Hammill Institute on Disabilities

Learning Disability Quarterly
2014, Vol. 37(2) 87–99

DOI: 10.1177/0731948713503963
ldq.sagepub.com
SAGE

UDL in the Middle School Science Classroom: Can Video Games and Alternative Text Heighten Engagement and Learning for Students With Learning Disabilities?

Matthew T. Marino, PhD[1], Chad M. Gotch, PhD[2], Maya Israel, PhD[3], Eleazar Vasquez III, PhD[1], James D. Basham, PhD[4], and Kathleen Becht, MS[1]

Abstract

This article examined the performance of 57 students with learning disabilities (LD) from four middle schools. Students were followed over the course of a school year in their inclusive science classrooms as they alternated between the use of traditional curricular materials for some units of study and materials that were supplemented with video games and alternative print-based texts to more closely align with Universal Design for Learning (UDL) guidelines during other units. Findings indicate that video games and supplemental text were effective at providing students with multiple means of representation and expression. The UDL-aligned units led to heightened levels of student engagement. There were no significant differences on posttest scores when students with LD were compared with peers without LD. Students' performance did not indicate significant differences between UDL-aligned units and those taught using traditional curricular materials. Findings suggest a need for alternative assessments to measure learning outcomes during UDL-aligned units. Implications for practice and areas of future research are discussed.

Keywords

learning disability, science, Universal Design for Learning, video games

Students with learning disabilities (LD) face a complex array of barriers as they enter inclusive middle school science courses. These include limited instructional diversity and science teachers with inadequate knowledge of effective pedagogical practices for teaching students with LD (Mastropieri et al., 2006). In addition, Lee and Erdogan (2007) pointed out that students with LD develop negative attitudes about science when they encounter complex expository texts and other instructional materials that limit their ability to access and comprehend scientific information. As a result, only 5% of students with disabilities enter the science, technology, engineering, and mathematics (STEM) workforce even though their individual attributes often lend themselves to success in these career paths (Leddy, 2010).

Marino (2010), in a comprehensive review of literature related to students with LD and technology use in secondary science classes, noted substantive empirical evidence suggesting that secondary science curricular materials often fail to engage students with LD because complex vocabulary and phenomenological constructs were presented using inaccessible media such as expository texts. Ineffective pedagogy also contributed to students' struggles. For example, secondary science teachers often failed to accurately assess students' declarative knowledge and procedural skills at the outset of inquiry activities. In addition, many teachers utilized ineffective instructional techniques and exhibited implicit social biases related to the types of students who could achieve at a high level in the sciences. The science teachers set low expectations for students with LD. Unfortunately, students with LD often achieve at a level commensurate with that expectation. This is evidenced in the 2011 National Assessment of Educational Progress (NAEP) eighth-grade science scores, where 68% of students with disabilities scored at the below basic level

[1]University of Central Florida, Orlando, USA
[2]Washington State University, Pullman, USA
[3]University of Illinois at Urbana–Champaign, USA
[4]University of Kansas, Lawrence, USA

Corresponding Author:
Matthew T. Marino, Department of Child, Family and Community Sciences, University of Central Florida, Orlando, FL 32816, USA.
Email: Matthew.marino@ucf.edu

compared to 31% of their peers without disabilities (Aud et al., 2012). It is clear that alternative curricular materials and pedagogical practices are necessary to meet the needs of students with LD in this context.

Since Rose, Meyer, and Hitchcock's (2005) seminal text was released, Universal Design for Learning (UDL) has garnered national and international attention as a curriculum and pedagogical design framework that proactively addresses the student diversity in today's inclusive classrooms. The Center for Applied Special Technology (CAST; 2011) articulated three core principles as necessary to align curricular materials with the UDL framework in Version 2.0 of their guidelines, which call for multiple means of (a) representation, (b) action and expression, and (c) engagement. Each of these intuitive principles appears critical to enhancing learning for students with LD in STEM domains.

Over the past several years, scholars have attempted to articulate the role of technology in UDL curricular materials. For example, King-Sears (2009) noted that UDL curricula provides teachers with the opportunity to proactively integrate intelligent pedagogy with technology so that students' can choose the most appropriate medium to access specific content. Edyburn (2010) extended this notion stating:

> Why is computer technology essential for a majority of 21st-century activities outside of school but optional for helping students achieve high standards within school? . . . To suggest that the potential of UDL can be achieved without technology is simply another way to maintain the status quo. (p. 38)

Taken within the context of the secondary science curriculum, this statement is especially poignant. In fact, teachers, researchers, and funding agencies devote a continually increasing amount of time and resources toward enhancing access to science education materials using technology (U.S. Department of Education, 2010). This effort aligns with an increased emphasis on the empirical investigation of the UDL framework (Gordon, Gravel, & Schifter, 2009; Rappolt-Schlichtmann, Daley, & Rose, 2012).

Educational video games are widely available resources that provide teachers with the means to create UDL-science curricular materials (Marino, Basham, & Beecher, 2011). These games allow repeated practice opportunities where students can interact with alternative representations of complex vocabulary and phenomena. Marino and Beecher (2010) pointed out that educational video games can contribute to a UDL-science curricula and align with current special education service delivery methods by enhancing the accessibility of the content and inextricably linking social aspects of game play to increasingly intensive instructional supports that correlate to response-to-intervention tiers.

At first glance, educational video games are seductive. The National Research Council (2011) reported that an increasing body of evidence suggests that educational video games have the potential to promote critical attributes associated with scientific literacy. Marino, Israel, Beecher, and Basham (2012), in a nation-wide study of middle school students and teachers in inclusive science classrooms, reported a widespread affinity for the adoption of educational video games as part of the science curriculum. However, it remains unclear whether this enthusiasm has empirical support.

A recent review of literature by Young et al. (2012) pointed out that the high degree of variability in current video game designs contributes to inconclusive findings in efficacy studies across educational contexts. They noted that although video games can be valuable educational assets, there is often a disconnect between the efficacy of the games and their effectiveness in the classroom. Further complicating matters is the fact that many educational video games lack clearly defined learning objectives and outcomes. Therefore, it is often difficult to examine whether these games contribute to students' learning (Marino et al., 2011).

The study reported here was designed to contribute to the limited literature in this area by examining how science curricular materials that align with the UDL framework contribute to learning outcomes for students with LD. This study is a critical step toward understanding the limitations associated with using traditional paper and pencil quantitative assessments to assess the efficacy of UDL curricular materials. In addition, the study provides preliminary insights regarding which attributes of the UDL curriculum are beneficial and the pedagogical factors that are necessary for successful implementation of a UDL-science curricula.

The research was guided by the following questions:

Research Question 1: Is there a relationship between the use of video games, alternative text, and the level of engagement of students with LD in inclusive middle school science classrooms?

Research Question 2: Are there differences in performance on paper and pencil posttests when students with LD participate in video game and alternate text enhanced units compared with traditional instruction units?

Research Question 3: Are there differences in performance during UDL-enhanced units on the paper and pencil tests when students with LD are compared to students in other reading ability groups?

Research Question 4: Is there a relationship between students' with LD use of UDL scaffolds in the game (e.g., on-demand tutorials) and their performance on the paper-and-pencil tests?

Two aspects of learner variability that are tracked in every school in the United States are students' reading ability and disability status. These served as the independent variables in the current study. The dependent variables under investigation were students' performance on a

Table 1. Sample Implementation Protocol.

Topics	Game availability	Curricular materials	
		Students with LD and below basic readers	NAEP reading score at basic or higher
Organisms	No game	RC	RC
Cells	Cell Command	RC + PCI supplement + game	RC + game
Classification	No game	RC	RC
Heredity and reproduction	Crazy Plant Shop	RC + PCI supplement + game	RC + game
Evolution	No game	RC	RC
Bacteria & viruses	You Make Me Sick!	RC + PCI supplement + game	RC + game
Protists and fungi	No game	RC	RC
Plants	Reach for the Sun	RC + PCI supplement + game	RC + game
Animals	No game	RC	RC

Note. Teachers implemented the topics in the order that was prescribed by their district. All followed the ABAB methodology. LD = learning disabilities; NAEP = National Assessment of Educational Progress; RC = Regular curriculum.

paper-and-pencil posttest, students' video game play statistics, and students' reported levels of engagement during the learning activities.

The standardized measure of students' reading ability in this study was the NAEP reading achievement level. In this study, students' NAEP scores were reported as below basic, basic, proficient, and advanced. Each reading category is accompanied by a grade-level description of the students' skill level and accompanying assessment cut score. While a full discussion of the NAEP reading ability levels is outside the scope of this manuscript, interested readers are encouraged to learn more at http://nces.ed.gov/nationsreportcard/reading/.

The special education teacher reported students' LD status. Students were considered LD if they were on an Individualized Education Program (IEP) with a primary classification of LD during the time of the study. Students on an IEP ranged across reading ability levels. Therefore, they were represented in two ways, as a student with LD and as a member of a reading ability group.

Method

This study employed a mixed-methods design (Creswell & Plano Clark, 2007). Once the quantitative and qualitative data sources were analyzed individually, they were examined collectively to draw conclusions about the efficacy of the curriculum. Participants in the study followed an ABAB model, with A representing non-UDL units and B representing UDL-aligned units. Table 1 shows a demonstrative model of the curriculum students followed.

As this 3-year project progressed, several factors emerged that shaped the research design. First, after working with teachers in 14 states, it was clear that there was tremendous variability in the scope and sequence at which middle school science content was taught. For example, in some school districts, students' progress was dictated by the calendar (e.g., September 14, students will start Unit 2), while in other districts progress was left completely at the discretion of the teacher. There was also variation in grade level at which the content was taught (e.g., life science in fifth grade vs. seventh grade).

In this study, we included a range of participants reflective of the diversity in today's inclusive classrooms. The original intention was to provide teachers with a series of predetermined UDL-aligned content specific investigations. However, teachers repeatedly indicated that they did not want a prescribed science curriculum. Instead, they wanted the flexibility to incorporate resources for their students while keeping the materials they were familiar with. Furthermore, many told us they would not be able to use a researcher-developed UDL curriculum because of prolonged state and district level adoption processes. Therefore, the researchers opted to supplement teachers' existing curricular materials during the UDL-aligned units with the educational video games from Filament Games and leveled print-based science books from PCI education for students with LD and those with below basic reading level scores.

Teachers used their regular curricular materials during the non-UDL units. Each class participated in at least two UDL-aligned units and two non-UDL units. Researchers involved in the study are cognizant of the fact that this retrofit approach is not consistent with the vision of UDL. However, this approach is consistent with the way the science teachers in the study plan their units and lessons on a daily basis.

Participants

One fifth-grade and four seventh-grade female science teachers from four states in the Pacific Northwest and Midwest were selected to participate in the study. The teachers were selected from an initial convenience sample of 150 teachers. Teachers were selected based on the

Table 2. Sample Characteristics.

Characteristic	% of students
Grade level	
5	16
7	84
NAEP reading level	
Below basic	23
Basic	13
Proficient	41
Advanced	23
LD	17

Note. NAEP = National Assessment of Educational Progress; LD = learning disabilities.

alignment of their curriculum scope and sequence with the study protocols, technology infrastructure, administrative consent, and student demographics related to the number of students with LD in their classrooms. Each science teacher had a minimum of 8 years teaching experience in inclusive science classrooms. All were highly recommended as exemplary teachers by their principals. Of the five teachers selected, three taught in suburban schools, one in an urban school, and one in a rural school. Student to teacher ratios averaged 19 to 1. Special education and reading teachers who worked with these science teachers provided information about the students' disability qualification and NAEP reading ability scores. The special education teachers and reading teachers were not participants in the study.

Student-level data across units were gathered from 341 students, both with and without disabilities, whose teachers consented to participate in this study. Females composed 49% of the participants in the study. Fifty-seven students with a primary IEP classification as LD participated in the study. The students ranged from 10 to 14 years of age. Students' reading scores were standardized across states using NAEP cut-level comparison charts. Student distribution by grade, NAEP reading ability level (i.e., below basic, basic, proficient, advanced), and IEP status are presented in Table 2. During preintervention surveys that were designed to assess familiarity with technology, 94% of students with LD reported that they used computers and mobile devices on a daily basis. Eighty-seven percent of the students were White. Free or reduced lunch rates for the students ranged from 18% to 67% with a mean of 56%. Less than 2% of the students were English language learners.

Materials

Paper-and-pencil pre-/posttest. The researchers developed 10 unit pre-/posttests. Pretests and posttests were identical for each unit. Each test contained between 20 and 24 items. The tests were designed to be difficult for middle school students to prevent a ceiling effect and severe negative skew that occurred during Phase I pilot testing. Tests were developed using a Delphi process that included science teachers, a science professor, and a senior representative of the National Science Teachers Association. All participants in this process were asked to review the test items, provide feedback, and rate the level of difficulty. Modifications were made until the tests had formats and items at each level of Bloom's Taxonomy. The format of the tests was consistent across topic areas in order keep the difficulty level as identical as possible given differences in content. Item-level analyses indicated internal consistency reliabilities of the posttests ranged from .77 to .85. Reducing the number of items across the original tests resulted in a secondary set of 10-item tests. Item reductions were derived using the same process described previously. The 10-item tests were developed to isolate the effects of the material presented in the video games during UDL-enhanced units. Items in non-UDL units were selected based on agreement among the test developers that the items were equivalent in difficulty level to the UDL-enhanced questions.

Video games. Students who participated in the study played at least two of the suite of four middle school life science games developed by Filament Games in collaboration with the first author. The games and topics were Cell Command (cell anatomy and functions); Crazy Plant Shop (genes and inheritance); You Make Me Sick! (bacteria and viruses); and Reach for the Sun (photosynthesis and plant life cycle). Although a detailed discussion of the UDL development cycle for each game is outside the scope of this article, an overview of one of the games, You Make Me Sick!, is provided to illustrate how they are intended to provide multiple means of representation, action, expression, and engagement. Each of the games underwent a similar development cycle that aligned with CAST UDL guidelines Version 2.0 (2011). UDL checkpoints from these guidelines are noted parenthetically in the subsequent discussion. Interested readers can gain further information at http://www.udlcenter.org/aboutudl/udlguidelines. A detailed analysis of how UDL checkpoints were included in one section of the game, the Pathogen Design Studio, is included in Table 3.

You Make Me Sick! was designed to teach students about common bacterial and viral transmission pathways and the benefits of healthy lifestyle choices. The game challenged players to make a virtual host (i.e., a person to infect), as sick as possible (UDL Checkpoint 8.1). Students first analyzed the host's attributes (e.g., overweight, smokes, is on an antibiotic) and then decided which pathogen (i.e., virus or bacteria) would be most effective. Students could choose from existing pathogens such as salmonella or rabies, or they could design their own from a detailed menu of pathogen properties (UDL Checkpoints 7.1 and 8.2). This provided students with varied levels of expertise equal

Table 3. Mapping Game Features in You Make Me Sick! to UDL Checkpoint Version 2.0.

Pathogen design studio features	Checkpoint	Description
Provide pictorial and verbal information about the intended host	5.1	Use multiple media for communication
Include well-known pathogens and the ability to engineer new pathogens	5.2	Use multiple tools for construction and composition
	5.3	Build fluencies with graduated levels of support for practice and performance
Structure the pathogen design process while allowing students to choose their own navigation path	6.1	Guide appropriate goal setting
	6.3	Facilitate mapping information and resources
Offer and track students' use of an advanced tutorial to scaffold VLE navigation proficiency and reduce cognitive load	6.2	Support planning and strategy development
	6.4	Enhance capacity for monitoring progress

access to the content. More advanced users could engineer their own pathogen, while less experienced players could choose an existing pathogen, thus providing multiple means of action and expression. Players then decided how to most efficiently infect the host and placed their pathogen in a virtual room (UDL Checkpoint 8.3). Students who struggled with this aspect of the game received guidance through an advanced tutorial (UDL Checkpoints 6.2 and 8.4). When the host came in contact with the pathogen, the player was virtually transported inside the host's body (UDL Checkpoints 3.1 and3.2). Over the next several minutes of game play, students transitioned from the macro environment outside the host's body to the cellular level, where they attempted to evade white blood cells and complete the infection process (see Figure 1). This journey through the host's body provided multiple means of representation, which was further reinforced with a virtual dictionary that included a read aloud option and page numbered links to the supplemental PCI science materials (UDL Checkpoints 1.1, 1.2, 1.3, 2.1, and 2.5).

The game included explicit instruction using animated tutorials that students could access at any time during game play (UDL Checkpoint 3.3). Students could also alter the sound and in some of the games the appearance of their avatar (UDL Checkpoint 7.1). The use of each of these features was tracked using individual login IDs. At the conclusion of the game, players received a certificate of virulence, which provided a score that teachers could use as an alternative assessment (UDL Checkpoints 6.4 and 9.3).

PCI science curricular materials. Students who were on an IEP and those who were below basic readers were provided with a supplementary print-based text (Lindsay & Cordova, 2007) from PCI Education that was used in concert with their general education science curriculum materials. This traditional print version of the text (i.e., no digital version was available during this study) presented alternative representations of science vocabulary and concepts in a format that was written and illustrated specifically for struggling readers. The average Flesch–Kincaid readability level across units was established at fifth grade. Students were encouraged to use whichever text source (i.e., their regular textbook or the PCI textbook) they preferred. However, all students chose the PCI materials over their traditional texts. The PCI text was meant to further align the teachers' science curriculum with the UDL framework by providing students with a choice of learning materials. It was not meant to be an add-on or create additional work for the students.

Student characteristics and attendance form. General education science teachers, with assistance from their special education and reading teacher colleagues, completed a student characteristics profile for each of their students. The profile included each students' (a) special education status (i.e., whether or not the student was on an IEP and if so, their disability classification) as well as any accommodations and/or modifications they received and (b) NAEP reading ability level, which was based on their most recent standardized reading test score.

Student postintervention focus group interviews. Teachers were asked to identify several students with and without disabilities to participate in postintervention focus group interviews. Interview protocols were developed collaboratively by the lead author and members of the development team at Filament Games. Interviews were conducted in person, over the phone, or via video conferencing software. Twelve students with a primary classification of LD and seven students who were not on an IEP provided assent and parental consent prior to participating in the interviews. Students were asked to share their thoughts about their perceptions of science, the alternative PCI text, the video games, and the assessments. Interviews lasted approximately 30 min. Interviews were conducted with students as soon as time permitted after the final UDL unit was finished. All interviews were conducted within 2 weeks of finishing the curriculum.

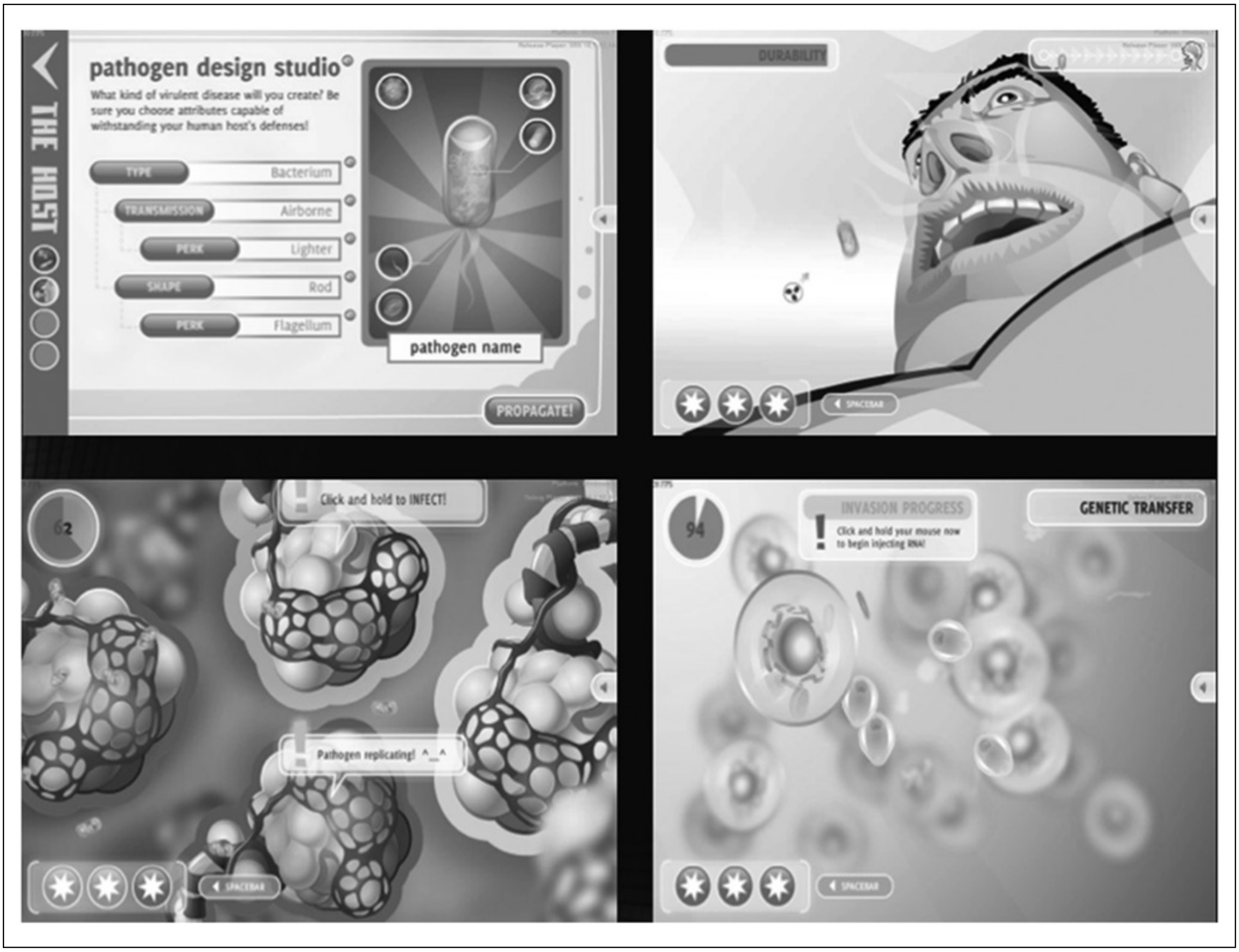

Figure 1. Multiple means of engagement and representation in the game *You Make Me Sick!*

Procedure

Teacher professional development. Prior to implementing the study, teachers participated in professional development sessions with members of the research and implementation team. The number and length of professional development sessions varied from 1 hr to 5 hr based on the self-reported technology experience and comfort level of the teacher. Professional development included how to access and play the video games, how to implement the games in the classroom, and how to administer and submit the tests and surveys using Scantron (i.e., bubble) sheets.

Classroom implementation. Teachers were assigned unique student ID numbers for each class period, which they assigned to each of their students. These ID numbers allowed researchers to track each student's game play and paper and pencil test and survey data. When the students first logged into the game, they were prompted to enter this number. They then set up user names and passwords for future access the games.

Students completed the preintervention survey at the beginning of the first unit, prior to any instruction. During the first UDL-enhanced unit, students completed the initial Internet login. Each student had his or her own computer. Login was followed by a computer-based introduction to the first game, which lasted approximately 10 min. This experience provided the backstory for the game and taught students essential game play mechanics, such as how to navigate in the game environment. Students spent the final day of instruction for each UDL-enhanced unit playing the video game, which was designed to reinforce key science concepts and vocabulary from the unit while stimulating inquiry skills and knowledge transfer. On the final day, students and teachers completed a postintervention survey. Each of the materials and activities within the game-enhanced and nonenhance units was highly varied. This deviation occurred at both the teacher and classroom level, where teachers altered their approaches based on the individual needs of their students. The inclusion of a series of game-enhanced and non-game-enhanced units allowed us to balance the number of experiments, ancillary supports, and hands-on activities across both conditions.

Fidelity of implementation. Fidelity of implementation was addressed in several ways. First, teachers received professional development and had access to an implementation guide. Second, the researchers obtained teacher self-reports regarding the students' game play. Finally, the lead researcher conducted 40 hr of observations in select classrooms during each phase of the study to determine how the teachers' instructional approach differed across conditions (i.e., game-enhanced vs. no game). In addition, the lead researcher conducted periodic Skype interviews with the teachers during the intervention processes. Students and teachers completed pre- and postintervention surveys using Scantron sheets. These data were combined with game play statistics from the Filament Games server on an excel spreadsheet, which was converted for statistical analysis using the SPSS statistical software Version 2.0 (2012).

Data Analysis

Research Question 1: Level of engagement during UDL units. Follow-up student semistructured focus group interviews were transcribed, coded, and analyzed using a basic interpretive qualitative methodology (Merriam, 2002). Categories and themes were determined through a constant comparative method (Glaser & Strauss, 1967) using data from transcribed interviews, game play statistics, and survey responses. The data were grouped first by regularities and then by irregularities into tentative categories and subcategories. To increase credibility of this data analysis, careful consideration was paid to both validity and reliability. To increase reliability, transcription included pauses, facial expressions when possible, and tone of voice. Following the initial analysis, a secondary analysis took place in a select number of classrooms. During this member checking process, the lead researcher presented initial results and asked teachers and students for help in clarifying interpretation of the results. Interviewees were presented with their statements and with resulting categories. They were asked whether their comments were interpreted correctly and if their comments fit appropriately within the emerging themes. These member checks served to enhance accuracy, credibility, and validity.

Research Questions 2 and 3: Repeated measures ANOVA. To examine test performance of students with LD and students across reading ability groups, a repeated measures factorial analysis of variance (ANOVA) was performed. Condition (two levels, enhanced and traditional instruction) and time (two levels, pretest and posttest) were cast as within-subject factors. Students' unit tests were aggregated within each combination of within-subject factors. This design resulted in four percentage-correct scores for each student. Reading ability (four levels: below basic, basic, proficient, and advanced) and learning disability status (two levels: LD and non-LD) functioned as between-subjects factors. An alpha level of .05 was set for judging statistical significance. Effect size (r) calculations aided interpretation of the statistical test results. Given the novelty of the study's investigation, we relied on Cohen's (1992) generic guidelines for judging r as small (.10), medium (.30), or large (.50). Parallel analyses were run, one with full sets of items on the unit tests, reflecting the full domain of the unit topic (e.g., cell biology), and one with reduced-item sets (10 items per unit) that reflected the aspects of the unit topic specifically addressed by the enhanced instruction. Conducting these parallel analyses allowed for insight into whether potential gains from the enhanced instruction were aligned only to the video games, perhaps reflecting a kind of practice effect, or were realized in a broader understanding of the topic.

Research Question 4: Multiple regressions. To examine the relationships between student performance on unit posttests and use of UDL scaffolds in the game, a series of multiple regressions were conducted. The unit posttest score (% of questions answered correctly) served as the dependent variable. Independent variables were entered in blocks to control for family-wise error. The first block contained the unit pretest score and LD status. These variables were included in the analyses to control for individual differences (e.g., prior knowledge) that could influence unit posttest scores. A sum of all uses of game UDL features—dictionary accesses, voice-over accesses, tutorial accesses, audio help toggles, and volume adjustments—was entered as a second block. The third block entered contained the number of levels completed by the student. The fourth and final block contained two interaction terms—learning disability status with use of UDL features and learning disability status with level completion. These interaction terms allowed for isolating the relationship of game play to test performance specifically for students with LD. Regression models were evaluated by the total variance captured (R^2) in posttest scores. Blocks of independent variables were evaluated for the change in R^2 resulting from their inclusion in the model and the associated improvement in model fit, as measured by the F statistic ($\alpha = .01$). As with the repeated measures ANOVA, parallel analyses using full and reduced-item sets took place.

Results

It should be noted that the findings reported here include aggregated statistics compiled from four of the beta versions of the video games. Meaning, the games were playable, but the final design, artwork, physical, and cognitive accessibility features had not been fully realized. The results reported here were used to inform these accessibility features in the gold builds of the games and provide future methodological considerations for efficacy assessments of UDL-aligned curricular materials.

Research Question 1

This question pertained to level of engagement of students with LD during UDL units. The results are disaggregated by themes that were identified during the interviews.

Alternative science text. Students with LD were asked to describe their thoughts on the supplemental PCI text to help determine whether the text increased the accessibility of the content and whether an increase in accessibility could lead to heightened engagement. Students indicated that they preferred the PCI text but did not enjoy reading as a way to learn about science. Rather, most students stated that they preferred to access scientific information using technology. For example, a seventh-grade female with LD stated, "I liked it a lot better than our other book." A male classmate added, "Yeah, I liked it too. The pictures were good. Not as good as the Internet though. I don't know why they make us use books in school." When asked to describe how he used the book the student responded, "We used it in class once in a while and then in study hall. I don't read at home unless my dad makes me. Most of the time I don't bring my book home." Another student stated, "I have a hard time with the words in science class. Sometimes it takes me forever to read our other book. I usually just give up." Her classmate added,

> Right on. I hate to read. I'd much rather do something with my hands . . . like build something or take something apart. Sometimes I can't put it back together . . . but I learn better. The other book (PCI) was better but I still like doing something better (than reading).

Similarly, another student explained, "I would prefer not to have a book," and her classmate added, "Me either. I can find almost everything I need on YouTube. Why we are still using books is beyond me."

Video games. Students expressed a clear affinity toward the video games and reported collaborative engagement during game play. For example, one female student noted, "I liked the games, especially the one where you made the guy sick. That was rad." When asked what she liked about it, she indicated, "You got to figure out what would make him sick . . . and then you had to blow it in his mouth and you could see his lungs and stuff." Her male classmate noted, "That part was hard, trying to get it in his mouth and everyone was yelling at the game and at each other and asking each other what they did to beat the game."

Interviewer: "What was your teacher doing?"

Student: "He was yelling with us."

Interviewer: "Does he do that often?"

Student: "No. [the vice principal] came in and told us to keep it down."

In a separate interview, the students described their experiences in class when playing the games. One student stated,

> We did them (played the games) by ourselves, but they would be better in teams . . . We did the games by ourselves but we kept talking to each other and that helped us figure out the way to do it.

Another male student stated,

> I liked the action parts of the game. Like when I had to spray the cheese with my virus. That was cool . . . and then that fat guy he just sucked it down like I knew he would. Bam. He's sick.

A male student from a different class stated:

> We were all yelling at each other about the game . . . like when I infected that dude with Salmonella. I was like "Oh boy, you're in for the hurt now," and my boy was like "Watch what my rabies is doin to ya sucker!"

When asked to describe her experiences with the game Crazy Plant Shop, one student responded, "That was my favorite . . . It was cool to see how genetics actually happened in a plant shop. We did a project after the game on genetically modified crops so we got to see how people really mess with plants." Many of the female students with LD reported that Crazy Plant Shop was their favorite. For example, a female student stated:

> I just liked that you had customers and they . . . asked you for stuff and then you got to figure out a way to make it for them. It is a lot like the store where my mom works . . . so I guess it seemed real.

The connection between virtual world activities in the games and students' personal lives emerged as another theme. Students spoke about how they shared the games with their families outside school. One student noted, "My mom was saying, 'Now you're playing video games in school? What's this world coming too?'" Another said, "I played with my dad. He was excited about the bacteria

Table 4. Average Percentage Correct Scores by Factor.

	Full-item sets			Reduced-item sets		
	Time			Time		
Factor	Pretest (%)	Posttest (%)	Overall (%)	Pretest (%)	Posttest (%)	Overall (%)
Condition						
UDL-aligned	35.1	50.6	42.9	18.1	36.2	27.1
Traditional	42.0	60.2	51.1	26.8	64.3	45.6
LD status						
LD	36.3	58.4	47.3	6.6	42.9	24.8
Non-LD	39.7	53.9	46.8	30.4	53.9	42.1
Reading ability						
Below basic	34.1	44.7	39.4	13.9	37.7	25.8
Basic	36.2	48.8	42.5	22.8	42.4	32.6
Proficient	37.6	52.5	45.0	25.7	51.8	38.7
Advanced	44.7	70.9	57.8	29.2	65.9	47.6
Overall	38.6	55.4		22.5	50.2	

Note. UDL = Universal Design for Learning; LD = learning disabilities.

game." A final student said, "I was showing my older brother and he was like . . . that is so cool we didn't have that when I was in middle school." Although the students expressed differences in how their families reacted to gaming in school, most were positive.

Paper-and-pencil assessments. A vast majority of the students expressed frustration about taking the tests. For example, one student stated, "The tests were hard. I'm just not good at tests and I get all nervous and then I don't know what to do." Another student who was asked whether he thought his score on the end of the unit test would be an accurate indication of what he really knew responded:

> I don't think so. I don't know why we have to take all these tests anyway. When I need to put together something, like the other day my mom got a new vacuum . . . and she was reading the directions for like an hour and couldn't put it together and I just looked at it and figured out how to do it in 10 minutes.

Another student stated:

> The problem with those tests is that all the words sound so much alike in my head. First I think I know the answer, then I think its something else, then I just get confused. Usually I give up when that happens.

Students were asked what they thought about using video game play statistics as an alternate means of assessment instead of a multiple choice test. One student said, "I don't know how fair it would be. I mean how could you say that a game and a test are equal?" A majority of students with LD stated that they would rather play a game than take a test. One male student with LD pointed out, "I tried a lot harder on the game than I do on the test . . . I liked the game."

Collaborative discourse as a preferred way of learning was another theme. A male student stated:

> I think talking about the games helped me more than the games. Does that make sense? Like when [Teacher] explained how the game was like this and here it was in the book and stuff and then she helped us understand that they went together.

The students made several suggestions when asked how to improve the games so that they would be more engaging. First, many students reported that they wanted to collaborate during the game. One student noted, "If you had a video chat box like Skype and could talk to other people all over the world that would be great." Another student added, "Oh yea. That would be cool. Like if we had a problem we could phone a friend who was a scientist." Other students reported that the use of cognitive tools in the game (e.g., dictionary, voice-over access tutorials) was beneficial. One female student explained, "I liked the dictionary in the game. Not many of us used it I don't think, but it helped me."

Research Questions 2 and 3

These questions covered performance across conditions and student characteristics. In the first analysis, using the full item sets, significant main effects were found for the within-subject factors time and condition. For time, posttest scores (Table 4) were significantly higher than pretest scores, at the aggregate, $F(1, 335) = 42.50$, $p < .001$, $r = .34$. For condition, average percentage correct scores, across pretests

Table 5. Model Summaries for the Prediction of Unit Posttest Scores by Game Play Characteristics.

				Full-item sets			Reduced-item sets		
Topic	Game	*n*	Model	R^2	ΔR^2	*p*	R^2	ΔR^2	*p*
Cells	Cell Command	241	1	.26	—	—	.09	—	—
			2	.27	.01	.09	.09	< .01	.50
			3	.28	.01	.05	.10	.01	.29
			4	.29	.01	.34	.11	.01	.30
Heredity and reproduction	Crazy Plant Shop	108	1	.17	—	—	.05	—	—
			2	.21	.04	.02	.06	.01	.40
			3	.21	< .01	.60	.07	.01	.29
			4	.22	.01	.89	.07	< .01	.99
Bacteria and viruses	You Make Me Sick!	214	1	.38	—	—	.21	—	—
			2	.38	< .01	.26	.21	< .01	.31
			3	.38	< .01	.97	.21	< .01	.56
			4	.38	< .01	.87	.22	.01	.45
Plants	Reach for the Sun	48	1	.59	—	—	.09	—	—
			2	.61	.02	.15	.10	.01	.72
			3	.63	.02	.19	.11	.01	.41
			4	.63	< .01	.71	.13	.02	.61

Note. Predictors for Model 1: unit pretest, LD status; Model 2: unit pretest, LD status, use of UDL features; Model 3: unit pretest, LD status, use of UDL features, levels completed; Model 4: unit pretest, LD status, use of UDL features, levels completed, LD status by use of UDL features (interaction), LD status by Levels completed (interaction).

and posttests, were significantly higher, $F(1, 335) = 13.16$, $p < .001$, $r = .19$, for units without game enhancement than for units with game enhancement. Looking across students, an overall effect was found for reading ability, $F(3, 335) = 6.86$, $p < .001$, $r = .14$. Significant differences were found between students at the advanced level and students at all other levels. In addition, students at the below basic level performed significantly worse than students at the proficient level. No significant differences in performance were observed between students at the below basic and basic levels or between students at the basic and proficient levels. Finally, there was no significant difference in percent-correct scores between those with and without LD ($p = .70$).

The interaction between the condition and time variables was found to be nonsignificant ($p = .29$, $r = .06$), indicating that there was no difference in improvements from pretest to posttest between units with traditional curricular instruction and those with UDL-aligned instruction at the aggregate. Furthermore, there was no relationship between change from pretest to posttest across condition and either LD status ($p = .08$) or reading ability ($p = .16$). That is, the three-way interactions of condition, time, and LD status and condition, time, and reading ability were nonsignificant.

Analysis with the reduced-item sets revealed the same pattern of statistical significance. Significant main effects were found for time, $F(1, 335) = 13.11$, $p < .001$, $r = .19$; condition, $F(1, 335) = 5.80$, $p = .02$, $r = .13$; and reading ability, $F(3, 335) = 3.67$, $p = .01$, $r = .10$. The effects in the model for LD status ($p = .14$), and the three-way interactions of condition, time, and LD status ($p = .15$) as well as condition, time, and reading ability ($p = .36$) remained nonsignificant.

Research Question 4

This question was about game play characteristics and test performance. Table 5 displays model summaries for the multiple regression analyses. Across the units analyzed, student pretest scores based on the full item sets along with LD status accounted for a low-to-moderate amount of the variation in their posttest scores ($R^2 = .17–.59$). Neither the block of variables capturing use of UDL features nor the block capturing level completion accounted for substantially more variance in scores on top of those accounted for by the pretest scores. Improvements in R^2 ranged from .01 to .04 and no model additions rendered statistically significant improvements to the model. Finally, LD status did not significantly interact with use of UDL features or level completion to produce changes in unit posttest scores. Substantially less predictive ability was found with the reduced-item sets. Full models were able to explain only between .07 and .22 of the variance.

Discussion

Similar to the findings of Young and colleagues (2012), this study found mixed results related to the game-enhanced UDL curricular materials. A preponderance of qualitative evidence supported the notion that students with LD were highly engaged during the UDL-science units. This finding was similar to previous studies (Marino, 2009; Marino et al., 2012). Students described talking with their peers, yelling and cheering each other on in class, and sharing their game experience with family. They made connections between the virtual worlds in the games and their experiences in class. In addition, they reported gaining an in-depth understanding as they interacted with scientific content in novel ways.

The quantitative analysis, when viewed independently from the qualitative data, appears to support the notion that the UDL units were not beneficial to students with LD. While posttest scores in game-enhanced instructional units showed a sizable improvement relative to pretest scores, even more improvement was demonstrated in units with only traditional instruction. It would be reasonable to argue that specific subgroups might benefit more from enhanced instruction, but have their gains washed out in the focus on the aggregate; however, this was not the case. Neither reading ability group nor learning disability status demonstrated a significant interaction across time and condition. Students with LD did show greater gains in test performance, but not to a statistically significant level.

The finding that the reduced-item sets did not uncover any points of significance suggests the analysis was not hampered by a measurement sensitivity issue. If knowledge gains were to occur, they should at least have been captured by the reduced-item tests, which were directly related to game play. The lack of significant findings of interest in either the ANOVA or the regression analyses, as a whole, suggests that game enhancement is doing nothing to improve student topical knowledge in these areas. This finding may indicate that game enhancement is not necessary across all of the topics included in typical middle school classrooms. Additional analysis of specific topics and the benefits of game enhancement is warranted.

Why would students and teachers react so positively to the UDL units if their test scores were not improving? It is possible sample effects influenced the results. The students in traditional instruction settings scored 7% to 9% points higher on pretests than students during the UDL units. There could also be a second-level variable casting an influence. Our sample consisted of students nested within classrooms. Within these classrooms an array of instructional quality and implementation issues could exist. Unfortunately, we did not have the power to conduct our analyses in a multilevel framework to address such issues.

Another consideration is that teachers' test review sessions, which occurred on the last day before the test during the traditional instruction units, had more of a positive effect on students' posttest performance than the games. Teachers taught to the test during the review sessions. In contrast, the video game sessions focused more on interactive problem solving than appropriate ways to respond to a test question. It is likely that the teachers' review was a more effective means to bolster performance on the paper and pencil test than the games, despite the fact that the students reported appreciating the content and gaining a more thorough understanding of the concepts during game play.

Study Limitations

This study included a mean intervention instructional time of 800 min, with approximately 100 min of time playing video games. This is reflective of approximately 14 days of classroom instruction, which is far less than the 9 to 12 weeks Gersten and Edyburn (2007) advocate for during intervention research. Unfortunately, this follows many school district pace guidelines. Some teachers reported that this might have affected the results of the study. This is especially salient given the fact that students were playing beta versions of the games.

Implications for Practice

Students' comments about their own test performance proved extremely helpful during the interpretation of these mixed results. Recall the young man who stated,

> The problem with those tests is that all the words sound so much alike in my head. First I think I know the answer, then I think its something else, then I just get confused. Usually I give up when that happens.

All of these middle school students with LD were highly cognizant of the fact that they were not proficient when demonstrating their knowledge on traditional assessments. This is a problem that must be addressed given that all schools in the United States will require successful completion of a high stakes science assessment to graduate from high school in the near future.

UDL provides curriculum developers and teachers with guidelines for designing and implementing instruction in a flexible manner that meets the needs of diverse learners (Rose et al., 2005). Students in this study reported a clear appreciation for curricular materials that meet students' preferences and learning needs. These UDL-aligned technologies should be included to the greatest extent possible. The study supported the notion that UDL-aligned curricula that incorporates video games can increase knowledge transfer between virtual and classroom learning. In addition, the games promoted collaborative learning and engagement. Educators should be mindful that the games used in this study had clearly articulated educational objectives that aligned with national benchmarks. Many games lack such attributes (Young et al., 2012).

Students overwhelmingly indicated that they appreciated the availability of options not typically included in traditional science instruction and textbooks within both the PCI curriculum (such as vocabulary supports and detailed graphics that illustrate critical concepts) and the video games (such as a dictionary, voice-over access, tutorials, and audio help toggles). Although the quantitative results did not indicate significant score differences for students who made use of these UDL-based gaming features, this may have been due to the limited time in game play, with only one full class per unit devoted to the game. Qualitative results pointed to the efficacy of these features. Students indicated that they liked having access to content in a more flexible and accessible manner. In addition to the UDL-based features that the students valued in the gaming environment, they also indicated a desire to include more collaboration during game play. Educators should consider discussion as a central component during UDL units.

Assessment is a critical component of the teaching and learning cycle. Research indicates that a focus on UDL principles in standards, instruction, and assessment can result in enhanced accessibility for expanded groups of users (Thompson, Johnstone, Anderson, & Miller, 2005). In the context of the current study, traditional paper-based assessments did not yield significant differences between UDL-aligned and traditional environments. However, more meaningful assessment results may have emerged using alternative assessment methods that correlate with the qualitative data. For example, modeling methods can capture different dimensions of student responses and have the ability to dynamically adapt the assessment instrument to the individual ability of specific students (Timms et al., 2012). Other assessment options include learning progressions in science, learning trajectories in mathematics, developmental continuums in reading, or learning maps. Educators are encouraged to identify a diverse range of highly correlated assessments during the curriculum development and implementation cycle.

The current study used video game click trails to identify which tools in the game were most used. While useful, this approach fails to identify the nuanced decision process players experience during a video game. More complex data collection systems, algorithms, and analytics that link specific click choices to educational objectives will yield more robust analysis in future educational video games. This will include game play prediction, remediation, and dynamic scripting, which identifies player's choices and alters the game to meet their specific educational needs. In addition, previous studies (e.g., Marino, 2009) noted that students with LDs often require explicit instruction and prompting to use technology-based tools to their potential.

These data collection systems have been proposed for use in both large-scale and classroom assessments. In both cases, they may provide more detailed information about student thinking than traditional paper-based models of assessment. This detailed information is particularly important in the classroom, where it can be used as the first step in a formative assessment process, to impact instructional decisions and provide feedback to students, ultimately improving student learning (Alonzo & Steedle, 2009).

Authors' Note

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the U.S. Department of Education.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through the Small Business Innovation Research (SBIR) program contract ED-IES-10-C-0023 to Filament Games. This material is also based upon work supported by the National Science Foundation under Grant IIP-1046229.

References

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education*, *93*, 389–421. doi:10.1002/sce.20303

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., . . .Zhang, J. (2012). *The condition of education 2012* (NCES 2012-045). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Center for Applied Special Technology. (2011). *UDL guidelines–version 2.0*. Retrieved from http://www.udlcenter.org/aboutudl/udlguidelines

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: SAGE.

Edyburn, D. L. (2010). Would you recognize universal design for learning if you saw it? Ten propositions for new directions for the second decade of UDL. *Learning Disability Quarterly*, *33*, 33–41.

Gersten, R., & Edyburn, D. (2007). Defining quality indicators for group designs in special education technology research. *Journal of Special Education Technology*, *22*(3), 3–18.

Glaser, B., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York, NY: Aldine.

Gordon, D. T., Gravel, J. W., & Schifter, L. A. (2009). *A policy reader in universal design for learning*. Boston, MA: Harvard Education Press.

King-Sears, M. (2009). Universal design for learning: Technology and pedagogy. *Learning Disability Quarterly*, *32*, 199–201.

Leddy, M. H. (2010). Technology to advance high school and undergraduate students in science, technology, engineering, and mathematics. *Journal of Special Education Technology*, *25*(3), 3–8.

Lee, M., & Erdogan, I. (2007). The effect of science-technology-society teaching on students' attitudes toward science and certain aspects of creativity. *International Journal of Science Education*, *11*, 1315–1327.

Lindsay, K., & Cordova, K. S. (2007). *Life science*. San Antonio, TX: PCI Education.

Marino, M. T. (2009). Understanding how adolescents with reading difficulties utilize technology-based tools. *Exceptionality*, *17*, 88-102.

Marino, M. T. (2010). Defining a technology research agenda for elementary and secondary students with learning and other high incidence disabilities in inclusive science classrooms. *Journal of Special Education Technology*, *25*(1), 1–28.

Marino, M. T., Basham, J. D., & Beecher, C. C. (2011). Using video games as an alternative science assessment for students with disabilities and at-risk learners. *Science Scope*, *34*(5), 36–41.

Marino, M. T., & Beecher, C. C. (2010). Conceptualizing RTI in 21st century secondary science classrooms: Video games' potential to provide tiered support and progress monitoring for students with learning disabilities. *Learning Disability Quarterly*, *33*(4), 299–311.

Marino, M. T., Israel, M., Beecher, C. C., & Basham, J. D. (2012). Students' and teachers' perceptions of using video games to enhance science instruction. *Journal of Science Education and Technology*. Advance online publication. doi:10.1007/s10956-012-9421-9

Mastropieri, M. A., Scruggs, T. E., Norland, J. J., Berkely, S., McDuffie, K., Tornquist, E. H., & Berkeley, S. (2006). Differentiated curriculum enhancement in inclusive middle school science: Effects on classroom and high-stakes tests. *The Journal of Special Education*, *40*, 130–137.

Merriam, S. B. (2002). *Qualitative research in practice*. San Francisco, CA: Jossey-Bass.

National Research Council. (2011). *Learning science through computer games and simulations*. Committee on Science Learning: Computer Games, Simulations, and Education, M. A. Honey & M. L. Hilton, Eds., Board on Science Education, Division of Behavioral and Social Sciences and Education.Washington, DC: National Academies Press.

Rappolt-Schlichtmann, G., Daley, S. G., & Rose, T. L. (2012). *A research reader in universal design for learning*. Boston, MA: Harvard Education Press.

Rose, D. H., Meyer, A., & Hitchcock, C. (2005). *The universally designed classroom. Accessible curriculum and digital technologies*. Boston, MA: Harvard Education Press.

Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. (2005). *Considerations for the development and review of universally designed assessments* (Technical Report 42). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Timms, M., Clements, D. H., Gobert, J., Ketelhut, D. J., Lester, J., Reese, D. D., & Wiebe, E. (2012). *New measurement paradigms*. Newton, MA: Community for Advancing Discovery Research in Education.

U.S. Department of Education. (2010). *Transforming American education learning powered by technology: National education technology plan 2010*. Washington, DC: Author.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., . . .Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, *82*, 61–89.