

ARTFuL: Adaptive Review Technology for Flipped Learning

Daniel Szafir, Bilge Mutlu

Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706, USA
{ dszafir, bilge }@cs.wisc.edu

ABSTRACT

Internet technology is revolutionizing education. Teachers are developing massive open online courses (MOOCs) and using innovative practices such as flipped learning in which students watch lectures at home and engage in hands-on, problem solving activities in class. This work seeks to explore the design space afforded by these novel educational paradigms and to develop technology for improving student learning. Our design, based on the technique of *adaptive content review*, monitors student attention during educational presentations and determines which lecture topic students might benefit the most from reviewing. An evaluation of our technology within the context of an online art history lesson demonstrated that adaptively reviewing lesson content improved student recall abilities 29% over a baseline system and was able to match recall gains achieved by a full lesson review in less time. Our findings offer guidelines for a novel design space in dynamic educational technology that might support both teachers and online tutoring systems.

Author Keywords

Massive open online course (MOOC); flipped learning; adaptive user interfaces (AUI); brain-computer interfaces (BCI); electroencephalography (EEG); adaptive content review; information recall; learning

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces – *input devices and strategies, evaluation/methodology, user-centered design*

INTRODUCTION

The rise of internet media is transforming the educational landscape by serving as a medium for educational material that can transcend the traditional barriers of institutional access. Many organizations, including universities that publish didactic multimedia lectures and companies such as Khan Academy¹ and

¹ <http://www.khanacademy.org/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2013, April 27–May 2, 2013, Paris, France.

Copyright 2013 ACM 978-1-4503-1899-0/13/04...\$15.00.

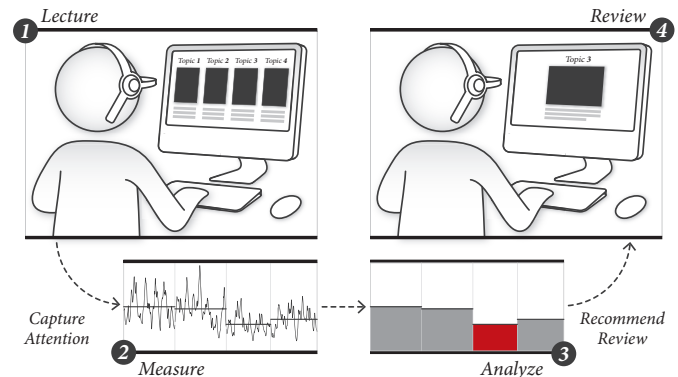


Figure 1. This work presents a novel educational system that (1) instructs users while (2) measuring attention across predefined lesson modules. Following the lesson, the system (3) analyzes the attention measurements to (4) adaptively determine review content that might best improve learning.

Udacity,² that offer videos specifically targeted at educational enrichment, have gained wide audiences by taking advantage of the ease of online posting to create massive repositories of free online educational material. Such online education places technology at the center of an emerging paradigm that has the potential to revolutionize classroom education, variously termed “flipped” [43] or “inverted” [23] learning. In these settings, students are assigned online lectures to watch on their own schedule, while classroom time is devoted to answering questions, working on problem sets, experimental activities, and getting help from the instructor.

Although online education may prove to be an effective supplement to traditional classroom education and a doorway to innovations such as flipped classrooms, there are still significant challenges surrounding this model that must be addressed. Just as “teachers need knowledge of how to organize the classroom to maximize student learning” [50], educational software needs knowledge of how to organize and present lesson material to optimize interactions with students who may have lost focus or may be multitasking. Unfortunately, current computer-based education technology, particularly online educational content that often takes the form of a video or PowerPoint lecture, is largely solitary, non-social, static, and lacking in monitoring capabilities that could offer contextual information on how well the student paid attention to the lecture. These drawbacks stand in sharp contrast to high-immmediacy, dynamic classrooms in which teachers monitor

² <http://www.udacity.com/>

students during instruction and students can seek clarifications and feedback. Despite current limitations, self-directed, media-based learning, and computer-based education in general, continue to rise in popularity. How might new technology address the challenges and support the unique opportunities that this novel educational paradigm affords?

One potential approach toward developing technology that supports online education is to utilize effective teacher techniques such as questioning students to encourage self-explanation, varying content order or presentation speed, and reviewing content with students in order to create personalized educational experiences. Using this model, we propose a novel solution by designing a computer-based education system, incorporating *adaptive content review* technology, which monitors student attention during initial lesson presentation and determines in real time the optimal review topic for the student (Figure 1). This design is implemented in a prototype system and evaluated in an experimental lesson scenario to determine how adaptive content review might affect actual and perceived student learning.

RELATED WORK

Our design was informed by previous research on supporting education using computer technology. Although the emerging phenomena of flipped learning represents a relatively nascent research area, online educational paradigms are part of a continuous evolution from older, more well-studied computer-based education practices. Below, we briefly review the progress of research supporting education with technology, from intelligent tutors to virtual classrooms, eventually leading to the innovation of flipped learning. We then examine the challenges unique to flipped learning scenarios.

Intelligent Tutoring Systems

Computer technology holds significant promise in supporting education via computer aided-instruction (CAI) systems, which provide students with feedback and hints on their answers, and more advanced intelligent tutoring systems (ITS), which offer a finer granularity of interaction by providing hints and scaffolding in problem sub-steps. Such computer-based educational systems have proven effective in reducing instruction time while increasing student learning gains and attitudes towards education [21]. Additionally, computer-based education has shown improvement over traditional classroom education in terms of academic achievement [1] and student motivation [40].

The rise of the internet and virtual classroom technologies such as Blackboard³ and SLOODLE⁴ [19], have pushed computer-based education into the mainstream, providing even more incentive for effective ITS technologies. Modern ITS applications, which rely on an embedded assessment of student knowledge to better present instructional material, have proven nearly as effective as human tutoring in improving learning gains [47]. To further improve ITS technology, recent research has explored additional methods of gauging student knowledge, engagement with the instruction, and affective

³ <http://www.blackboard.com/>

⁴ <http://www.sloodle.org/moodle/>

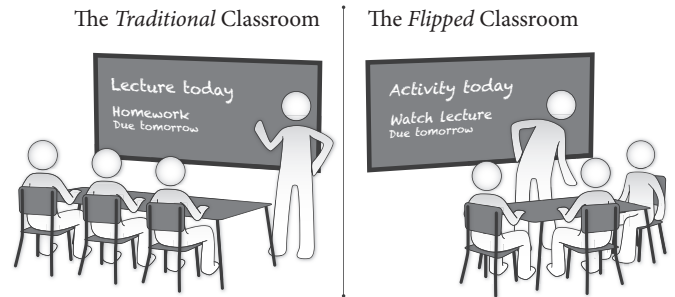


Figure 2. In flipped learning, activities that normally take place in the classroom happen during students' own time, and vice versa (adapted from [20]).

state, including measuring student eye movements, posture, heart rate, skin conductance, and electroencephalography signals [9, 30, 49]. ITS incorporating such additional measures must consider tradeoffs, balancing learning gains with potential negative effects such as cost of technology and discomfort that may arise from requiring students to wear sensors. However, decreasing sensor costs and advances in sensor technology, including miniaturization and utilizing wireless data transfers, help to mitigate these concerns.

Recently, hybrid classroom-ITS approaches have arisen, creating blended classrooms and distance learning environments in which instruction and example problems are shown in online videos, PowerPoint presentations, and online ITS [22]. While new monikers, including "e-Learning 2.0," "Adaptive Educational Hypermedia Systems," and "Learning Management Systems" have arisen to describe the current state of the art in online classroom environments, such systems are still in the experimental stage, offering limited capabilities for content authoring and course administration [26].

Flipped Learning

One emerging model for utilizing internet technology to support education is the *flipped learning* paradigm in which events that traditionally happen within the classroom take place during the students' own time, while work that is usually considered individual homework happens collaboratively in the classroom (Figure 2). Such inverted classrooms are motivated by many factors including devoting class time toward encouraging critical thinking [22] and student collaboration [42], supporting different student learning styles [23], and addressing the needs of students who have grown up with online media technology [6, 10, 17]. Although flipped learning is still an experimental strategy in need of additional study and objective learning results, studies suggest that it increases student perceptions regarding college classes [23], cooperative and innovative learning [42], and perceived learning [6].

While online educational media and flipped learning paradigms hold great promise toward increasing student learning, offering advantages over traditional education by allowing students to personalize their learning, research also suggest that students may rate online education significantly lower than standard classroom settings in measures of content, interaction, participation, faculty preparation, and communication [39]. These findings suggest that four categories of limitations that must be overcome for online educational media to reach its

full potential. The first category relates to technical challenges implicit in any technology-based educational approach such as students with older computers or slow internet connections getting frustrated with low-quality or interrupted online lectures. The second category involves social attitudes toward online educational technology including difficulties overcoming initial student skepticism toward online education [6, 43]. The third category represents usability concerns such as ergonomic problems that could arise from students spending long hours watching a computer screen [43]. The final—and perhaps the most important—limitation surrounds the interaction; viewing online lectures is inherently asynchronous, static, and unidirectional. Such an experience is severely limited compared with dynamic classroom settings in which students can interrupt, ask questions, and gain feedback, and teachers can monitor how well students pay attention to class material.

The lack of student monitoring capabilities in online education represents both a potential pitfall and a promising design opportunity for developers. In current online systems, teachers are unable to observe students during instruction and thus have no means to verify whether students paid attention to the lesson, understood the material, or even watched the lecture at all. One of the hallmarks of the current “Millennial” generation of students is the prevalence of multitasking [10]; students are accustomed to “watching TV, talking on the phone, doing homework, eating, and interacting with their parents all at the same time” [17]. Although flipped learning will likely appeal to the Millennial preferences for teamwork and technology use, multitasking students might simultaneously engage in other activities while watching or listening to online content [25]. Unlike traditional classroom settings that involve less division of attention [13, 31], such multitasking may negatively affect student learning and memory. Additional research is needed to better understand the potential drawbacks of multitasking on learning, particularly in online contexts. Further investigation is also needed to better understand how novel technology might support emerging educational paradigms such as flipped learning to overcome these challenges.

SYSTEM DESIGN

To investigate how technology might support flipped learning and ITS applications, we designed and implemented a novel system that passively monitors student attention to educational material in real time in order to suggest the optimal review topic. This research builds on advances in human factors and educational psychology and informs flipped learning, offline computer-based tutoring systems, and even traditional classroom instruction.

The system design, informed by teacher behaviors, involves three components: an instruction component, an attention monitoring component, and a supervisor component (Figure 3). The instruction component uses multimedia instructional content to teach students by guiding them through several modular subtopics within a lesson. The attention monitoring component, based on technology that measures electrical signals produced by neuronal activity, gauges student attention during various parts of the lecture. The supervisor component manages the interaction between the student and the lesson

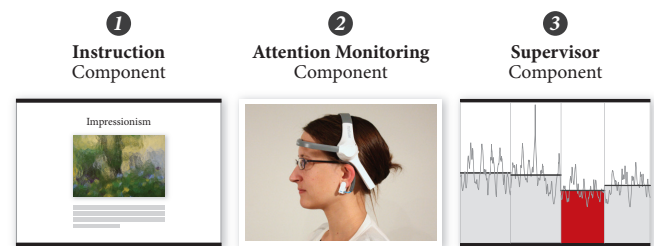


Figure 3. The instruction component (1), attention monitoring component (2), and supervisor component (3) work together to infer student attention during instruction and determine the optimal review content for students.

by adapting the review content based on measurements from the attention monitoring component. In our current proof-of-concept design, students only review a single lecture topic, which is selected by the supervisor on the basis of having the lowest average attention score amongst lecture modules.

Instruction

The instruction system draws on the idea of presenting educational lectures that are divided into coherent, semi-independent subtopics. Existing computer-based education systems have already been constructed around the assumption that lesson plans can be composed of tiered modules [29]. These modules might be adapted from textbooks, coded by teachers, identified by crowdsourcing using students themselves, or even parsed automatically using artificial intelligence techniques.

Our prototype instruction component, based on a four-module lesson plan, offers four minutes of instruction for each module, leading to a total instructional time of 16 minutes. The lecture length was informed by studies recommending that optimal instruction periods last 15–20 minutes [11].

The inherent modularity of the lesson design affords computer-based education the unique potential to adapt lessons to varying student needs, abilities, and goals. Current ITS systems commonly require students to work through problem substeps and continuously assess student achievement [46], which might be used to adapt content based on student performance. One potential means of adapting educational content is in the form of custom *content review*. Since each lesson module represents a general learning “goal” that the student must master, a post-hoc review of modules that lost student interest and attention might reinforce educational concepts that had been missed along the way.

Reviewing material has been described as an effective teaching strategy that can augment learning and break up lecture monotony [37]. However, review periods must be confined in scope due to limitations in the time allotted for education and because excessive review might be ineffective, bore the student, or cause negative perceptions regarding the interaction. Our design, based on the notion that students might benefit the most from review on subtopics to which they paid the least attention, attempts to optimize content review time for individual students. Consequently, the system required a means of monitoring how well students paid attention to each lesson module to determine which subtopic a student might benefit the most from reviewing.

Attention Monitoring

Several technologies are currently available for monitoring user attention and engagement with computer interfaces including gaze and eye-tracking systems, video recording devices, and a variety of sensors for measuring users' biological data.

Literature in human factors, particularly on adaptive automation, informed the development of the attention monitoring component of our design. Monitoring user attention is particularly important in adaptive automation, which considers the user as a system operator who supervises tasks such as air traffic control [24, 34], air navigation [33], and human-robot teams [27]. Researchers have developed many new technologies for monitoring operator cognitive workload and attention in these contexts, including heart-rate variability [5], galvanic skin response [4], and electroencephalography (EEG) [8, 36, 51]. From these potential technologies, EEG was chosen as the basis of the attention monitoring component due to the promise studies have shown of utilizing neural signals to identify subtle shifts in user alertness, attention, perception, and workload in laboratory, simulation, and real-world contexts [3, 8, 12, 32, 41, 44, 45, 48, 51]. ITS research has previously identified EEG as a potential signal from which to infer the difficulty of the instructional material [7, 30].

EEG Technology

EEG technology utilizes electrodes placed on the scalp to measure electrical activity created by positive and negative charges in the cerebral cortex following changes in membrane conductance as neurotransmitters are released. One of the advantages of using EEG is high temporal resolution, which offers the ability to correlate EEG data with stimuli in the external world. Unfortunately, EEG offers low spatial resolution, making it difficult to determine which part of the brain created the signals. Further, because EEG data represents a vast generalization of actual brain activity, small changes in user states can be difficult to perceive. Finally, in non-invasive, commercially available EEG headsets that are more appropriate for end-user applications such as education, electrode signals are highly susceptible to noise as well as to influence by extraneous signals such as electromyography (EMG)—electrical signals that originate from muscles in the scalp and face. Despite these disadvantages, EEG data, generally sequenced into frequency bands, has been shown to provide insight into cognitive states, including task engagement/attention, working modality, and perception of user/machine errors [12, 15, 36, 51], as well as user mood and emotions, such as anxiety, surprise, pleasure, and frustration [8, 14, 28].

By designing technology that can measure and react to such neural signals, researchers hope to create “bio-cybernetic loops” [36], in which users can either give direct mental commands to *active* systems or have a *passive* system respond and adapt itself to shifts in user states. Although research into active brain-computer interfaces (BCIs) has succeeded in various ways such as allowing users to control a mouse cursor or wheelchair, these systems often require extensive tuning and training on both the part of the system and the user, are slower than traditional input methods, and are rarely generalizable across multiple users [2, 35]. As a result, active BCIs are gener-

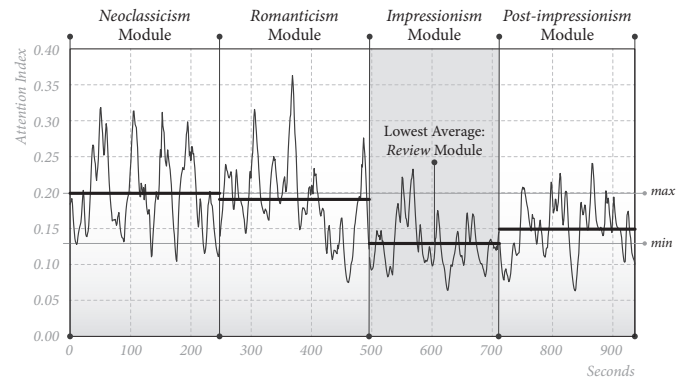


Figure 4. The supervisor system filters and averages EEG levels to produce attention scores for each subtopic within a lecture. User data shown here indicates that the *Impressionism* module should be reviewed.

ally only used by individuals with disabilities who are unable to use standard input methods and for whom the time and effort required for training have a significant payoff. Research into passive BCIs, in which brain activity is monitored to gain additional context into user activity and current state, appears more promising for typical users. Passive BCI systems have already been built that can aid in the detection of user and machine errors and support varying levels of user cognitive load [14, 51]. Further, our previous work has demonstrated the promise of utilizing passive BCI to support education through instructional agents that can re-engage users following drops in their attention [44]. Our current work extends research into BCIs by using EEG as a tool for monitoring student attention in online educational environments found in flipped classroom and MOOC scenarios.

Attention-Monitoring Hardware

In this research, it was paramount that all aspects of the system could feasibly be used in both actual classroom scenarios and self-directed, media-based learning settings. As a result, the attention monitoring component uses the low-cost Neurosky Mindwave headset as a means of measuring user EEG data. The wireless headset is relatively low-fidelity compared with alternative BCI devices; the Mindwave gathers data using a single electrode for signal input and an additional electrode for grounding. However, the Mindwave is significantly more practical for real-world scenarios, as both electrodes can remain dry ensuring that the headset can quickly and easily be put on and taken off, as opposed to other devices, which often require the placement of many more electrodes and the use of special conductive gels.

The Mindwave gathers EEG measurements from the FP1 region of the cortex which is known to manage learning, mental states, and concentration [16]. The hardware uses the A1 region for grounding and filtering the signal via common-mode rejection and additionally utilizes notch filters, analog and digital low- and high-pass filters, and proprietary algorithms to remove EMG artifacts and other noise. The effectiveness of this filtering was verified through pretests that assessed the presence of common artifacts such as eye-blinks. The device samples at a rate of 512 Hz and is sensitive to frequencies in the range of 3–100 Hz, which are broken into *alpha*, *beta*, *theta*, and *gamma* waves using Fast Fourier Transforms.

Supervisor

The supervisor software system uses the output from the Mind-wave EEG headset to monitor user attention during the presentation of educational material and select review content based on low attention scores. First, EEG levels for alpha, beta, and theta frequencies were sampled from the attention monitoring component at 512 Hz, averaged across a one-second window to produce 1 Hz signals, and combined to form an attention index using the formula $A = \beta/(\alpha + \theta)$ [12, 36, 44]. This attention index was filtered using median filtering to remove values likely produced by EMG artifacts and other noise:

$$A(t) = \tilde{V}_w$$

where $A(t)$ is the attention index at time t , calculated by taking the median of V , a vector of the previously measured indices across a given time window w (in this research a time window of 5 recordings, corresponding to measurements from the previous 5 seconds, was used based on pre-test results).

The index was further smoothed using an exponentially-weighted moving average (EWMA) to enable the software to pick out general attentional trends as opposed to moment-by-moment changes in attention or those produced by artifacts:

$$S(t) = \begin{cases} A(t) & : t = 1 \\ c * A(t-1) + (1-c) * S(t-1) & : t > 1 \end{cases}$$

where S corresponds to the smoothed attention value produced using an EWMA, t is time, A is the median-filtered index, and c is a regularization constant that is inversely proportional to the relative weighting of less recent events (a value of .2 was used in this study based on pre-test results).

Using the smoothed attention index, the system is able to infer student attention levels during the presentation of educational material, provided that the presentation contains sections delineated *a priori*. To determine the student attention level for a given module of educational content, the supervisor calculates the mean of the attention indices recorded during the presentation of that section. This information can be passed to computer systems or even human instructors to help them gauge the effectiveness of lessons. In our proof-of-concept design, the supervisor component uses attention information to select review topic by choosing the lesson module with the lowest average attention values (Figure 4).

HYPOTHESES

Our design was based on the premise that reviewing content to which students paid the least attention will improve student learning. The hypotheses below seek to capture the relationships between various methods of reviewing content and student learning.

Hypothesis 1. In a given learning task, a review focused on topics that had low EEG-monitored attention levels will increase learning performance compared with a no review baseline, while a review focused on topics that had high EEG-monitored attention levels will not increase learning performance over this baseline.

Hypothesis 2. In a given learning task, a review focused on topics that had low EEG-monitored attention levels will match learning performance gains achieved by a full review of lesson topics in less time; a review focused on topics that had high EEG-monitored attention levels will not produce an increase in learning performance equivalent to a full review of all topics.

EVALUATION

To investigate the effects of *adaptive content review* on educational outcomes, we implemented three alternative education systems designed around various methods of providing content review. The first design provides *no content review*, the second provides *maladaptive content review* based on reviewing concepts to which students initially paid the most attention, and the third provided a *full review* of all lesson concepts. These baselines were developed as parallels to current MOOC system capabilities; after viewing an online lecture, students can choose to review nothing, watch the entire presentation again, or choose to view individual parts of the lecture. A laboratory experiment evaluated the effectiveness of the adaptive design compared with these three alternatives.

Experimental Design

To test our hypotheses, we designed and conducted a gender-balanced 4×2 between-participants study, which manipulated the content review of a computer-based educational system that instructed participants regarding four topics within a given lecture. Independent variables included the *type of review* received and *participant gender*. Dependent variables included participants' cognitive learning performance measured by their recall of the lesson content, their perceptions of the educational software, and their self-reported learning.

In the experiment, the instruction component taught art history, chosen as a lesson topic that participants were unlikely to have strong prior familiarity with, while still representing a real-world learning task. Four modules comprised the lecture: neoclassicism, romanticism, impressionism, and post-impressionism. In all conditions, the computer-based educational system presented each topic for four minutes, leading to a total instructional time of 16 minutes. Each four minute module consisted of a two minute segment giving an overview of the topic itself, followed by two one-minute segments highlighting a single painting from that art period. The system used text, images, and a prerecorded female voice to provide participants with lecture content.

The experiment manipulated how a computer-based education system presented a review period following the initial instructional period. In order to alter review periods across participants, we created four one-minute review segments, each of which corresponded to one of the four lesson modules. No new information was presented during the review segments. Instead, each review segment gave a brief overview of the important points initially taught during the corresponding module and included both paintings presented for that art period.

Using these review segments, the independent review variable included four levels: (1) *no review*, (2) *maladaptive review*, which presented the review segment on the topic that had

the highest average EEG attention levels, (3) *adaptive review*, which presented the review segment on the topic that had the lowest average EEG attention levels, and (4) *full review*, which presented all four review segments. In order to control for time differences between the no review, maladaptive review, and adaptive review conditions, participants in the no review condition were instructed to listen to one minute of classical music instead of spending one minute on review. Time differences were not controlled for participants in the full review condition, thus participants who received the full review interacted with the system for three additional minutes. Student prior knowledge of the lecture material was controlled for by means of an initial quiz prior to receiving instruction.

Experimental Procedure

The experimental protocol consisted of eight main phases: (1) introduction, (2) initial quiz, (3) instruction, (4) distractor, (5) review, (6) distractor, (7) evaluation, and (8) survey. In the first phase, following informed consent, the researcher gave the participant a brief description of the experiment and brought the participant into a sound-controlled room. Here, the researcher familiarized the participant with the Maeda Path Game⁵, which was used later on in the experiment as a distractor task. Each participant was given the chance to complete one level of the game and to ask any questions regarding playing the game. Once the participant indicated they understood how to play the game, the researcher aided the participant in putting on the wireless EEG headset and ensured good connectivity. The researcher then left the room and the participant started interacting with the computer-based educational software, which guided the participant through the next seven phases.

First, the software welcomed the user and gave an overview of the lesson plan. Next, the participant was instructed to take a ten-question, multiple-choice quiz with randomized question order to assess their prior knowledge of art history during the four art periods covered by the lesson plan. The instruction phase followed the initial quiz, represented by a lecture consisting of four minutes of instruction for each art period.

After the instruction phase, participants played the Maeda Path Game for two minutes as a distractor task. A game task was chosen due to the popularity of such online games with Millennial students [38], and to simulate how, in most educational settings, review does not immediately follow the initial presentation of educational material.

A brief review period followed the first distractor phase. During this time, in the no review condition, participants simply listened to one minute of “Adagio un Poco Mosso,” a piece by Beethoven whom participants were told was a contemporary of many of the painters they had learned about. In the other three conditions, participants received review according to their condition, as described in the experimental design. Following the review period, participants played the Maeda Path Game for an additional two minutes as a further distractor



Figure 5. A participant in our experiment interacts with the educational technology, which monitors user attention to an art history lecture.

task to separate the review period from the evaluation which followed.

In the evaluation, participants took a 25-question, multiple-choice quiz, which tested their ability to recall the information presented to them during the instruction and review phases. The quiz consisted of five questions for each art period and five generalized questions that required knowledge of at least two art periods to answer correctly. As an example, the lecture instructed students that Neoclassicism began during the Age of Enlightenment, and one of the evaluation questions was: “In which of the following periods did Neoclassicism begin?” Question order was randomized for each participant, and no question appeared in both initial and final quizzes. Following the recall evaluation, each participant took a post-experiment questionnaire to obtain subjective evaluations of participant experiences, as well as demographic information. After questionnaire, the researcher re-entered the room, asked the participant to remove the headset, debriefed the participant, and compensated the participant \$5 for their time. The entire procedure took approximately 30 minutes.

Participants

A total of 48 participants (24 males and 24 females) took part in this experiment. Each of the four conditions was gender balanced (six males and six females). All participants were native English speakers recruited from the University of Wisconsin–Madison campus. Average participant age was 24.25 ($SD = 8.84$) with a range of 18–60. On a seven-point item, participants reported a moderate prior familiarity with computer-based or online education ($M = 4.04$, $SD = 1.75$), indicating the increasing prevalence of such technology. Figure 5 shows a participant interacting with our software.

Measurement and Analysis

Objective and subjective measurements captured the outcomes of the manipulations described above. Objective measurements included *information recall*, the percentage of correct answers on the post-experiment quiz, and *learning*, the normalized percentage difference in scores between the number of correct answers on the initial and final quizzes. Participant

⁵ <http://www.levitated.net/daily/levMaedaPath.html>

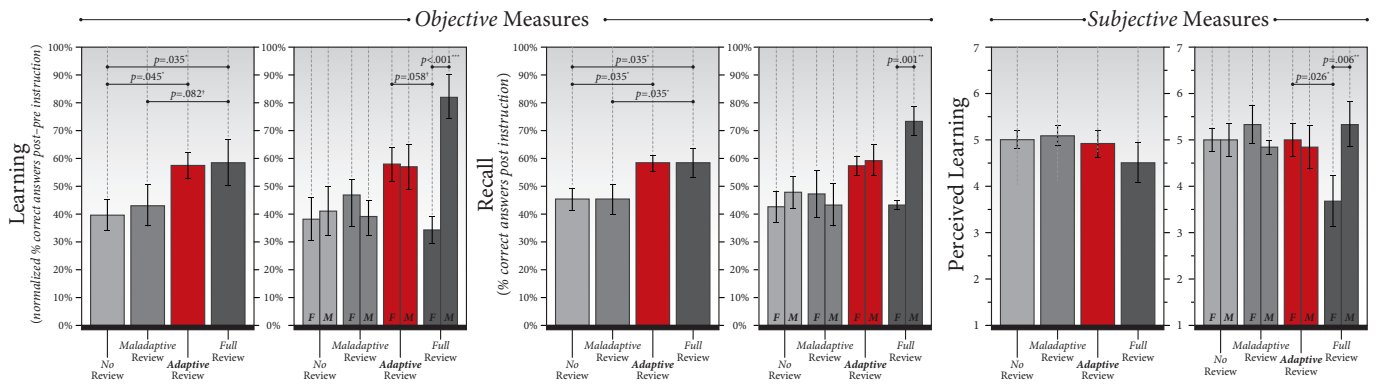


Figure 6. Results illustrate that adaptive review had a positive effect on learning and recall regardless of gender, while a maladaptive review system produced no effect. A full review improved learning and recall results overall, but was not effective across genders, leading to low learning, recall, and perceived learning scores for females. Only marginal and significant p values are reported. (†), (*), (**), and (***) denote $p < .10$, $p < .05$, $p < .01$, and $p < .001$, respectively.

EEG measurements were analyzed to complete a preliminary evaluation regarding the assumptions underlying our design. Subjective measurements used seven-point rating items to measure participant responses on a post-experiment questionnaire, which asked participants to rate the perceived competency of instruction, the usefulness of the review, their perceived learning, and whether participants would like to use the system again in the future.

An analysis of the data from objective and subjective measurements evaluated the hypotheses using a two-way analysis of variance (ANOVA), with the main experimental manipulation and participant gender as two fixed effects. Scheffé’s method was used for contrast tests. To control for Type II errors for establishing a “no difference” in comparisons, we followed guidelines suggested by [18], in particular an alpha level of 0.25 (i.e., $p > .25$).

Results⁶

Our experiment was based on the concept that different forms of content review would effect learning, thus we first confirmed the main effect of review content on student cognitive performance. An analysis of objective data found a significant main effect of type of review on information recall, $F(3,40) = 3.18$, $p = .034$, $\eta_p^2 = .193$, and a marginal main effect of type of review on learning, $F(3,40) = 2.53$, $p = .071$, $\eta_p^2 = .160$. Information recall scores were on average 45.33% ($SD = 14.30\%$), 45.33% ($SD = 19.25\%$), 58.33% ($SD = 11.11\%$), and 58.33% ($SD = 18.25\%$) in the no review, maladaptive review, adaptive review, and full review conditions, respectively. Average learning scores between conditions were 39.71% ($SD = 19.77\%$), 43.14% ($SD = 26.23\%$), 57.35% ($SD = 16.85\%$), and 58.33% ($SD = 29.51\%$) in the no review, maladaptive review, adaptive review, and full review conditions, respectively. We additionally analyzed the data for effects of gender and found a marginal effect of gender on both information recall, $F(1,40) = 3.92$, $p = .055$, $\eta_p^2 = .089$, and learning, $F(1,40) = 3.06$, $p = .088$, $\eta_p^2 = .071$. Males performed marginally better in both measures, averaging 56.00% ($SD = 18.61\%$) in information recall

and 54.90% ($SD = 26.52\%$) in learning, while females averaged 47.67% ($SD = 14.10\%$) in information recall and 44.36% ($SD = 21.24\%$) in learning. The results demonstrated a significant interaction between type of review and gender on information recall, $F(3,40) = 3.16$, $p = .035$, $\eta_p^2 = .191$, and on learning, $F(3,40) = 4.44$, $p = .009$, $\eta_p^2 = .266$. Prior familiarity with the material used in the experiment was low; the average pre-instruction quiz score was 1.52 ($SD = 1.53$). No significant differences in pre-instruction quiz scores were found across conditions or genders.

We utilized contrast tests to confirm our hypotheses. Hypothesis 1 predicted that participants who received adaptive content review would demonstrate increased cognitive learning performance over participants who received no review, while maladaptive review would not produce any learning gains over the no review baseline. The results confirmed both aspects of this hypothesis. Contrast tests revealed a significant difference in information recall between participants who received adaptive review and those that received no review, $F(1,40) = 4.77$, $p = .035$, $\eta_p^2 = .107$, with participants in the adaptive condition outperforming those with no review by 29%, regardless of gender. Participants who received adaptive reviewed also significantly outperformed those who received no review in learning, $F(1,40) = 4.29$, $p = .045$, $\eta_p^2 = .097$, regardless of gender. Participants in the no review and maladaptive conditions did not perform significantly differently in terms of information recall, $F(1,40) = 0.00$, $p = 1.00$, $\eta_p^2 = 0.00$, or learning, $F(1,40) = 0.16$, $p = .689$, $\eta_p^2 = .004$.

Hypothesis 2 predicted that participants who received an adaptive review of topics would achieve equivalent learning results to those who received a full review of all topics. Additionally, Hypothesis 2 predicted that participants who received a full review would outperform those who received no review and those who received a maladaptive review. The analysis partially supported this hypothesis. No significant difference was found between participants in adaptive and full review conditions in measures of information recall, $F(1,40) = 0.00$, $p = 1.00$, $\eta_p^2 = 0.00$, or learning, $F(1,40) = 0.01$, $p = .909$, $\eta_p^2 = 0.00$. These results demonstrate that students who received adaptive review were able to achieve equivalent perfor-

⁶ Figure 6 highlights the major results of the study. Only marginal and significant effects are reported.

mance results as those who received a review of all content, even though they spent 75% less time on review. As predicted by Hypothesis 2, the analysis revealed significant differences between full and maladaptive review in information recall $F(1,40) = 4.77, p = .035, \eta_p^2 = .107$. However, the results showed only a marginal difference between full and maladaptive review in learning, $F(1,40) = 3.18, p = .082, \eta_p^2 = .074$, providing partial support for the hypothesis.

Contrast tests across types of review and gender found that, while full review helped males, it led to the lowest scores of information recall and learning across all conditions for females. Males significantly outperformed females in the full review condition in information recall, $F(1,40) = 12.70, p = .001, \eta_p^2 = .241$, and in learning, $F(1,40) = 15.88, p < .001, \eta_p^2 = .284$. Female learning in the adaptive condition marginally surpassed that in the full condition, $F(1,40) = 3.81, p = .058, \eta_p^2 = .087$, although no significant difference was found between these conditions in information recall, $F(1,40) = 2.77, p = .104, \eta_p^2 = .065$. A full review of lesson material did not lead to significant differences for females compared with no review in information recall, $F(1,40) = .006, p = .937, \eta_p^2 = 0.00$, or learning, $F(1,40) = 0.11, p = .747, \eta_p^2 = .003$, nor were any differences found for females between full and maladaptive review conditions in information recall, $F(1,40) = 0.23, p = .637, \eta_p^2 = .006$, or learning, $F(1,40) = 1.12, p = .297, \eta_p^2 = .027$.

A preliminary analysis evaluated the underlying assumption that an attention index could provide a good predictor of which lecture sections students might benefit from reviewing. Regression analysis, using average EEG attention values as a predictor variable for recall scores, assessed this premise. This analysis only included EEG values for participants in the no review condition because the previous analysis revealed that the additional review content received in the other three conditions might significantly alter student recall abilities. In constructing the model, the analysis normalized the EEG predictor variable and applied a *log* transformation to both predictor and response variables to obtain linearity. For students in the no review condition, EEG-monitored attention levels accounted for 25.19% of the variance in student recall abilities and served as a marginal predictor of recall scores, $\beta = .187, t(10) = 2.17, p = .055$, providing partial support for our premise.

Overall, we found no significant main effects of type of review and gender on students' perceived learning, their perceptions regarding the competence of the instruction, whether they found the review to be useful, and their willingness to use the system again. However, we found a marginal interaction effect between type of review and gender on perceived learning, $F(3,40) = 2.83, p = .051, \eta_p^2 = .141$. This effect mirrored the gender differences in objective cognitive performance surrounding the full review condition, with males reporting significantly higher learning than females in the full review condition, $F(1,40) = 8.40, p = .006, \eta_p^2 = .140$, and females in the full review condition reporting significantly less learning than those in the adaptive review condition, $F(1,40) = 5.39, p = .026, \eta_p^2 = .094$. Moreover, females reported learning

significantly less after receiving full review compared with maladaptive review, $F(1,40) = 8.40, p = .006, \eta_p^2 = .140$, and when comparing full review with no review, $F(1,40) = 5.38, p = .026, \eta_p^2 = .094$, even though there were no significant objective differences in learning performance for females across these conditions.

DISCUSSION

Our results confirmed Hypothesis 1 as adaptive review significantly increased recall and learning gains compared with the no review baseline, while students who received maladaptive review did not achieve cognitive performance gains over the baseline in these measures. Hypothesis 2 was partially confirmed; our results found no difference in learning or recall between the full and adaptive conditions, while a significant difference was found between full review and maladaptive review in terms of recall. However, our results reported only a marginally significant difference between full review and maladaptive review in terms of learning.

The results demonstrate that simply reviewing material is not enough to improve student learning. We found no difference between the no review and maladaptive review conditions, likely because participants in the maladaptive condition received a review of material to which they had initially paid attention, and thus did not require review on. Reviewing unnecessary material might have detrimental effects on student learning, leading to boredom or frustration, which could be one factor behind females having the least recall in the full review condition. This objectively poor performance by females after full review, noted by females in their subjective, perceived learning reports, highlights the notion that reviewing everything is also a non-optimal strategy. The experimental results support the hypothesis that adaptively reviewing content optimizes the time spent on review as adaptive content review led to high learning results in both genders.

Limitations & Future Work

While the technology achieved its intended goal of improving student learning by monitoring user attention in real time and optimizing review content based on lapses in attention, open questions remain regarding the assumptions underlying the design, attention monitoring system, and current review abilities.

The design is based on the underlying premise that the optimal review topic for a student will be the topic to which they paid the least attention. While the experimental results provide initial support for this premise, it may not always hold. For instance, one factor behind a student losing interest in a lecture topic might be that the lesson is covering material with which the student is already familiar. In such conditions, reviewing material based on low attention values may not always be optimal. Future work might attempt to confirm that students lost attention not because of prior existing knowledge by combining our method with other embedded assessment techniques.

Due to technological limitations inherent in EEG, the attention monitoring component is prone to be affected by other signals such as muscle artifacts. Although the monitoring system

employed filters to remove such artifacts and the supervisor used smoothing and averaging techniques to create long term trends, which should alleviate problems due to short-term signal variations caused by EMG, more investigation is necessary to validate that the attention index indeed represents underlying cognitive activity free from extraneous signals. Our preliminary analysis reveals that EEG attention levels were a marginal predictor for student recall abilities, however a means of fully validating EEG levels for use in such contexts has yet to be determined. The capabilities of our technology might be increased by exploring other means of determining review content instead of simply averaging attention across pre-defined lesson subsections. Future technology might benefit from more robust signal analysis methods including examining EEG slopes, regression lines, or height and frequency of local maxima and minima.

The current design requires lessons to be divided into subtopics, requires a pre-recorded content review section for each subtopic, and only selects a single topic for review. Although many lessons and textbooks do follow a modular paradigm, adapting complex and creative lessons to this format may not always be easy. Further, in certain circumstances students might benefit from a review of multiple topics, or even from no review at all if sufficient attention was initially paid to the lecture. These limitations reflect the proof-of-concept system status and might be addressed by future research in several ways. First, advances in intelligent summarization technology could work toward creating automatic sub-topic review sections based on pre-existing educational content. Further, robust supervisor systems might benefit from baseline attention measurements, allowing the system to select multiple review topics whose attention values fall below baseline levels. Finally, future systems may not necessarily require pre-demarcated subtopics. Instead, systems may identify individual content items across the entire learning process and merge them into a coherent review, creating a truly customized educational experience.

Our experiment represented a limited interaction consisting of a single lecture based on a single method of gauging student attention. To more fully explore the potential benefits in utilizing EEG-based review technology, future studies might examine long-term learning effects or incorporate additional measurements such as gaze or traditional embedded assessment to corroborate the EEG signal. Additionally, future studies might investigate differences in learning across subtopics for students receiving adaptive EEG-based review. Although we found no significant differences across topics within the adaptive condition, we note that this analysis is limited due to our sample size and the lack of *a priori* knowledge regarding which subtopic students in the adaptive condition would review.

Our design represents a prototype that highlights one potential means for technology to support novel educational contexts. One final avenue for future research might include designing tools that can visualize student attention data for the use of human educators, rather than computer systems. Such technology would be of immense value for teachers who could use it to evaluate and improve their own lessons as well as gain

insights into individual student and class needs. Additional visualization methods for large-scale data will aid in developing these tools.

Research and Design Implications

This work has important theoretical, methodological, and practical design implications for education researchers and developers who wish to explore the use of technology to support novel educational paradigms such as flipped learning. First, the results, in which adaptive content review achieved significant learning performance while maladaptive review proved no better than having no review at all, demonstrate an apparent link between levels of attention and test performance. Additional analysis, which revealed a marginal association between EEG attention values and information recall, further strengthens this theory. Second, the design methodology outlines an effective strategy for developing educational tools modeled after human educator behaviors, as well as demonstrating the potential of using novel technologies such as EEG to facilitate learning in self-directed education. Finally, this research might inform future explorations into self-directed learning technologies by providing a model for a practical system to improve student learning that could be deployed in real-world contexts.

CONCLUSION

The rise of online media and computer-based education is fundamentally altering education, creating a novel design space for human-computer interaction researchers. This research sought to explore how novel technologies might support learning in the new design space offered by flipped classroom and MOOC settings. Our work presents the design and implementation of a novel system that utilizes real-time attention monitoring to adaptively determine optimal content review for students. This design significantly improved learning outcomes, demonstrating the potential for adaptive educational software in creating effective user interactions.

ACKNOWLEDGMENTS

The University of Wisconsin–Madison Graduate School provided support for this research. We would like to thank Allie Terrell and Brandi Hefty for their help in our research.

REFERENCES

1. Aleven, V., and Koedinger, K. R. An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science* 26, 2 (2002), 147–179.
2. Ayaz, H., Shewokis, P., Bunce, S., Schultheis, M., and Onaral, B. Assessment of cognitive neural correlates for a functional near infrared-based brain computer interface system. *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (2009), 699–708.
3. Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., and Craven, P. L. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine* 78, Supplement 1 (2007), B231–B244.
4. Boucsein, W., Haarmann, A., and Schaefer, F. Combining skin conductance and heart rate variability for adaptive automation during simulated ifr flight. In *Engineering Psychology and Cognitive Ergonomics*, vol. 4562. 2007, 639–647.

5. Byrne, E. A., and Parasuraman, R. Psychophysiology and adaptive automation. *Biological Psychology* 42, 3 (1996), 249–268.
6. Cannod, G. C., Burge, J. E., and Helmick, M. T. Using the inverted classroom to teach software engineering. *Computer Science and Systems Analysis Technical Reports* (2007).
7. Cirett Galán, F., and Beal, C. Eeg estimates of engagement and cognitive workload predict math problem solving outcomes. *UMAP* (2012), 51–62.
8. Cutrell, E. Tan, D. BCI for passive input in HCI. In *Proc CHI'07* (2007).
9. D'Mello, S., Graesser, A., and Picard, R. Toward an affect-sensitive autotutor. *IEEE Intelligent Systems* 22, 4 (2007), 53–61.
10. Frand, J. L. The information-age mindset: Changes in students and implications for higher education. *EDUCAUSE Review* 35, 5 (2000), 15–24.
11. Frederick, P. J. The lively lecture: 8 variations. *College Teaching* 34, 2 (1986), 43–50.
12. Freeman, F. G., Mikulka, P. J., Prinzel, L. J., and Scerbo, M. W. Evaluation of an adaptive automation system using three eeg indices with a visual tracking task. *Biological Psychology* 50, 1 (1999), 61–76.
13. French, D. P. ipods: Informative or invasive?. *J College Science Teaching* 36, 1 (2006), 58–59.
14. George, L., and Lecuyer, A. An overview of research on 'passive' brain-computer interfaces for implicit human-computer interaction. In *Proc ICABB'10 Workshop on Brain-Computer Interfacing and Virtual Reality* (2010).
15. Gevins, A., and Smith, M. E. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science* 4, 1 (2003), 113–131.
16. Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., and Rush, G. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors* 40 (1998), 79–91.
17. Jonas-Dwyer, D., and Pospisil, R. The millennial effect: Implications for academic development. *Transforming Knowledge into Wisdom: Holistic Approaches to Teaching and Learning* (2004), 195–207.
18. Julnes, G., and Mohr, L. Analysis of no-difference findings in evaluation research. *Evaluation Review* 13, 6 (1989), 628–655.
19. Kemp, J. W., Livingstone, D., and Bloomfield, P. R. Slooodle: Connecting vlc tools with emergent teaching practice in second life. *British J Educational Technology* 40, 3 (2009), 551–555.
20. KNEWTON. The flipped classroom infographic, March 2012. <http://www.knewton.com/flipped-classroom/>.
21. Kulik, C.-L. C., and Kulik, J. A. Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior* 7, 1–2 (1991), 75–94.
22. Lage, M. J., and Platt, G. The internet and the inverted classroom. *J Economic Education* 31, 1 (2000), 11–11.
23. Lage, M. J., Platt, G. J., and Treglia, M. Inverting the classroom: A gateway to creating an inclusive learning environment. *J Economic Education* 31, 1 (2000), 30–43.
24. Langan-Fox, J., Cauty, J. M., and Sankey, M. J. Human-automation teams and adaptable control for future air traffic management. *Intl J Industrial Ergonomics* 39, 5 (2009), 894–903.
25. Long, S. R., and Edwards, P. B. Podcasting: making waves in millennial education. *J Nurses in Staff Devel* 26, 3 (2010), 96.
26. Meccawy, M., Blanchfield, P., Ashman, H., Brailsford, T., and Moore, A. Whurle 2.0: Adaptive learning meets web 2.0. In *Times of Convergence. Technologies Across Learning Contexts*, vol. 5192. 2008, 274–279.
27. Miller, C., and Parasuraman, R. Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors* 49, 1 (2007), 57–75.
28. Molina, G., Tsoneva, T., and Nijholt, A. Emotional brain-computer interfaces. In *Proc ACII'09* (2009), 1–9.
29. Moore, A., Brailsford, T. J., and Stewart, C. D. Personally tailored teaching in whurle using conditional transclusion. In *Proc ACM Hypertext and Hypermedia* (2001), 163–164.
30. Mostow, J., Chang, K., and Nelson, J. Toward exploiting eeg input in a reading tutor. In *Proc AIED'11* (2011), 230–237.
31. Naveh-Benjamin, M., Kilb, A., and Fisher, T. Concurrent task effects on memory encoding and retrieval: Further support for an asymmetry. *Memory & Cognition* 34 (2006), 90–101.
32. Nijholt, A., Bos, D. P.-O., and Reuderink, B. Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Computing* 1, 2 (2009), 85–94.
33. Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., and Barnes, M. Theory and design of adaptive automation in aviation systems. *Technical Report NAWCADWAR-9023-60* (1992).
34. Parasuraman, R., and Wickens, C. D. Humans: still vital after all these years of automation. *Human Factors* 50, 3 (2008), 511–520.
35. Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W., Ramoser, H., Schlogl, A., Obermaier, B., and Pgegenzer, M. Current trends in Graz Brain-Computer Interface (BCI) research. *IEEE Trans Rehabilitation Engineering* 8, 2 (2000), 216–219.
36. Pope, A. T., Bogart, E. H., and Bartolome, D. S. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology* 40, 1-2 (1995), 187–195.
37. Potter, A. M. Statistics for sociologists: Teaching techniques that work. *Teaching Sociology* 23, 3 (1995), 259–263.
38. Prensky, M. Digital natives, digital immigrants part 2: Do they really think differently? *On the Horizon* 9 (2001), 1–6.
39. Ryan, M., Carlton, K., and Ali, N. Evaluation of traditional classroom teaching methods versus course delivery via the world wide web. *J Nursing Education* 38, 6 (1999), 272–277.
40. Schofield, J. W. *Computers and Classroom Culture*. Cambridge University Press, 1995.
41. Solovey, E., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., and Jacob, R. Brainput: enhancing interactive systems with streaming fnirs brain input. In *Proc CHI'12* (2012), 2193–2202.
42. Strayer, J. How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research* (2009), 1–23.
43. Strayer, J. F. *The effects of the classroom flip on the learning environment: a comparison of learning activity in a traditional classroom and a flip classroom that used an intelligent tutoring system*. PhD thesis, Ohio State University, 2007.
44. Szafir, D., and Mutlu, B. Pay attention!: designing adaptive agents that monitor and improve user engagement. In *Proc CHI'12* (2012), 11–20.
45. Tan, D., and Nijholt, A. *Brain-Computer Interfaces: applying our minds to human-computer interaction*. Springer, 2010.
46. VanLehn, K. Intelligent tutoring systems for continuous, embedded assessment. *The future of assessment: Shaping teaching and learning* (2008), 113–138.
47. VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist* 46, 4 (2011), 197–221.
48. Vi, C., and Subramanian, S. Detecting error-related negativity for interaction design. In *Proc CHI'12* (2012), 493–502.
49. Woolf, B., Bursleson, W., Arroyo, I., Dragon, T., Cooper, D., and Picard, R. Affect-aware tutors: recognising and responding to student affect. *Intl J Learning Technology* 4, 3 (2009), 129–164.
50. Yilmaz-Tuzun, O. Preservice elementary teachers beliefs about science teaching. *J Science Teacher Education* 19 (2008), 183–204.
51. Zander, T. O., Kothe, C., Jatzev, S., and Gaertner, M. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In *Brain-Comp Int.* 2010, 181–199.