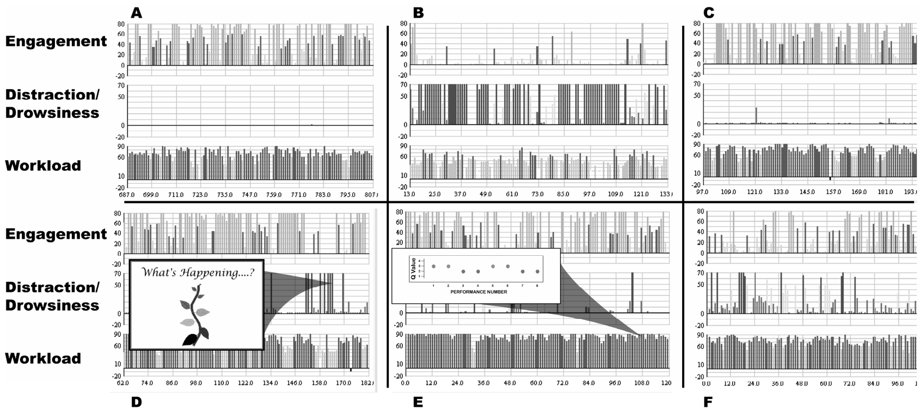


Fig. 3. Dynamics of WL, D and E for Six Students on IMMEX Tasks. This figure shows 10 minute segments of the B-Alert cognitive metrics while students performed IMMEX problems.

3.2 Increases in Problem Solving Skills Are Not Accompanied by Decreases in Cognitive Workload or Engagement

We next measured the seconds needed to solve the first, second and third cases of *Paul's Pepperoni Pizza* (n=7) and calculated the average WL and E across these three performances. As shown in Table 1, while the time needed to complete the task significantly decreased, there were no significant changes in either WL or E.

Table 1. Changes in Time on Task, WL and E With Problem Solving Experience

Performance	Speed (seconds)	WL	E
1	422 ± 234	.629 ± .07	.486 ± .09
2	241 ± 126	.625 ± .08	.469 ± .08
3	136 ± 34	.648 ± .06	.468 ± .09

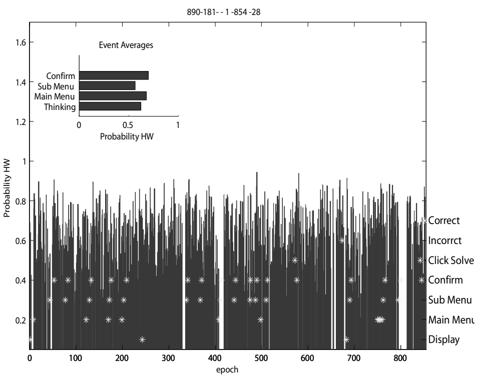
3.3 Students Apply Similar Workload to Similar Problems and More Workload to More Difficult Problems

Five students also performed 3 cases of *Phyto Phyasco* which is also a middle school IMMEX problem. There were no significant differences between the WL (0.64 ± .05 vs. 0.63 ± .05, p =.42) and E (0.51 ± .07, 0.51 ± .04, p = .92) across the two problem sets. Two individuals also solved the more difficult high school chemistry problem Hazmat. For both of these individuals the WL was significantly greater for the three cases of *Hazmat* than for *Paul's Pepperoni Pizza*. (Subject 103: 0.76 ± .02 vs. 0.71 ± .03, p< 0.001; Subject 247: 0.57 ± .02 vs. 0.49 ± .03, p< 0.005).

Five of the students missed one or more of the cases in the problem set and a paired samples test was performed to determine if, at a performance level, differences existed in WL, E, or DT when the subjects were correctly, or incorrectly, solving a problem. None of these differences were significant.

3.4 The Navigation and Decision-Related Events in IMMEX May Be Behaviorally Relevant

We next increased the granularity of the analysis by dividing performances into segments related to problem framing, test selections, confirmation events where the student decides whether to select data, and closure where the student decides on the problem solution. We then compared the WL and E values across the different events within the different single IMMEX performances. The WL and E values at each subtask boundaries [7] (e.g. Main Menu, Sub Menu, etc.), as well as the epochs immediately before and after the event were averaged across the problem set. As shown in Figure 4 there were often significant differences among these averages at the different events. These differences, however, were neither uniform nor predictable across individuals or tasks.



Mean Thinking = 0.628
Mean Main Menu = 0.681
Mean Sub Menu = 0.568
Mean Confirm = 0.699
Significance Thinking vs Main Menu = 0.101
Significance Thinking vs Sub Menu = 0.014
Significance Thinking vs Confirm = 0.003
Significance Main Menu vs Sub Menu = 0.009
Significance Main Menu vs Confirm = 0.571
Significance Sub Menu vs Confirm = 0.000

Fig. 4. Linking Cognitive Workload Indices with IMMEX-related Events. Left: The timestamps of IMMEX data requests and displays are integrated with the EEG workload indices and then plotted against the one-second epochs of the task. The upper left histograms average the workload indices for each of the IMMEX events including the one second prior to and after the event. Right: Table of significant differences between WL events.

We more closely examined events from one student who performed the IMMEX mathematics problem *Paul's Pepperoni Pizza*. The particular student being illustrated missed solving the first case, correctly solved the second case, and then missed the third case indicating that an effective strategy had not yet been formulated.

The problem framing event was defined as the period from when the Prologue first appeared on the screen until the first piece of data information is chosen. For this subject the HWL decreased from the first to the third performance (.72 ± .11 vs. .57 ± .19, $t = 28.7$, $p < .001$), and engagement increased .31 ± .30 vs. .49 ±.37, $t = 4.3$, $p <.001$). The decreased workload was similar to that observed in other subjects; the increasing E may relate more to the student missing the problem. During the decision-making process, students often demonstrated a cycling of the B-Alert cognitive indexes characterized by relatively high workload and low engagement which then switched to lower workload and higher engagement (Figure 5). The cycle switches were often, but not always, at boundaries associated with the selection of new data.

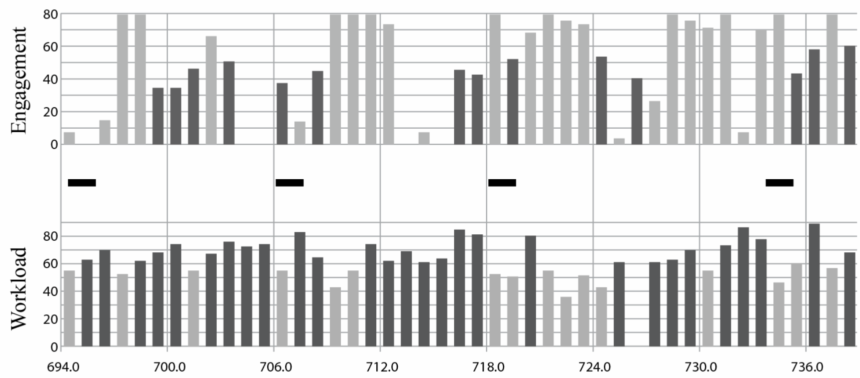
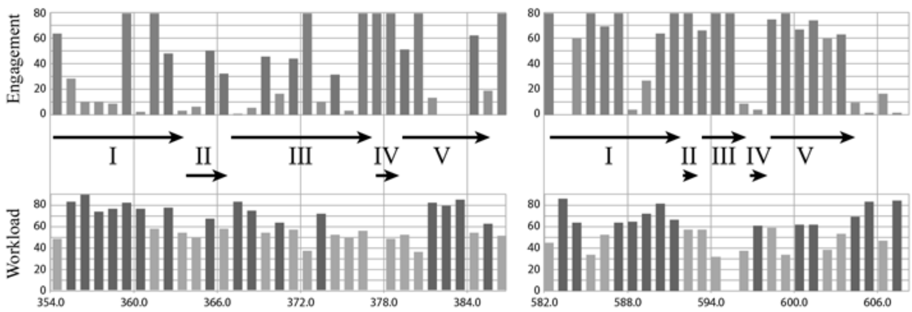


Fig. 5. Fluctuations in WL and E during Problem Solving. The bars indicate the epochs where the student made test selections.

The closing sequences of a problem are complex where the student first makes an irrevocable decision to attempt a solution. Then, the he must make a selection choice from an extensive list of possible solutions. Finally, they must confirm their choice. After that they receive feedback on their success / failure; the students have two such solution attempts. The dynamics of WL and E for one student's first and second solution attempts of *Paul's Pepperoni Pizza* are shown in Fig. 6.

In the 10 seconds before solving the problem (epochs 354 – 364 (I)) there was WL which decreased as the student made his decision (II, III). Two seconds before the student confirmed his choice (epoch 377, IV) there was an increase in engagement which was maintained as the student realized that the answer was incorrect (V).



(a) (b)

Fig. 6. (a) Workload and Engagement Events Related to Problem Closure on the First Attempt. (b) Workload and Engagement Events Related to Problem Closure on the Second Attempt.

The workload and engagement dynamics were different on the second solution attempt. Here there was less WL and more E in the 10 seconds leading up to the decision to solve the problem (Epochs 582- 592, (I, II). At epoch 593 the choice to continue was confirmed, and two seconds before making this decision engagement increased and was maintained during the selection and confirmation process. Between epochs 593 and 596 an incorrect answer was chosen and confirmed (III, IV). At epoch 597 the selection was made and the student learned of the incorrect answer (V).

4 Discussion

We have described a web-based data acquisition architecture and event interleaving process that allows us to begin to map EEG-derived cognitive indices to behaviorally relevant aspects of the students problem solving. An unusual feature of these studies was the application of these technologies to every-day classroom activities that are quite distinct from highly controlled laboratory tasks. In this regard the studies mirrored, and experienced similar challenges of aligning WL measures with subtask boundaries, that were reported by Iqbal et al., [7], and Lee & Tan, [8]

As expected, WL increased when students were presented with problem sets of greater difficulty. Less expected, however, was the finding that as skills increased, the levels of WL did not decrease accordingly; suggesting significant mental commitment may be involved during strategic refinement. Given the anticipated differences between individual students' experience and knowledge we have focused our studies on comparing differences within individuals as skills are developed, rather than extensively compare across individuals.

By restricting the analyses to the seconds surrounding relevant problem solving events such as menu navigation and decision making more refined views of the changing dynamics of WL and E were obtained as skills were refined. Nevertheless, these measurements still accounted for only a small portion of the cognitive workload of the total performance suggesting the need for a finer grained analysis between these events. To this end, we have begun recording videos of the problem solving process as well as of the user on a second by second basis and interleaving them with EEG cognitive indices through log files generated by the problem solving application, the video recording software and the EEG acquisition system. With this more refined system we anticipate being able to link the majority of the WL and E fluctuations to observable events.

Acknowledgment. Supported in part by grants from the National Science Foundation (NSF-ROLE 0231995, DUE Award 0126050, ESE 9453918) and the U.S. Department of Education (R305H050052).

References

[1] Anderson, J.R.: Acquisition of Cognitive Skill. Psychological Review 89, 369–406 (1982)

[2] Berka, C., Levendowski, D.J., Cvetinovic, M.M., Davis, G.F., Lumicao, M.N., Popovic, M.V., Zivkovic, V.T., Olmstead, R.E.: Real-Time Analysis of EEG Indices of Alertness, Cognition and Memory Acquired with a Wireless EEG Headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)

[3] Berka, C., Levendowski, D.J., Ramsey, C.K., Davis, G., Lumicao, M.N., Stanney, K., Reeves, L., Harkness, R.S., Tremoulet, P.D., Stibler, K.: Evaluation of an EEG-Workload Model in an Aegis Simulation, Biomonitoring for Physiological and Cognitive Performance during Military Operations, edited by John Caldwell, Nancy Jo Wesentsten. In: Proceddings of SPIE, vol. 5797, pp. 90–99 (2005)

[4] Ericsson, K.A.: Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. Academic Medicine 79(10), S70–S81 (2004)

[5] Fabiani, M., Gratton, G., Coles, M.G.: Event-related Brain Potentials. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (eds.) Handbook of Psychophysiology, pp. 53–84. Cambridge University Press, Cambridge (2000)

[6] Haider, H., Frensch, P.A.: The Role of Information Reduction in Skill Acquisition. Cognitive Psychology 30, 304–337 (1996)

[7] Igbal, S.T., Adamczyk, P.D., Zheng, X.S., Bailey, B.P.: Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, Portland, Oregon. USA (2005)

[8] Lee, J.C., Tan, D.S.: Using a Low-cost Electroencephalogramph for Task Classification in HCI Research. In: UIST '06. Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (Montreux, Switzerland), October 15-18, pp. 81–90. ACM Press, New York, NY (2006)

[9] Poythress, M., Russell, C., Siegel, S., Tremoulet, P.D., Craven, P.L., Berka, C., Levendowski, D.J., Chang, D., Baskin, A., Champney, R., Hale, K., Milham, L.: Correlation between Expected Workload and EEG Indices of Cognitive Workload and Task Engagement. In: Proceedings of 2nd Annual Augmented Cognition International Conference, San Francisco, CA (In press)

[10] Schneider, W., Shiffrin, R.M.: Controlled and Automatic Human Information Processing I: Detection, Search, and Attention. Psychological Reviews 84, 1–66 (1977)

[11] Stevens, R., Casillas, A.: Artificial Neural Networks. In: Mislevy, R.E., Williamson, D.M., Bejar, I. (eds.) Automated Scoring, pp. 259–312. Lawrence Erlbaum, Mahwah (2006)

[12] Stevens, R., Johnson, D.F., Soller, A.: Probabilities and Predictions: Modeling the Development of Scientific Competence. Cell Biology Education, vol. 4(1), pp. 42–57. The American Society for Cell Biology (2005)

[13] Stevens, R., Soller, A., Cooper, M., Sprang, M.: Modeling the Development of Problem Solving Skills in Chemistry with a Web-Based Tutor. In: Lester, J.C., Vicari, R.M., Paraguaca, F. (eds.) Intelligent Tutoring Systems. 7th International Conference Proceedings, pp. 580–591. Springer-Verlag, Berlin Heidelberg, Germany (2004)

[14] Stevens, R., Wang, P., Lopo, A.: Artificial Neural Networks can Distinguish Novice and Expert Strategies during Complex Problem Solving. JAMIA 3(2), 131–138 (1996)

[15] VanLehn, K.: Cognitive Skill Acquisition. Psychology 47, 513–539 (1996)