

Incorporating effective e-learning principles to improve student engagement in middle-school  
mathematics

Mulqueeny, K., Kostyuk, V., Baker, R.S., Ocumpaugh, J.

**Abstract**

**Background**

The expanded use of online and blended learning programs in K-12 STEM education has led researchers to propose design principles for effective e-learning systems. Much of this research has focused on the impact on learning, but not how instructional design impacts student engagement, which has a critical impact both on short-term learning and long-term outcomes. Reasoning Mind has incorporated the e-learning principles of personalization, modality, and redundancy into the design of their next-generation blended learning platform for middle-school mathematics, named Genie 3. In three studies, we compare student engagement with the Genie 3 platform to its predecessor, Genie 2, and to traditional classroom instruction.

**Results**

Study 1 found very high levels of student engagement with the Genie 2 platform, with 89% time on-task and 71% engaged concentration. Study 2 found that students using Genie 3 spent significantly more time in independent on-task behavior and less time off-task or engaged in on task conversation with peers than students using Genie 2. Students using Genie 3 also showed more engaged concentration and less confusion. Study 3 found that students using Genie 3 spent 93% of their time on-task, compared to 69% in traditional classrooms. They also showed more engaged concentration and less boredom and confusion. Genie 3 students sustained their

engagement for the entire class period, while engagement in the traditional classroom dropped off later in the class session.

## **Conclusions**

The incorporation of evidence-based e-learning principles into the design of the Genie 3 platform resulted in higher levels of student engagement when compared to an earlier, well-established platform that lacked those principles, as well as when compared to traditional classroom instruction. Increased personalization, the use of multiple modalities, and minimization of redundancy resulted in significant increases in time on-task and engaged concentration, but also a decrease in peer interaction. On the whole, this evidence suggests that capturing students' attention, fostering deep learning, and minimizing cognitive load leads to improved engagement, and ultimately better educational outcomes.

**Keywords:** Blended Learning, E-Learning, Middle-School Mathematics, Time On-Task, Engaged Concentration, BROMP

## **Background**

As online and blended learning continues to see rapid expansion in K-12 (e.g., Horn and Staker 2011), particularly in STEM (Science, Technology, Engineering, and Mathematics) fields (e.g. Heffernan and Heffernan 2014; Koedinger and Corbett 2006), a growing body of research has begun to explore and develop design principles to ensure its efficacy (e.g. Betrancourt 2005; Clark and Mayer 2011; Garrison and Anderson 2003; Govindasamy 2002). Most of this research is based on empirical investigations of individual principles. This research strategy provides strong evidence about design features in isolation, but it is less informative when it comes to understanding how they function in concert.

In addition, little research has examined how these design principles influence not just immediate domain learning, but engagement as well, which may mediate long-term student outcomes. Indeed, many studies of the cognitive benefits of design principles do not consider how the removal of potentially appealing factors may (negatively) impact student engagement (see, for instance, Harp and Mayer 1998). Particularly in real-world settings where educational software must compete with many other activities, determining whether or not design features are engaging students is a necessary component of evaluating their effectiveness, leading many to investigate behavioral and affective indicators of student engagement in STEM learning systems. While findings suggest that some time off-task can refocus bored or frustrated students (Baker et al. 2011; Sabourin et al. 2011), students who completely disengage with educational software, spending large amounts of time off task, show lower learning (Goodman 1990), both in the short-term and in the long-term, the latter a result of aggregate effects from a loss of practice opportunities (Cocea et al. 2009). Other disengaged behaviors such as carelessness and gaming the system are also associated with poor learning outcomes (Cocea et al. 2009; Pardos et al. 2014). Furthermore, findings suggest engaged concentration (or Csikszentmihalyi's "flow") is positively associated with learning, while boredom leads to poor learning outcomes (Craig et al. 2004; D'Mello and Graesser 2012; Pardos et al. 2014). Confusion and frustration, have more complex relationships to learning; while necessary for learning (D'Mello et al. 2014), spending a considerable amount of time confused or frustrated is associated with worse outcomes (e.g., Liu et al. 2013). In addition, both behavioral engagement and affect are associated with long-term student participation in STEM; for example, boredom, confusion, engaged concentration, gaming the system, and carelessness in middle school mathematics are predictive of eventual college attendance (San Pedro et al. 2013), and gaming the system and carelessness are predictive of

whether or not a student enrolls in a STEM degree program (San Pedro et al. 2014). Researchers are beginning to explore the relationship between design features and engagement (e.g. Baker et al. 2009; Doddannara et al. 2013; D'Mello et al. 2014), but to date few studies have explored the causal impact of well-known and widely-used design principles on engagement.

One of the more comprehensive discussions of designing for multimedia learning systems has been put forward by Clark and Mayer (2011), who present eight principles based on previous research. These include: (1) *Personalization*: Use a conversational style, polite speech, and virtual coaches (Moreno et al. 2001). (2) *Multimedia*: Use words and graphics, not words alone (Halpern et al. 2007). (3) *Contiguity*: Align words to corresponding graphics (Moreno and Mayer 1999). (4) *Coherence*: Limit extraneous information (Mayer et al. 2001) (5) *Modality*: Present words as audio, rather than text (Low and Sweller 2005). (6) *Redundancy*: Explain visuals with spoken word or text, not both (Mayer and Moreno 2003). (7) *Segmenting*: Present lessons in small, well-spaced units (Mayer and Chandler 2001). (8) *Pretraining*: Ensure that learners know the names and characteristics of key concepts (Kester et al. 2006). Each of these principles is designed to enhance learning by focusing students' attention and limiting cognitive load. Using design to focus the students' attention on the critical task of learning mathematics should minimize the attentional resources needed to inhibit distractions (Mayer and Moreno 2003), allowing for greater and more prolonged attention to be paid to learning.

In this paper, we investigate whether design changes that reflect three of Clark and Mayer's (2011) principles—*Personalization, Modality, and Redundancy*—improve student engagement with an online STEM learning system, Reasoning Mind (RM). RM, developed by the nonprofit company of the same name, currently provides blended learning instruction in mathematics to over 100,000 students in the United States. RM works with expert teachers to

design online learning experiences that re-create best-practices for instruction (Khachatryan et al. 2014), providing elementary and middle school curricula that focuses on fostering deep understanding of core mathematical topics necessary for students' later success in algebra. Having instruction delivered by the computer frees teachers to conduct the sort of targeted interventions with struggling students that research suggests is most effective at improving student learning (Bush and Kim 2014; Waxman and Houston 2012). In this paper, we study whether design changes to this platform, which reflect Clark and Mayer's principles, have a positive impact on student engagement.

### The Reasoning Mind Blended Learning Systems

In this article, we will study two generations of Reasoning Mind's online learning systems: Genie 2 and Genie 3, used by elementary and middle school students. RM developed the Genie 2 platform in 2005, designing instruction in line with the practices of expert teachers (Khachatryan et al. 2014) within a system where students receive immediate, individualized feedback while learning, primarily from a pedagogical agent known as the Genie. However, the design of this platform did not purposefully incorporate research-based principles of e-learning such as those in Clark and Mayer (2011). As of 2013, RM has been piloting the next-generation Genie 3 platform, which explicitly incorporates three instructional design principles previously found to increase learning gains in online instruction (Clark and Mayer 2011). The improvements made in Genie 3 to incorporate the *personalization*, *modality*, and *redundancy* principles are outlined in Table 1.

#### **Genie 2**

The Genie 2 platform presents students an online environment named RM City (see Figure 1), where different buildings represent different types of learning activities. *Guided Study* is the main learning mode, where students study curriculum objectives. In *Homework*, students enter answers to mathematics problems chosen by the system based on the student's progress and prior performance. These problems are printed out and given as homework by the teacher. The student completes the problems at home and then types in the answers at the beginning of class. The *Office* allows teachers to assign individual objectives from Guided Study or practice problems for material that the student needs to focus on. The *Wall of Mastery* provides students opportunities to challenge themselves with more difficult problems. Throughout the system, the Genie acts as an empathic virtual guide, providing solutions and encouragement.

In *Guided Study*, where students spend the majority of their time in the RM system, the curriculum is divided into a series of objectives, or mathematical topics, for students to complete. Examples of objectives include "Numerical Expressions with Parentheses," "Comparing Fractions with Different Denominators," and "Rounding Decimals." Each objective consists of a sequence of pedagogical stages: warm up, theory instruction, a notes test, a series of increasingly difficult practice problems, and a review. As students progress through each stage of an objective, their progress is charted on a virtual map. Objectives contain animated stories with a recurring cast of characters, as well as illustrations and animations that closely correspond to the problems students are solving. All instruction is delivered through text, with optional narration that reads out the text on the screen. Illustrations and accompanying explanatory text are positioned closely to facilitate comprehension (Moreno and Mayer 1999).

Objectives are strictly sequenced based on prerequisite skills, but a student's progress through an objective is self-paced, allowing the student to navigate forward and backward to

review the material. Upon successful completion of an objective – defined by an accuracy cutoff – the student proceeds to the next objective in the sequence. When a student fails an objective, the system uses automatic diagnostics and remediation to fill in gaps in understanding that are hindering progress. Strong students move quickly and are challenged with more difficult problems.

### **Genie 3**

The Genie 3 platform has a less cartoon-like home screen than Genie 2 (see Figure 3). It includes some of the same learning activities from the earlier platforms (*Guided Study*, *Homework*, and *Wall of Mastery*) but it also adds the *Test Center*, where students complete tests and quizzes, and the *Math Journal*, a repository of the key rules and definitions provided by the system from the lessons the student has completed.

In the Genie 3 platform, *Guided Study* is redesigned to simulate ideal classroom experiences provided by expert teachers. As such, students control a customizable avatar (see Figure 4), completing daily lessons in a virtual small-group session with a simulated tutor and two simulated peers. While Genie 2's characters, including the Genie, are used mainly for motivation, positive feedback, and emphasis of key points, Genie 3's rotating cast of three tutors and seven peers act as full pedagogical agents (cf. Forsyth et al. 2014). The tutors lead the instruction of the lessons, ask the student or virtual peers to solve problems, and prompt the real student to evaluate the virtual peer's solution or work collaboratively, solving individual parts of a multi-step problem. Virtual peers model positive attitudes toward mathematics, demonstrate common misconceptions, and play a motivational role, encouraging the real student, sympathizing with difficulties, and emphasizing the value of persistence and hard work. All of

the agents use an informal conversational style, in line with the principle of *personalization* (Moreno et al. 2001), and are narrated by voice actors, supporting multiple *modalities* (Low and Sweller 2005).

The small-group lesson environment uses a shared virtual white board, where diagrams, problems, key definitions, rules, and statements are written, and where students work to solve problems. No other text is presented; spoken narration is used to carry the majority of instruction, a design choice in keeping with the principles of *modality* (Low and Sweller 2005) and *redundancy* (Mayer and Moreno 2003). Diagrams and illustrations are paired closely with explanatory labels and text, in line with the *continuity* principle (Moreno and Mayer 1999). Lessons are broken up into pedagogical segments corresponding to classroom lessons. Typically, lessons include a warm-up, introduction of new material, a series of practice problems, and review. Completion of the lesson is tracked on a progress bar at the bottom of the screen.

### Measuring Student Engagement

Engagement is a concept that has been defined in many ways (see review in Fredricks et al. 2004). Finn and Zimmer (2012) outline four components of engagement thought to impact student learning and achievement: academic, social, cognitive, and affective. Both academic and social engagement are comprised of behavioral indicators (treated as a single construct in Fredricks et al. 2004). The former refers to behaviors related to the learning process, while the latter reflects whether or not the student follows written and unwritten rules for classroom behavior. Cognitive engagement involves the use of mental resources to comprehend complex ideas. Affective engagement is the emotional response and feelings of involvement in school.

Previous research has often examined these constructs using survey methods. For example, Finn et al. (1991) administered a questionnaire to teachers, finding that academic behaviors that reflect effort and initiative are positively correlated with end of year achievement test scores ( $r= 0.40$  to  $0.59$ ), while inattentive behavior is negatively correlated with achievement ( $r = -0.52$  to  $-0.34$ ). More recently, research has found a significant relationship between academic and social engagement in fourth and eighth grades and high school graduation (Finn and Zimmer 2012).

In this paper, student engagement measures (discussed more thoroughly in next section) are investigated in series of three field observation studies that investigate the effect of the Reasoning Mind mathematics curricula on the prevalence of these indicators. Study 1 reports on observations of student engagement that were conducted when students were using the Genie 2 platform. Study 2 uses the same observation method to compare the engagement of students using Genie 2 to those using Genie 3. Finally, Study 3 compares students using Genie 3 to students in a traditional mathematics classroom (with no technological support).

## **BROMP Field Observations**

Quantitative field observations of student engagement were collected using the Baker Rodrigo Ocumpaugh Monitoring Protocol (or BROMP), an established observation method with over 150 certified coders in four countries (Ocumpaugh et al. 2015). BROMP has been used to investigate behavioral and affective indicators of student engagement in a number of different online learning environments (e.g., Baker et al. 2010, 2014; Paquette et al. 2014; Pardos et al. 2014; Rodrigo et al. 2008), including research on college attendance and engagement within ASSISTments (San Pedro et al. 2013, 2014).

Within this method, trained observers repeatedly record observations of educationally relevant behavior and affect of students individually, in a pre-determined order, ensuring roughly equal samples of each student's behavior. Observers record the first behavior and affect they see, but have up to 20 seconds to make that decision. In this study, behavior codes included *On Task – Independent* (i.e. working alone on an assigned task), *On Task – Conversation* (i.e. discussing work with a peer or teacher), *Off Task* (i.e. not working on their assigned task), and *Gaming the System* (i.e. systematic guessing or use of hints to obtain answers rather than learning)—all of which are typically coded for during BROMP observations. However in Studies 2 and 3, the category of *On-Task Conversation* was split into two categories: *On-Task – Proactive Remediation*, which was coded when students received individual or small group interventions from the teacher (Miller et al., 2015), and all other *On-Task – Conversation* behaviors. Affective states included *Engaged Concentration, Boredom, Confusion, Frustration, and Delight* (D'Mello et al. 2010). Cases where the student had stepped out of class or their behavior or affect were otherwise impossible to classify or outside the coding scheme were coded as *Other* and are not included in the analysis. All observers in this study were BROMP-certified (Ocumphaug et al. 2015), meaning that they had obtained an acceptable inter-rater reliability (Cohen's Kappa  $> 0.6$  on each coding scheme) with a previously-certified BROMP coder during training sessions identical to the observations performed in all three studies.

### Research Questions

Research Question 1: Is the Reasoning Mind program more engaging than traditional classroom instruction? We hypothesize that students using these blended learning systems (both Genie 2 and Genie 3) will show greater levels of student engagement than students participating in traditional, face-to-face instruction.

Research Question 2: Is the Genie 3 platform more engaging than Genie 2? We hypothesize that the improvements to Genie 3 in the domains of *personalization*, *modality*, and *redundancy* (Clark and Mayer 2011) will lead to improved student engagement.

### **Study 1**

In Study 1, students who are using the Reasoning Mind Genie 2 platform were observed. In this first, pilot study, only one condition was observed; it is included here as it gave a baseline for engagement in the most established version of the RM blended learning system, and inspired the remaining two studies.

Fifth-grade students from three different schools in the Texas Gulf Coast region were observed while using Genie 2 as their regular mathematics curriculum. Two schools were in urban areas with large class sizes (approximately 25 students each), and served predominantly minority populations (one mostly Latino and the other African American). The third was a suburban charter school with smaller classes (approximately 15 students each) and a predominantly White population. For each of the three schools, two classes were observed for one class period. Due to a data collection error, student IDs were not linked to the observations, in any form; as such, it is infeasible to conduct statistical significance tests of engagement without violating independence assumptions. However, each student was sampled an equal number of times, making averaging across students feasible.

### **Results and Discussion**

Results are given in Table 2. The overall incidence of behavior and affect indicates high engagement. Students were *on-task* 89% of the time, which is higher than values observed in Cognitive Tutor classrooms in U.S. suburban middle schools (Baker et al. 2004) or traditional

classrooms (Lloyd and Loper 1986; Lee et al. 1999). *Gaming the system*, where students misuse the software in order to succeed without learning, is almost non-existent, suggesting that students are taking the program seriously. Patterns of affect also indicated high levels of engagement, with students exhibiting high levels of *engaged concentration* (71%) and relatively low levels of *boredom* (10%). Low-to-moderate levels of *confusion* (9%) and *frustration* (7%) are on par with previous studies (Pardos et al. 2014) and suggest that students are being challenged to learn new material. These results demonstrate that Genie 2, which has been used annually by tens of thousands of elementary students over the last ten years, is already quite engaging.

## Study 2

Study 2 compares student engagement with the Genie 2 platform to the newly developed Genie 3. As explained above, this platform's design, which targets middle-school students, offers continuity with the Genie 2 platform, but incorporates improvements in several research-based e-learning principles, particularly including personalization, multimedia, and modality.

Study 2 employs a quasi-experimental design. Teachers within the same school were assigned to teach with the traditional or Genie 3 curriculum by the school principal, and students were non-randomly assigned to each group for the school year. Both groups were observed once in the fall semester and again in the spring. The observation procedure was similar to Study 1. Two BROMP-certified coders conducted the observations. In this study, observers did not code for *Gaming the System*, which was all but non-existent in Study 1, but they did include an additional behavioral category. Because one of the anticipated benefits of blended learning is that it frees teachers to engage more frequently in targeted interventions, BROMP observers also coded for *On Task – Proactive Remediation*, as discussed above. These cases were previously

coded as *On Task – Conversation*, so we would expect a comparable reduction in that behavior compared to Study 1.

The subjects we observed in this study were sixth-grade students in a small, central Texas city, with a student population that is one-third Latino and one-third White. We observed six classes (126 students in the fall and 125 in the spring) using the new Genie 3 platform and six classes (122 students in the fall and 123 in the spring) using the Genie 2 platform. The two groups performed equivalently on a pre-test measure of key topics in sixth-grade mathematics. The Genie 3 group scored an average of 32.63% ( $SD = 17.06$ ), while the Genie 2 group averaged 32.72% correct ( $SD = 14.49$ ), an effect size (Cohen's  $d$ ) of 0.006.

### Results and Discussion

2,966 observations were collected from the Genie 3 classrooms (1,570 in the fall and 1,396 in the spring) and 2,764 observations were collected from Genie 2 classrooms (1,510 in the fall and 1,254 in the spring). Average distributions for these codes are given in Figure 5.

Proportional data are constrained and tend not to be normally distributed, this was particularly the case here, with very many students having either very high or very low proportions of engagement in any one behavior or affective category. Table 3 shows the measurements of skewness and kurtosis for each of the behaviors and affects observed. Applying the rule of thumb that the ratio of skewness and kurtosis to the corresponding standard error should be within  $\pm 2.58$ , only the *on task – independent* behavior in Genie 2 classes had suitably low kurtosis, but all distributions were skewed beyond normality. Because the proportional data was not normally distributed, we applied an arcsine transformation (calculating the arcsine of the square root) to the proportion of observations classified in a given behavioral or affective

category. This transformation is used to normalize the distribution of proportional data, which are limited to values between 0 and 1. By extending this range, we expand the difference between extreme values (near 0 and 1) and compress the difference between central values (near 0.5; McDonald 2014). With more normally distributed data, we were able to perform an analysis of variance (ANOVA) for each behavior and affective state to compare the frequency in each group of students.

The ANOVA found that, for both semesters, students using Genie 3 spent more time in *on task – independent* (88.9% vs. 75.7%,  $F(1,494) = 61.22, p < 0.001$ ) and less in *on-task conversations* (2.4% vs. 8.3%,  $F(1,494) = 62.44, p < 0.001$ ) and off-task (5.4% vs. 12.0%,  $F(1,494) = 33.37, p < 0.001$ ). There was not a significant difference in the time spent in *on task - proactive remediation* (Fall: 3.3% vs. 4.0%,  $F(1,494) = 1.11, p = 0.293, ns$ ).

Similarly, we used an arcsine transformation of the proportion of each student's observations classified as each affective state. ANOVA found significantly higher levels of *engaged concentration* for Genie 3 students (86.8% vs. 82.3%,  $F(1,494) = 9.90, p < 0.005$ ) and less *confusion* (1.0% vs. 5.0%,  $F(1,494) = 46.19, p < 0.001$ ). There were not significant differences in *boredom* (12.2% vs. 12.5%,  $F(1,494) = 0.27, p = 0.603, ns$ ), *frustration* (0.0% vs. 0.2%,  $F(1,494) = 1.15, p = 0.283, ns$ ), or *delight* (0.0% vs. 0.0%,  $F(1,249) = 1.00, p = 0.318, ns$ ).

In addition to determining whether engagement indicators differed across these two conditions, we were also interested in determining whether temporal dynamics might be influencing these results. Specifically, we evaluated whether engagement indicators shifted over the course of a lesson-period by comparing average distributions in the first 30 minutes of class

and the second 30 minutes. Again, we performed an arcsine transformation of the proportional data, and used ANOVA to compare each behavior and affect category in the chosen timeframe.

There was a significant increase in the time *Off Task* in the Genie 3 classes during the second half of the class (3.8% vs. 7.3%,  $F(1,469) = 5.18, p < 0.05$ ). This corresponded to a moderate decrease in *On Task – Independent*: (90.5% vs. 85.6%,  $F(1,469) = 3.51, p = 0.062, ns$ ), but there were no changes in other behavior rates (*On Task – Conversation*: 2.0% vs. 3.2%,  $F(1,469) = 1.92, p = 0.166, ns$ ; *On Task – Proactive Remediation*: 3.8% vs. 3.9%,  $F(1,469) = 0.19, p = 0.665, ns$ ), nor among the affective states (*Engaged Concentration*: 88.7% vs. 85.0%,  $F(1,469) = 2.38, p = 0.124, ns$ ; *Boredom*: 10.3% vs. 13.5%,  $F(1,469) = 2.07, p = 0.151, ns$ ; *Confusion*: 0.9% vs. 1.5%,  $F(1,469) = 0.47, p = 0.494, ns$ ; *Frustration*: 0.00% vs. 0.02%,  $F(1,469) = 1.10, p = 0.294, ns$ ; *Delight*: No Variance).

These results contrast to the Genie 2 classes, where neither the behavioral nor the affective indicators of engagement changed from one half hour to the next, for any construct (*On Task – Independent*: 76.7% vs. 79.6%,  $F(1,465) = 2.09, p = 0.149, ns$ ; *On Task – Conversation*: 8.0% vs. 7.2%,  $F(1,465) = 0.61, p = 0.437, ns$ ; *On Task – Proactive Remediation*: 3.37% vs. 3.36%,  $F(1,465) = 0.02, p = 0.890, ns$ ; *Off Task*: 12.0% vs. 9.8%,  $F(1,465) = 2.14, p = 0.144, ns$ ; *Engaged Concentration*: 84.0% vs. 83.0%,  $F(1,465) = 0.02, p = 0.881, ns$ ; *Boredom*: 11.1% vs. 12.3%,  $F(1,465) = 0.01, p = 0.934, ns$ ; *Confusion*: 4.69% vs. 4.66%,  $F(1,465) = 0.75, p = 0.388, ns$ ; *Frustration*: 0.10% vs. 0.08%,  $F(1,465) = 0.01, p = 0.942, ns$ ; *Delight*: 0.1% vs. 0.0%,  $F(1,465) = 1.72, p = 0.191, ns$ ).

At the end of the school year, students in both groups completed the same assessment of key topics in sixth-grade mathematics that was given as a pre-test in the fall. Genie 3 students

improved significantly more than Genie 2 over the course of the year, improving 25.41 percentage points on average compared to 16.47 ( $t(209) = 4.60, p < 0.001$ ), an effect size of 0.63 standard deviations.

The improvements in the Genie 3 platform over Genie 2 improved students' *independent time on-task* and *engaged concentration* while reducing *time off-task*. The lack of a difference in time spent in *proactive remediation* suggests that design differences, which were aimed at changing the students' engagement, did not impact the frequency in which teachers offered help to individual students. This is not surprising, since teachers typically seek to maximize the time they can spend delivering this kind of instruction, regardless of the educational software students are using. The more frequent occurrence of *on-task conversation* in Genie 2 classes is likely due to the use of audio instruction in Genie 3 that requires students wear headphones, making it more difficult for students to talk to each other. It is possible that this is also the cause for the change in time *on-task* as well.

### **Study 3**

The design and procedure was identical to that of Study 2, except for the use of a traditional instruction control, rather than the Genie 2 platform, and the number of subjects. Students were arbitrarily assigned to classes (i.e. not randomly), and the principal assigned two teachers to the traditional instruction condition and two to the Genie 3 condition, again not randomly. The traditional, face-to-face instruction classes included teacher lectures, individual worksheet exercises, whole-class work on an overhead projector, and work in pairs and small groups. Teachers did not use a complete, published curriculum, but pulled material from a

variety of sources. The use of multimedia materials, such as videos or smart boards, was not observed, and the classrooms did not have computers present.

We observed twelve sixth-grade classrooms at one middle school in a majority Latino, urban Texas school district. In the fall, six classes (118 students) used the Genie 3 curriculum and four classes (95 students) received traditional classroom instruction. In the spring the same six classes using Reasoning Mind (109 students) and six classes using the traditional curriculum (132 students) were observed.<sup>1</sup> The two groups did not significantly differ on a Reasoning Mind-developed pre-test measure of key topics in sixth-grade mathematics, the Genie 3 group answered an average of 32.00% of questions correct ( $SD = 13.11$ ), while the traditional instruction group averaged 28.37% questions correct ( $SD = 17.03$ ), an effect size (Cohen's  $d$ ) of 0.23.

### Results and Discussion

Observers collected 3,085 observations from Genie 3 classes (1,649 in the fall and 1,436 in the spring) and 2,879 observations from traditional classrooms (1,131 in the fall and 1,748 in the spring). Average distributions for these observations are given in Figure 6. They show, broadly, that Genie 3 students spent more time on task and in engaged concentration.

Table 4 shows the measurements of skewness and kurtosis for each of the behaviors and affects observed in Study 3. Applying the rule of thumb that the ratio of skewness and kurtosis to the corresponding standard error should be within  $\pm 2.58$ , only four of the ten distributions had suitably low kurtosis, but all distributions were skewed beyond normality. As in Study 2, we used an arcsine transformation of the proportion of each student's observations classified as each behavioral category. ANOVA results show that the average proportions of all behavioral

categories were significantly different when comparing the Genie 3 students to those in traditional classrooms. Genie 3 students spent more time in *on task – independent* (84.5% vs. 60.9%,  $F(1,452) = 198.92, p < 0.001$ ), more time in *on task – proactive remediation* (7.3% vs. 0.0%,  $F(1,452) = 61.00, p < 0.001$ ), less time in *on-task – conversation* (1.5% vs. 7.8%,  $F(1,452) = 106.17, p < 0.001$ ), and less time *off-task* (6.7% vs. 31.3%,  $F(1,452) = 380.99, p < 0.001$ ) than students in the traditional classroom.

As with the behavioral data, ANOVA results show that the two groups are significantly different in terms of their affective engagement measures. Genie 3 students showed higher levels of *engaged concentration* (74.8% vs. 66.2%,  $F(1,452) = 23.72, p < 0.001$ ), less *boredom* (23.3 % vs. 30.5%,  $F(1,452) = 19.34, p < 0.001$ ), less *confusion* (1.8% vs. 3.1%,  $F(1,452) = 16.02, p < 0.001$ ) and less *delight* (0.0% vs. 0.3%,  $F(1,452) = 7.02, p < 0.01$ ) than students in the traditional classroom. Only *frustration* did not show a significant difference between conditions (0.1% vs. 0.0%,  $F(1,452) = 3.73, p = 0.054, ns$ ).

As in Study 2, we compared the behavior and affect observed in the first half hour of class against the second half hour (See Figure 7). In this case, results show that Genie 3 students sustain high engagement throughout their lessons. In this condition, there were no significant changes in any behavioral or affective category from the first to second 30-minute period (*On Task – Independent*: 87.2% vs. 82.9%,  $F(1,411) = 1.96, p = 0.162, ns$ ; *On Task – Conversation*: 1.1% vs. 1.3%,  $F(1,411) = 0.13, p = 0.724, ns$ ; *On Task – Proactive Remediation*: 6.2 % vs. 9.1%,  $F(1,411) = 1.41, p = 0.236, ns$ ; *Off Task*: 5.5 % vs. 6.6%,  $F(1,411) = 0.23, p = 0.633, ns$ ; *Engaged Concentration*: 76.2% vs. 76.6%,  $F(1,411) = 0.50, p = 0.482, ns$ ; *Boredom*: 21.4% vs. 21.9%,  $F(1,411) = 0.09, p = 0.763, ns$ ; *Confusion*: 2.2% vs. 1.5%,  $F(1,411) = 0.73, p = 0.395, ns$ ; *Frustration*: 2.9% vs. 2.2%,  $F(1,411) = 0.67, p = 0.413, ns$ ; *Delight*: No variance).

Students in traditional classrooms, however, showed decreased engagement over the course of a lesson-period. Here, *independent on task behavior* dropped significantly in the second half hour, from 65.5% to 58% ( $F(1,452) = 9.91, p < 0.005$ ), while *off-task behavior* increased from 28.2% to 34.2% ( $F(1,452) = 9.02, p < 0.005$ ). There was a moderate increase in *on-task conversation*, from 6.2% to 7.8% ( $F(1,452) = 3.88, p = 0.05, ns$ ) and no change in *proactive remediation* (0.00% vs. 0.04%,  $F(1,452) = 1.00, p = 0.318, ns$ ). Among affective states, the traditional classroom students saw a significant increase in *delight* (0.0% vs. 0.3%,  $F(1,452) = 5.98, p < 0.05$ ), however this should not be seen as a positive development, as further investigation discovered that all cases of *delight* corresponded with *off-task behavior*. There was a moderate increase in *boredom*, from 29.0% to 33.1% ( $F(1,452) = 3.05, p = 0.081, ns$ ), while *engaged concentration* (67.8% vs. 63.9%,  $F(1,452) = 2.15, p = 0.144, ns$ ), *confusion* (3.2% vs. 2.7%,  $F(1,452) = 0.00, p = 0.950, ns$ ), and *frustration* (0.0% vs. 0.0%, no variance) did not change.

At the end of the school year, students in both groups completed the same assessment of key topics in sixth-grade mathematics that was given as a pre-test in the fall. Genie 3 students improved significantly more than traditional instruction over the course of the year, improving 21.70 percentage points on average compared to 9.81 ( $t(221) = 6.03, p < 0.001$ ), an effect size of 0.81 standard deviations.

The Genie 3 students demonstrated consistently higher levels of engagement than students in the traditional classroom, and while engagement, particularly time on task, decreased over the course of the lesson in the traditional classrooms, Genie 3 students sustained engagement throughout the entire class.

## Conclusions

In terms of Research Question #1: all three of these studies found very high levels of student engagement when using the Reasoning Mind blended learning program, in support of our initial hypothesis. In both the Genie 2 elementary school platform and the Genie 3 middle school platform, RM students demonstrated over 65% of engaged concentration and over 85% time on-task. These high levels of engagement are sustained for an entire hour-long mathematics lesson. Continued engagement creates a greater opportunity for student learning, and as discussed above, increases achievement (Finn et al. 1991) and odds of high school graduation (Finn and Zimmer 2012). Several of the e-learning principles discussed in the introduction (Clark and Mayer 2011) serve to capture a student's focus, hold it, and minimize cognitive load (Chandler and Sweller 1991; van Merriënboer and Sweller 2005). A few principles are used similarly in both Genie 2 and 3. By the very nature of blended learning, both employ multimedia, with a mixture of words and graphics (Halpern et al. 2007), although Genie 3 uses significantly more audio than Genie 2. Lessons in both platforms use segmenting (Mayer and Chandler 2001) to allow for frequent mental breaks as the lesson is split up into chunks that are more easily digested and allow students to see the progress they are making. Contiguity (Moreno and Mayer 1999) and coherence (Mayer et al. 2001) limit the cognitive load of instruction by using visual information to support comprehension and restricting unnecessary information that would require the use of additional cognitive resources to inhibit (Pasolunghi, Cornoldi, and Liberto 1999).

Pre-post measures of mathematics achievement also found that increased engagement corresponded with greater mathematics learning gains. In both studies 2 and 3, Genie 3 students improved their performance approximately 10 percentage points more than the comparison groups. These findings lend further support to previous findings that increases in student

engagement correspond with better learning outcomes (Craig et al. 2004; D'Mello and Graesser 2012; Pardos et al. 2014).

Research Question #2 considered the differences in the design of the two Reasoning Mind platforms. Several principles are emphasized in Genie 3 over Genie 2 that may account for the differences observed in Study 2, where students were significantly more engaged in Genie 3, again, supporting our hypothesis. Applying the modality and redundancy principles, Genie 3 presents the instruction predominantly as audio, and supplements it with text. This serves two purposes: to enhance learning through dual processing (Mayer and Moreno 1998) and to limit distractions to learning as students use headphones to listen to their own instruction. The personalization of Genie 3, in which students create their own avatar and engage in an informal virtual tutoring session with full pedagogical agents encourages greater engagement than the cartoonish Genie 2 platform (Moreno et al. 2001). Further research is necessary to determine which of these improvements had the greatest impact on student engagement.

The Genie 2 platform did offer some advantages over Genie 3. For instance, although boredom was not significantly different between Genie 2 and Genie 3 in Study 2, there appeared to be less boredom in Genie 2 in Study 1. Future research should monitor boredom in particular to see if boredom is lower for Genie 2 within specific populations. Also concerning is the significantly lower levels of on-task conversation seen in Genie 3 compared to Genie 2, which suggests that there is little collaborative learning going on in the classroom. This is likely caused by the use of headphones to present spoken instruction. While this design choice had the benefit of reducing distractions, this benefit may come at the cost of discouraging conversation with peers. As such, it may be desirable for future versions of Genie 3 to add features which connect students with their actual peers, not just virtual ones.

It is also unclear why the consistency of behavior observed in Genie 3 students in Study 3 was not seen in Study 2. However, it is notable that the change in Study 2 amounted to an increase in *off-task* behavior of 3.5%, which, since it was almost double the rate of the first half hour, represents a statistical, but not necessarily a practical change in behavior. However, further study and replication should determine whether the consistency throughout the lessons is replicated in both Genie 2 and Genie 3.

Some limitations of the studies in this paper are that students were not randomly assigned to groups and there were no baseline measures of engagement. While this is typical of *in situ* studies, where researchers must be willing to work around the primary needs of the school, it does limit our ability to make full conclusions as to causality. On the other hand, since schools rarely make classroom assignments on a random basis, these results may be more typical of field conditions than a fully random trial would have been. Further large-scale studies will attempt to determine if the high levels of student engagement seen among Reasoning Mind students generalizes more broadly, but these results offer promising support for the engagingness of blended learning systems, particularly when they incorporate appropriate design principles.

#### List of Abbreviations

ANOVA: Analysis of Variance

BROMP: Baker-Rodrigo-Ocumpaugh Monitoring Protocol

RM: Reasoning Mind

STEM: Science, Technology, Engineering, and Math

#### Endnotes

1. Note the discrepancy in the number of classes in the traditional classroom condition  
was due to an emergency at the school

## References

- Baker, RS, Corbett, AT, Koedinger, KR (2004). Off-task behavior in the cognitive tutor classroom: When students game the system. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 383-390.
- Baker, RS, D'Mello, SK, Rodrigo, MMT, Graesser, AC (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223-241.
- Baker, RSJd, de Carvalho, AMJA, Raspas, J, Aleven, V, Corbett, AT, Koedinger, KR (2009). Educational software features that encourage and discourage "gaming the system". *Proceedings of the 14<sup>th</sup> International Conference on Artificial Intelligence in Education*. 475-482.
- Baker, RSJd, Moore, G, Wagner, A, Kalka, J, Karabinos, M, Ashe, C, Yaron, D (2011). The dynamics between student affect and behavior occurring outside of educational software. *Proceedings of the 4<sup>th</sup> Bi-Annual International Conference on Affective Computing and Intelligent Interaction*, 14-24.
- Betrancourt, M (2005). The animation and interactivity principles in multimedia learning. In RE Mayer (Ed.) *The Cambridge Handbook of Multimedia Learning* (pp. 287-296). New York, NY: Cambridge University Press.
- Bush, J & Kim, M (2014). Evaluation of Reasoning Mind Final Report 2013-14. Program Evaluation. *Dallas Independent School District*.
- <http://www.dallasisd.org/cms/lib/TX01001475/Centricity/domain/98/evaluation/13-14/finalrpts/EA14-512-2-Reasoning-Mind-Program-Final.pdf>. Accessed 1 March 2015.

- Chandler, P, & Sweller, J (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293-332.
- Clark, RC & Mayer, RE (2011). E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning (3<sup>rd</sup> ed.). San Francisco, CA: *Pfeiffer*.
- Cocea, M, Hershkovitz, A, Baker, RSJd (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate? *Proceedings of the 14<sup>th</sup> International Conference on Artificial Intelligence in Education*, 507-514.
- Craig, S, Graesser, A, Sullins, J, Gholson, B (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3), 241-250.
- D'Mello, S & Graesser, AC (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145-157.
- D'Mello, S, Lehman, BA, Person, N (2010). Mining collaborative patterns in tutorial dialogues. *Journal of Educational Data Mining*, 2(1), 1-37.
- D'Mello, SK, Lehman, B, Pekrun, R, Graesser, AC (2014). Confusion can be beneficial for learning. *Learning & Instruction*, 29(1), 153-170.
- Doddannara, LS, Gowda, SM, Baker, RSJd, Gowda, SM, de Carvalho, AMJB (2013). Exploring the relationships between design, students' affective states, and disengaged behaviors within an ITS. In HC Lane, K Yacef, J Mostow, P Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 31-40). Germany: Springer Berlin Heidelberg.
- Finn, JD, Folger, J, Cox, D (1991). Measuring participation among elementary grade students. *Educational and Psychological Measurement*, 51, 393-402.

- Finn, JD & Zimmer, KS (2012). Student engagement: What is it? Why does it matter? In SL Christenson, AL Reschly, C Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 97-131). New York, NY: Springer Science+Business Media.
- Forsyth, C, Li, H, Graesser, AC (2014). Learning by communicating in natural language with conversational agents. *Current Directions in Psychological Science*, 23(5), 374-380.
- Fredricks, JA, Blumenfeld, PC, Paris, AH (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109.
- Garrison, DR & Anderson, T (2003). E-learning in the 21<sup>st</sup> century a framework for research and practice. New York, NY: *RoutledgeFalmer*.
- Goodman, L (1990). Time and learning in the special education classroom. Albany, NY: *SUNY Press*.
- Govindasamy, T (2002). Successful implementation of e-learning pedagogical considerations. *Internet and Higher Education*, 4, 287-299.
- Halpern, DF, Graesser, A, Hakel, M (2007). 25 learning principles to guide pedagogy and the design of learning environments. Washington, DC: *Association for Psychological Science Task Force on Life Long Learning at Work and at Home*.
- Harp, SF & Mayer, RE (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90, 414-434.
- Heffernan, NT & Heffernan, CL (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470-497.

- Horn, MB & Staker, H (2011). The rise of K-12 blended learning. Resource document. Innosight Institute. <http://www.christenseninstitute.org/wp-content/uploads/2013/04/The-rise-of-K-12-blended-learning.pdf>. Accessed 10 March 2015.
- Kester, L, Kirschner, PA, van Merriënboer, JJG (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology*, 31, 167-185.
- Khachatryan, GA, Romashov, AV, Khachatryan, AR, Gaudino, SJ, Khachatryan, JM, Guarian, KR, Yufa, NV (2014). Reasoning Mind Genie 2: An intelligent tutoring system as a vehicle for international transfer of instructional methods in mathematics. *International Journal of Artificial Intelligence in Education*, 24(3), 333-382.
- Koedinger, KR & Corbett, AT (2006). Cognitive tutors: Technology bringing learning science to the classroom. In RK Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 61-77). New York: Cambridge University Press.
- Lee, SW, Kelly, KE, Nyre, JE (1999). Preliminary report on the relation of students' on-task behavior with completion of school work. *Psychological Reports*, 84, 267-272.
- Liu, Z, Pataranutaporn, V, Ocumpaugh, J, Baker, RSJd (2013). Sequences of frustration and confusion, and learning. *Proceedings of the 6<sup>th</sup> International Conference on Educational Data Mining*, 114-120.
- Lloyd, JW & Loper, AB (1986). Measurement and evaluation of task-related learning behavior: Attention to task and metacognition. *School Psychology Review*, 15(3), 336-345.
- Low, R & Sweller, J (2005). The modality effect in multimedia learning. In RE Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 147-158). New York: Cambridge University Press.

- Mayer, RE & Chandler, P (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390-397.
- Mayer, RE, & Moreno, R (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-20.
- Mayer, RE & Moreno, R (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43-52.
- Mayer, RE, Heiser, J, Lonn, S (2001). Cognitive constraints on multimedia learning: Whe presenting more material results in less learning. *Journal of Educational Psychology, 93*, 187-198.
- McDonald, JH (2014). Handbook of Biological Statistics. Baltimore, MD: *Sparky House*.
- Miller, WL, Baker, RS, Labrum, MJ, Petsche, K, Wagner, AZ (2015). Automated detection of proactive remediation by teachers in Reasoning Mind classrooms. *Proceedings of the 4<sup>th</sup> Conference on Learning Analytics and Knowledge*.
- Moreno, R & Mayer, RE (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.
- Moreno, R, Mayer, RE, Spires, H, Lester, J (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*, 177-214.
- Ocumphaug, J, Baker, RS, Rodrigo, MMT (2015). Baker Rodrigo Ocumphaug Monitoring Protocol (BROMP) 2.0 Technical and Training Manual. Technical Report. New York,

- NY: *EdLab*. Manila, Philippines: *Ateneo Laboratory for the Learning Sciences*.
- <http://www.columbia.edu/~rsb2162/BROMP.pdf>. Accessed 24 March 2015
- Paquette, L, Baker, RS, Sao Pedro, MA, Gobert, JD, Rossi, L, Nakama, A, Kauffman-Rogoff, Z (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. In S Trausan-Matu, KE Boyer, M Crosby, K Panourgia (Eds.) *Intelligent Tutoring Systems* (pp. 1-10). Switzerland: Springer International Publishing.
- Pardos, ZA, Baker, RS, San Pedro, M, Gowda, SM, Gowda, SM (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, 1(1), 107-128.
- Pasolunghi, MC, Cornoldi, C, Liberto, S (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. *Memory & Cognition*, 27(5), 779-790.
- Rodrigo, MMT, Baker, RSJd, D'Mello, S, Gonzalez, MCT, Lagud, MCV, Lim, SAL, Macapanpan, AF, Pascua, SAMS, Santillano, JQ, Sugay, JO, Tep, S, Viehland, NJB (2008). Comparing learners' affect while using an intelligent tutoring systems and a simulation problem solving game. *Proceedings of the 9<sup>th</sup> International Conference on Intelligent Tutoring Systems*, 40-49.
- Sabourin, J, Rowe, JP, Mott, BW, Lester, JC (2011). When off-task is on-task: The affective role of off-task behavior in narrative-centered learning environments. In G Biswas, S Bull, J Kay, A Mitrovic (Eds.), *Artificial Intelligence in Education* (pp. 534-536). Germany: Springer Berlin Heidelberg.

- San Pedro, MOZ, Baker, RS, Bowers, AJ, Heffernan, NT (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. *Proceedings of the 6<sup>th</sup> International Conference on Educational Data Mining*, 177-184.
- San Pedro, MOZ, Baker, RSJd, Gowda, SM, Heffernan, NT (2013). Towards an understanding of affect and knowledge from student interaction with an intelligent tutoring system. In HC Lane, K Yacef, J Mostow, P Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 41-50). Germany: Springer Berlin Heidelberg.
- San Pedro, MOZ, Ocumpaugh, JL, Baker, RS, Heffernan, NT (2014). Predicting STEM and non-STEM college major enrollment from middle school interaction with mathematics educational software. *Proceedings of the 7<sup>th</sup> International Conference on Educational Data Mining*, 276-279.
- van Merriënboer, JJG, & Sweller, J (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2), 147-177.
- Waxman, HC & Houston, WR (2012). Evaluation of the 2010-2011 Reasoning Mind Program in Beaumont ISD. [http://www.reasoningmind.org/wp-content/uploads/2013/11/2010-2011\\_Beaumont\\_ISD\\_Independent\\_Evaluation.pdf](http://www.reasoningmind.org/wp-content/uploads/2013/11/2010-2011_Beaumont_ISD_Independent_Evaluation.pdf). Accessed 25 March 2015.

## Illustrations and Figures

Figure 1. Genie 2 Home Screen

Figure 2. Genie 2 Guided Study

Figure 3. Genie 3 Home Screen

Figure 4. Genie 3 Guided Study

Figure 5. Average distribution of behavior and affect, Genie 2 vs. Genie 3

Figure 6. Average distribution of behavior and affect, Genie 3 vs. traditional instruction

Figure 7. Average distribution of behavior and affect by half-hour, Genie 3 vs. traditional  
instruction

## Tables

Table 1. Improved implementation of e-learning principles in the Genie 3 platform.

<b>E-Learning Principle</b>	<b>Genie 2</b>	<b>Genie 3</b>
Personalization	“The Genie” is a limited pedagogical agent who only provides motivation, in text only, but does use a conversational style.	Virtual tutors and peers are full pedagogical agents and virtual coaches. They use conversational but polite speech, and are voiced by human narrators. Students have customizable avatars.
Modality	Entirely text-based with optional text-to-speech.	Prioritizes narrative, auditory instruction over text.
Redundancy	Audio is matched to on-screen text word for word.	Spoken narration and written text complement each other, but do not match.

Table 2. Count and percentage of observations of each behavior and affective state.

<b>BROMP Category</b>		<b>N</b>	<b>%</b>
Behavior	On Task	243	82%
	On Task Conversation	20	7%
	Off Task	31	10%
	Gaming	2	1%
Affect	Engaged Concentration	194	71%
	Boredom	27	10%
	Confusion	24	9%
	Frustration	19	7%
	Delight	9	3%

Table 3. Skewness and kurtosis of behavior and affect distributions for Study 2.

		Genie 3		Genie 2	
		Skewness (SE = 0.16)	Kurtosis (SE = 0.31)	Skewness (SE = 0.16)	Kurtosis (SE = 0.31)
Behavior	On Task - Independent	-1.98	7.10	-1.11	<b>0.60</b>
	On Task - Conversation	3.02	9.99	2.34	7.40
	On Task - Proactive Remediation	2.50	6.64	4.05	16.18
	Off Task	3.82	21.63	1.89	4.36
Affect	Engaged Concentration	-2.05	5.18	-1.09	0.97
	Boredom	2.19	5.69	1.53	2.90
	Confusion	3.96	17.50	2.34	5.98
	Frustration	15.75	248.00	12.64	167.15
	Delight	-	-	15.75	248.00

Acceptable ratios of skewness and kurtosis to SE are within  $\pm 2.58$ . Measures in boldface type are within the acceptable range.

Table 4. Skewness and kurtosis of behavior and affect distributions for Study 3.

		Genie 3		Traditional Classroom	
		Skewness (SE = 0.16)	Kurtosis (SE = 0.32)	Skewness (SE = 0.16)	Kurtosis (SE = 0.32)
Behavior	On Task - Independent	-1.67	2.71	-0.45	<b>0.10</b>
	On Task - Conversation	2.30	4.34	1.89	3.79
	On Task - Proactive Remediation	2.90	9.87	15.07	227.00
	Off Task	3.15	11.81	0.70	0.62
Affect	Engaged Concentration	-0.97	<b>0.22</b>	-0.72	<b>-0.21</b>
	Boredom	1.15	0.82	0.78	<b>0.03</b>
	Confusion	5.32	32.85	1.97	3.62
	Frustration	9.56	97.66	-	-
	Delight	-	-	6.29	41.13

Acceptable ratios of skewness and kurtosis to SE are within  $\pm 2.58$ . Measures in boldface type are within the acceptable range.