

# **A Real-Time Model to Assess Student Engagement during Interaction with Intelligent Educational Agents**

## **Abstract**

Adaptive learning is an educational method that utilizes computers as an interactive teaching device. Intelligent tutoring systems, or educational agents, use adaptive learning techniques to adapt to each student's needs and learning styles in order to individualize learning. Effective educational agents should accomplish two essential goals during the learning process – 1) monitor engagement of the student during the interaction and 2) apply behavioral strategies to maintain the student's attention when engagement decreases. In this paper, we focus on the first objective of monitoring student engagement. Most educational agents do not monitor engagement explicitly, but rather assume engagement and adapt their interaction based on the student's responses to questions and tasks. A few advanced methods have begun to incorporate models of engagement through vision-based algorithms that assess behavioral cues such as eye gaze, head pose, gestures, and facial expressions. Unfortunately, these methods require a heavy computation load, memory/storage constraints, and high power consumption. In addition, these behavioral cues do not correlate well with achievement of high-cognitive tasks, as we will discuss in this paper. As an alternative, our proposed model of engagement uses physical events, such as keyboard and mouse events. This approach requires fewer resources and lower power consumption, which is also ideally suited for mobile educational agents such as handheld tablets and robotic platforms.

In this paper, we discuss our engagement model which uses techniques that determine behavioral user state and correlate these findings to mouse and keyboard events. In particular, we observe three event processes: total time required to answer a question; accuracy of responses; and proper function executions. We evaluate the correctness of our model based on an investigation involving a middle-school after-school program in which a 15-question math exam that varies in cognitive difficulty is used for assessment. Eye gaze and head pose techniques are referenced for the baseline metric of engagement. We then conclude the investigation with a survey to gather the subject's perspective of their mental state throughout the exam.

We found that our model of engagement is comparable to the eye gaze and head pose techniques. When high-level cognitive thinking is required, our model is more accurate than the eye gaze and head pose techniques due to the use of outside variables for assistance and non-focused gazes during questions requiring deep thought. The large time delay associated with the lack of eye contact between the student and the computer screen causes the aforementioned algorithms to incorrectly declare the subjects as being disengaged. Furthermore, speed and validity of responses can help to determine how well the student understands the material, and this is confirmed through the survey responses and video observations. This additional information will be used in the future to better integrate instructional scaffolding and adaptation with the educational agent.

## **I. Introduction**

The purpose of this paper is to discuss a reliable, noninvasive method of monitoring academic

engagement within the domain of computer-based education (CBE). In successful classroom settings, teachers are able to observe the student's engagement in real-time and employ strategies to reengage the student, which, in effect, improves attention, involvement and motivation to learn<sup>1</sup>. This is also true during one-on-one tutoring sessions due to the fact that tutors are able to track engagement in real-time as well. In general, teachers are able to determine engagement by following behavioral cues from students such as direction of attention, posture, facial expressions, and responsiveness to instructional activity<sup>2</sup>. This behavioral engagement is a crucial component in education because it is often related to the academic achievement of a student<sup>3,4</sup>.

Currently, educational software is a widely used method of instruction inside the classroom and at home. Research has shown that CBE actually improves academic achievement<sup>5</sup> and student motivation<sup>6</sup> when compared to traditional classroom instruction. Using CBE reduces the amount of instructional time required and increases the student's attitude towards learning<sup>7</sup>. Although research has shown CBE as being a highly effective learning tool, it pales in comparison to a human tutor<sup>5</sup>. Therefore, CBE should be used as a supplement to traditional instruction and not as a replacement<sup>8</sup>. In this investigation, we will determine how a computer-based system can monitor student engagement in a manner comparable to that of real classroom teachers.

## II. Related Work

CBE only focuses on comprehension of material<sup>9</sup> and not real-time engagement, which is essential for optimal academic achievement. Comprehension of material is determined solely by the validity of answer selections. Many standardized tests today, such as the SAT and GRE, adapt to the students based exclusively on their responses. This type of evaluation is known as computerized adaptive testing (CAT)<sup>10</sup>. If the student answers a question correctly, he/she is given a more difficult problem. If the student answers a question incorrectly, he/she is given a problem of less difficulty. However, for an educational system to be optimum, it must ensure that the student is actively and continuously engaged. Computer-based tools only focus on comprehension because of the difficulty associated with determining cognitive states. Due to the variability of behavior, characteristics, and environment, computational methods with the capability of identifying the behavioral cues associated with engagement have yet to be developed<sup>1</sup>.

As an alternative to measuring engagement in real time, scales have been created to evaluate motivation once the student has completed a system<sup>11</sup>. The problem with this method is that an educational agent will not be able to adapt to the educational needs of the student once the learning session has already been completed. The art of adaptation requires real-time information processing, which scales are unable to deliver.

A more promising alternative to measuring engagement is through electroencephalography (EEG) signal measurements. EEG signals are able to identify subtle shifts in alertness, attention, and workload in real time<sup>12</sup>. Szafir and Mutlu used an EEG headset to monitor engagement in an educational setting through storytelling<sup>1</sup>. When the EEG signals would begin to drop during narration, adaptive behavioral cues (verbal and non-verbal) would be used to re-engage the students. EEG measurements have the advantage of being non-invasive, well studied, and low

cost<sup>1</sup>; however, wearing a headset creates a controlled testing setup, which does not convey a natural learning environment. This ultimately has the potential to cause unnecessary distractions and distort results.

In efforts to create a non-invasive tool to monitor engagement in real time and within a natural learning environment, a viable option would be to use eye gaze and head pose to determine behavioral user state. Asteriadis et al. was able to develop a system using head pose and movement, direction of gaze, as well as measurements of hand gesture expressivity to determine six user states in an e-learning environment: attentive, full of interest, frustrated/struggling to read, distracted, tired/sleepy, and not paying attention<sup>13</sup>. The developed system was able to effectively detect reading- and attention-related user states very well when subjects were asked to read/watch an electronic document (web page, multimedia presentation, video clip). However, this system was not tested in a complex problem solving or test-taking environment.

### III. Design

In this paper, we discuss a novel model of student engagement based solely on mouse and keyboard events that leverages previous eye gaze and head pose research. Events are composed of mouse left/right clicks and keystrokes. Three event processes are observed to identify a common pattern associated with both an engaged and disengaged student: total time, response accuracy, and proper event execution. This data is collected as the students take a 15-question math test.

#### A. Total Time

The difficulty of a problem is determined by how much time is needed to submit a well thought out answer. Difficulty is directly proportional to the amount of time needed to respond. However, this exact allotment of time per question is unknown due to the subjectivity of classifying the difficulty level of problems. Therefore, in this investigation we used a distribution of the time taken by the entire group of students to determine the ideal time needed to adequately answer a particular problem. The ideal response time falls within the interquartile range (IQR) of the data. If the student answers within the lower quartile, his/her response is classified as fast. If the student answers within the upper quartile, his/her response is classified as slow. If the student answers within the IQR, his/her response is classified as average.

Next, we calculate the total amount of time needed to complete the entire 15-question test. This ideal test time is determined by a distribution of the time taken by the entire group of students to adequately complete the test. The ideal test time will fall within the IQR of the data. If the student answers within the lower quartile, his/her test time is classified as fast. If the student answers within the upper quartile, his/her test time is classified as slow. If the student answers within the IQR, his/her test time is classified as average.

#### B. Response Accuracy

Response accuracy is defined as the correctness of the submitted answer. As mentioned previously, this technique is currently widely used with standardized CAT. If a student answers a

question incorrectly, his/her accuracy is classified as incorrect. If a student answers a question correctly, his/her response will be classified as correct.

### C. Proper Function Execution

Initially, the set of functions (begin test, next question, previous question) needed to complete the assessment are listed. Next, the event(s) associated with each function are identified as shown in Table I. The two events that are monitored during this portion of the investigation are the location of left/right mouse clicks and keystrokes. If a key or combination of keys is used that falls within the list of needed keys to execute a function, the student is classified as being on-task. However, if a key or combination of keys is used that falls outside of the list of needed keys to execute a function, the student is classified as being off-task. Similarly, if the mouse is clicked at a location that falls within the list of needed clicks to execute a function, the student is classified as being on-task. However if the mouse is clicked at a location that falls outside of the list of needed clicks to execute a function, the student will be classified as being off-task.

Table I. Functions and Events

Functions	Keystroke	Mouse Left Click Location
Begin Test	n/a	$(x - 686)^2 + (y - 654)^2 \leq 74^2$
Next Page	n/a	$(x - 1125)^2 + (y - 680)^2 \leq 42^2$
Previous Page	n/a	$(x - 245)^2 + (y - 676)^2 \leq 43^2$
Select A	A	$x: 199 - 424; y: 474 - 626$
Select B	B	$x: 448 - 673; y: 474 - 626$
Select C	C	$x: 696 - 921; y: 474 - 626$
Select D	D	$x: 945 - 1170; y: 474 - 626$

Due to the fact that the student may accidentally press the wrong key or click the wrong place on a page, we monitor the events over a period of samples. Each event sample consists of  $n = 8$  events. If more than  $p = 25\%$  of the sample is classified as being off-task, then the entire sample will be classified as off-task. For example, if the student has 7 events that are classified as being on-task and 1 event that is classified as being off-task, this 1 event is ignored and the student is classified as being on-task.

The inequality used to determine when a series of events is on-task is shown in Equation 1. The subset of  $n$  sequential events is categorized by  $E$  which is defined as  $\{x_m + x_{(m+1)} + \dots + x_{(m+7)}\}$ . Each event that effectively executes a function according to Table I yields a value of 1, and each event the does not execute a function yields a 0. Equation 1 is computed at every sample until there is no longer a subset of  $n$  events to evaluate.

$$\text{Equ. 1:} \quad \frac{1}{n} \sum_{i=m}^{m+(n-1)} x_i < p \quad \text{OR} \quad \frac{1}{n} \sum E < p, \quad \forall m = 1, 2, \dots, n$$

### C. Eye Gaze

The six user-states that Asteriadis et al. defined will be broken down into two basic categories – engaged and disengaged<sup>13</sup>. Attentive, full of interest, and frustrated will be classified as engaged, while distracted, sleepy, and not paying attention will be classified as disengaged. All the data

that is classified as disengaged will be used to derive the percent error of Asteriadis et al.'s eye gaze and head pose model as shown in Equation 2.

Equ. 2: 
$$\text{Percent Error} = \frac{T_{\text{disengaged}}(\text{according to eye gaze model})}{T_{\text{total}}} \times 100\%$$

## IV. Experimental Design

### A. Evaluation Scenario

We designed and conducted an experiment in which participants completed a computer-based math test of varying difficulty to explore the trends developed over time associated with engagement in CBE. A total of 13 participants (6 males and 7 females) took part in this experiment. The participants were recruited from an afterschool program in Atlanta, GA. Two students were in the 6<sup>th</sup> grade, five students were in the 7<sup>th</sup> grade, and six students were in the 8<sup>th</sup> grade. This evaluation consisted of 2 segments: the initial validation of the engagement model performed during a pilot study followed by the application of model.

- **Derivation/Validation:** This segment was performed first to gather data and derive the model. This portion of the test consisted of only low difficulty questions, which required low-level cognitive thinking (pilot study).
- **Application:** This segment was performed last to observe how the engagement model and eye gaze performed when only high difficulty questions were given, which required high-level cognitive thinking.

Each student was given no prior information of the material presented on the tests; however, the questions were taken from the state of Georgia's Criterion-Referenced Competency Tests (CRCT) <sup>14</sup>. The students were placed in a normal testing environment within a school. They were told to be very engaged throughout the test. Each student was provided pencil and paper, which was placed next to the keyboard and mouse; however, the students were allowed to move the pencil and paper as they wanted. The recorded mouse and keyboard events were then analyzed to determine total time, response accuracy, and proper event execution.

We designed three 15-question math tests to assess our hypotheses – one for each grade level. Boardmaker Plus is the software that was used to create this simulated math program <sup>15</sup>. The tests were designed for students between the 6<sup>th</sup> and 8<sup>th</sup> grade. Some questions on the test were low difficulty and required low-level cognitive thinking to complete. Most, if not all, of those problems can be computed quickly using mental math because they only require one step to answer. However, other questions were high difficulty and required high-level cognitive thinking to complete. Most, if not all, of those problems cannot be computed quickly using mental math because they require multiple steps to answer. In many cases, pencil and paper may be needed to develop an answer. We then used a web camera to monitor eye gaze and pose, which was used to estimate behavioral user state throughout the test. Event processing was used simultaneously to monitor and record the associated mouse and keyboard events.

### B. Hypotheses

Two hypotheses have been developed for our system based on the current research that measures behavioral user state through eye gaze and head pose.

Hypothesis 1. The student is engaged if his or her series of events (or combination of events) are classified as:

- On-task and correct (regardless of speed)
- On-task, slow or average, and incorrect

Hypothesis 2. Eye gaze and head pose will not be an accurate measure of user state/engagement for the high difficulty questions. The use of pencil and paper will create false-negatives since eye gaze will be directed towards the paper instead of the computer screen.

## V. Results

### A. Total Time

The total time needed to complete each question was calculated and shown in Fig. 1 for the 7<sup>th</sup> and 8<sup>th</sup> grade. Because there were only two students in the 6<sup>th</sup> grade, all of their responses were automatically classified as being of average speed. Using a boxplot, we were able to properly divide the remaining data into its respective quartiles and categorize any outliers as slow or fast. Due to the nature of the box and whisker plot, there will always be a similar distribution between the average, slow, and fast categories as shown in Fig. 2c.

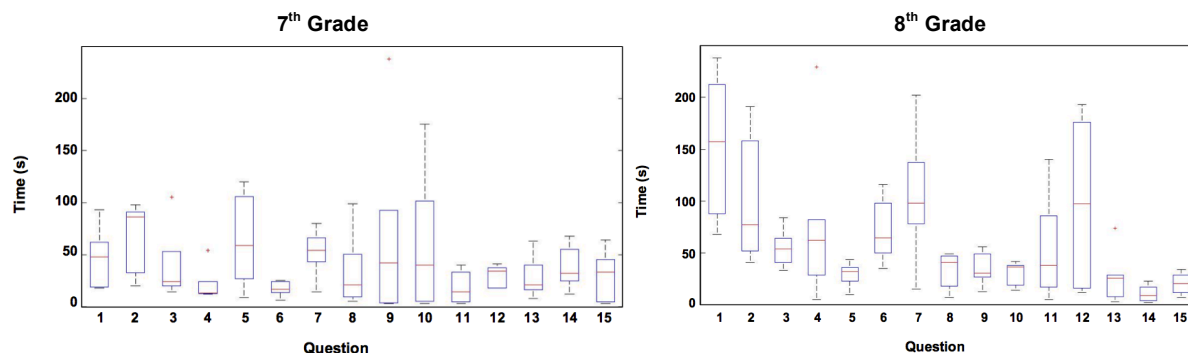


Figure 1. Total time required per question for 7<sup>th</sup> grade (left) and 8<sup>th</sup> grade (right).

### B. Response Accuracy

The students answered 45% of the question correctly and 55% of the questions incorrectly as shown in Fig. 2d. This was expected due to the difficulty of the testing material.

### C. Proper Function Execution

Each subject was given the choice to use either the keyboard or mouse to select his or her response. In result, 4% of the keyboard and mouse input was classified as being off-task as

shown in Fig. 2b. By using Table I, we were able to determine if the mouse clicks and keystrokes occurred within the necessary constraints to successfully navigate through the test. Fig. 2a shows the combinations of events that model engagement and how often each combination occurred during this study.

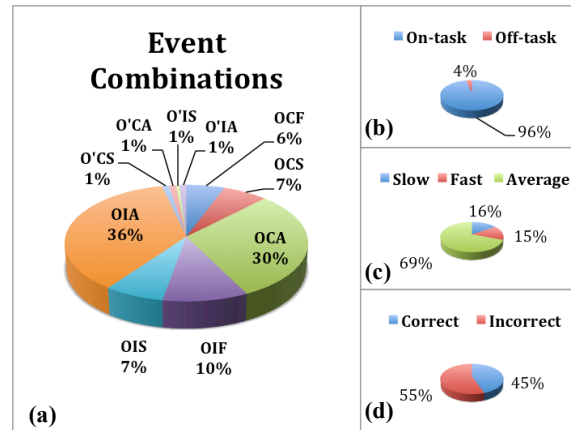


Figure 2. (a) This chart shows the how often we received each combination of events throughout both tests (S=slow, A=average, F=fast, C=correct, I=incorrect, O=on-task, O'=off-task). (b) On-task vs. off-task events. (c) Speed of responses throughout test. (d) Correct vs. incorrect responses.

#### D. Eye Gaze

The eye gaze and head pose technique had an average of a 24.2% error for the 6<sup>th</sup> grade test, a 32.1% error for the 7<sup>th</sup> grade test, and a 34.8% error for the 8<sup>th</sup> grade test. However, for the students who scored considerably higher than their peers, they exhibited up to a 65% error. Fig. 3 shows the relationship between the subject's test score and the amount of time his or her gaze is elsewhere in the room.

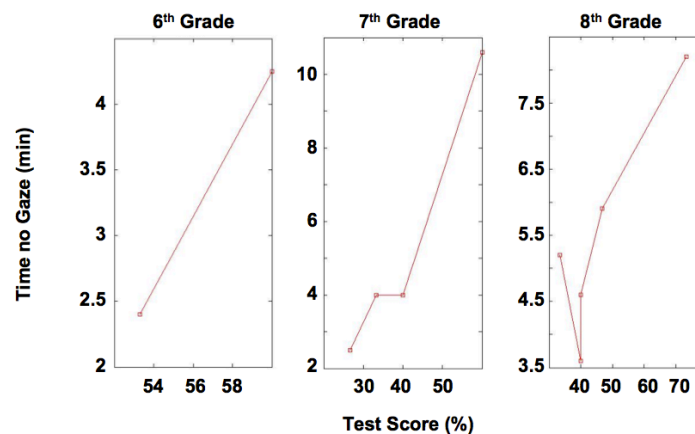


Figure 3. This graph shows the relationship between the subjects' test scores and the amount of time that eye gaze was not on the computer screen.

## VI. Conclusion

If a student is being off-task, he or she is disengaged (regardless of speed and/or response) due to the low occurrence of this in the engagement model. However, more tests need to be conducted to verify this assumption. Furthermore, validity of responses alone is not enough information to determine user-state as exhibited in Fig. 2b. However, speed along with the validity of responses can help to determine more information about the engaged student. If the student is on-task and has a series of fast responses with a series of correct answers (OCF), the student may need questions of higher difficulty. The results show that 6% of the sample was OCF. If the student is on-task and has a series of slow responses with a series of correct answers (OCS), the student may understand the material and require more time to think. The results show that 7% of the sample was OCS. If the student is on-task and has a series of slow responses with a series of incorrect answers (OIS), the student may lack understanding and need questions of lesser difficulty. The results show that 7% of the sample was OIS. This additional information will be used in the future to better integrate instructional scaffolding and adaptation with the device.

This work also suggests that eye gaze and head pose technique is not an effective measure of engagement when high-level cognitive thinking is required, as stated in the hypotheses. During the test, the subjects consistently looked down at the paper to write out the multistep problems and calculate the answers by hand. We also observed that other subjects looked at random objects in space to perform mental math. In fact, we observed that the longer that the student looked away from the computer screen, the higher he or she performed on the test. The 8<sup>th</sup> grade subject who scored the highest looked away from the screen for 8.2 minutes, which was 47.4% of the entire test time. The 7<sup>th</sup> grade subject who scored the highest looked away from the screen for 10.6 minutes, which was 65.4% of the entire test time. The large time delay associated with the lack of eye contact from the human to the computer screen caused Asteriadis et al.'s eye gaze technique to incorrectly declare the subjects as being distracted or disengaged. However, using the event model, we were able to correctly categorize the students as being engaged. By monitoring the time delay/speed, accuracy of responses, and proper event execution associated with each question, we are able to expand the eye gaze model proposed by Asteriadis et al. to a complex problem solving environment<sup>12</sup>.

## VII. Future Work

This investigation is only a starting point for where we would like to go in the future. We have developed a basis to what engagement looks like with keyboard and mouse input; however, there are cases when the model will fail. For example, when the student is taking a long time to input a response, this model would consider the subject to be engaged and assume that the student is either thinking or working the problem out on pencil and paper. What if the student is actually talking to a peer and still manages to submit an answer before the computer categorizes him or her as disengaged? For situations like this, we would like to integrate a robotic platform into this intelligent tutoring system to reinforce engagement.

More specifically, the long-term goal is to create an adaptive robotic tutor using a humanoid robot in conjunction with a touchscreen device. Therefore, we would like to conduct a similar experiment that will effectively transfer the mouse/keyboard model to a touch-screen device. The



mouse clicks will be comparable to a stylus and/or touch screen. The physical keyboard events will be comparable to the events of a virtual keyboard. Mouse movements, which are evident in the CBE setting, will be obsolete once the robotic platform with touchscreen capabilities is utilized.

Also, now that a model of engagement is created and we are able to accurately determine behavioral user state, we need to implement adaptive tutoring. By utilizing behavioral strategies to maintain the student's attention when engagement decreases, we will be able to keep the students engaged continuously. Possible behavioral strategies to implement include, but are not limited to gestures, expressions, eye contact, posture, proximity, tone, pitch, and volume.

## References

- [1] D. Szafrir and B. Mutlu, "Pay Attention! Designing Adaptive Agents that Monitor and Improve User Engagement," in Proc. ACM/SigCHI Conference on Human Factors in Computing, 2012.
- [2] D. M. Christophel, "The relationships among teacher immediacy behaviors, student motivation, and learning." Communication Education, vol. 39, no. 4, pp. 323-40, 1990. [Online]. Available: <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ414768>
- [3] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," REVIEW OF EDUCATIONAL RESEARCH, vol. 74, no. 1, pp. 59-109, Jan. 2004. [Online]. Available: <http://dx.doi.org/10.3102/00346543074001059>
- [4] C. R. Greenwood, B. T. Horton, and C. A. Utley, "Academic Engagement: Current Perspectives on Research and Practice," School Psychology Review, vol. 31, no. 3, pp. 328-349, 2002.
- [5] V. Aleven, "An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor," Cognitive Science, vol. 26, no. 2, pp. 147-179, Apr. 2002. [Online]. Available: [http://dx.doi.org/10.1016/S0364-0213\(02\)00061-7](http://dx.doi.org/10.1016/S0364-0213(02)00061-7)
- [6] J. W. Schofield. In Computers and Classroom Culture. Cambridge University Press, 1995.
- [7] C.-L. C. Kulik and J. A. Kulik, "Effectiveness of computer-based instruction: An updated analysis," Computers in Human Behavior, vol. 7, pp. 75 – 94, 1991.
- [8] J. Lowe, "Computer-based education: Is it a panacea?" Journal of Research on Technology in Education, vol. 34, pp. 163–71, 2002.
- [9] J. R. Anderson, A. T. Corbett, K. R. Koedinger, and R. Pelletier, "Cognitive Tutors: Lessons learned," The Journal of the Learning Sciences, vol. 4, pp. 167–207, 1995.
- [10] H. Wainer, N. J. Dorans, R. Flaugher, B. F. Green, R. J. Mislevy, L. Steinberg, and D. Thissen, *Computerized Adaptive Testing: A Primer*, 2nd ed. Lawrence Erlbaum, Apr. 2000. [Online]. Available: <http://www.worldcat.org/isbn/0805835113>
- [11] V. P. Richmond, J. C. McCroskey, and A. D. Johnson, "Development of the nonverbal immediacy scale (NIS): Measures of self-and other perceived nonverbal immediacy," Communication Quarterly, vol. 51, no. 4, pp. 504-517, 2003. [Online]. Available: <http://dx.doi.org/10.1080/01463370309370170>

[12] C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau, G. Davis, V. T. Zivkovic, R. E. Olmstead, P. D. Tremoulet, and P. L. Craven, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," May 2007. [Online]. Available: <http://www.ingentaconnect.com/content/asma/asem/2007/00000078/a00105s1/art00032>

[13] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias, "Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment," *Multimedia Tools and Applications*, vol. 41, no. 3, pp. 469-493, Feb. 2009. [Online]. Available: <http://dx.doi.org/10.1007/s11042-008-0240-1>

[14] Criterion-Referenced Competency Tests (CRCT) [Online] Available: <http://www.doe.k12.ga.us/Curriculum-Instruction-and-Assessment/Assessment/Pages/CRCT.aspx>

[15] Boardmaker Plus [Online]. Available: <http://www.mayer-johnson.com/boardmaker-plus-v-6/>