Testing the impact of a pre-instructional digital game on middle-grade students' understanding of photosynthesis

Katie McMillan Culp, Wendy Martin, Margaret Clements, Ashley Lewis Presser

Center For Children & Technology/Education Development Center, Inc.

96 Morton Street, 7th floor, New York, NY 10014

October, 2013

For more information,
visit http://cct.edc.org or http://possibleworlds.edc.org

Preferred citation: Culp, K.M., Martin, W., Clements, M., & Presser, A.L. (2013). *Testing the impact of a pre-instructional digital game on middle-grade students' scientific misconceptions.* Possible Worlds Technical Report and Working Paper Series. New York, NY: Center for Children & Technology, Education Development Center, Inc. Retrieved from http://possibleworlds.edc.org.

Testing the impact of a pre-instructional digital game on middle-grade students' understanding of photosynthesis

Katie McMillan Culp, Wendy Martin, Ashley Lewis Presser, Margaret Clements
Center For Children & Technology/Education Development Center, Inc.
October, 2013

1. Introduction

School-age children and teenagers in the United States report investing significant amounts of time in playing digital games (Lenhart et al., 2008; Rideout, Foehr, & Roberts, 2010). Learning scientists and game designers have become interested in designing digital games for learning in part because they interpret this investment of time as evidence that these children and teens bring a high level of motivation and persistence to gameplay (Papastergiou, 2009; Prensky, 2003), and these factors have been shown to be important to improving student achievement in other contexts (Alderman, 2013; Brophy, 2010; Pintrich, 1999).

Surveys suggest that teachers in US K–12 schools have expanded their use of digital games for learning over the past several years, but in limited ways (Millstone, 2012; PBS/Grunwald Associates, 2011). K–5 teachers use games more than do teachers of the later grades, and digital games are most often used to practice component skills of reading and arithmetic. Digital games appear to be used less frequently with middle- and high-school–aged students, and used less to support learning of more advanced content—such as difficult science concepts, applied mathematical problem-solving, or historical thinking—even though these older students are the youth who play digital games most intensely outside of school and play more complex leisure games. David Buckingham (2009) calls to this type of gap the "new digital divide," referring to the stark contrast between students' engagement with electronic games and other digital media outside of the classroom and their very limited engagement with these technologies inside the school building. This "new" divide joins the prior, and still extant, digital divide created by inequitable access to technology and to opportunities to use it creatively and actively, which persists in the US and elsewhere (Helsper & Eynon, 2009; International Telecommunications Union, 2013).

Some researchers have called for a move away from an emphasis on investigations of whether games for learning "work" as a genre. Instead, they have recommended shifting research priorities to focus on identifying design features of games and their implementation contexts that effectively advance specific forms of engagement or progress (Clark, Tanner-Smith, & Killingsworth, 2013). Others maintain that the investment required to broadly develop and deploy digital games justifies a desire for strong evidence of effectiveness (Connolly, Stansfield, & Hainey, 2009; Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; de Freitas & Oliver, 2006). The present study was conducted with the expectation that experimental designs can contribute to both conversations, by producing both rigorous evidence of effectiveness and insight into whether key features of the pedagogical context may influence the impact of gameplay on learning.

This study responds to the need for more, rigorous, early-stage tests of the efficacy of digital games as supports for learning. It tests the impact of a digital game and associated classroom activities on middle-grade students' understanding of the photosynthetic process. It also addresses the need for further study of digital games in the context of classroom instruction to target core curricular concepts (Papastergiou, 2009), and the need for further investigation of how pedagogical context may shape the impact of gameplay on learning (Connolly et al., 2009). Photosynthesis is a commonly taught topic that addresses core concepts in both biology and energy transfer. It also is the subject of widespread scientific misconceptions (Driver, Squires, Rushworth, & Wood-Robinson, 1993; Schneps, Sadler, Woll, & Crouse, 1989), and is recognized by teachers as being difficult to teach effectively.

This study contributes to the growing body of research on the efficacy of digital games as supports for building science knowledge at the middle-grade level. It tests the impact of an intervention that locates digital gameplay as the anchor of a multi-step instructional sequence that unfolds over several days. This is one of the first experimental studies to test whether and how instructional quality in the classroom may mediate the impact of digital games on learning outcomes. These findings will allow us to better understand whether and how digital games can help students to master chronically difficult scientific concepts in real classroom contexts.

1.1 Theoretical background

1.1.1 Prior research on the impact of digital games on learning. The evidence base regarding the effectiveness of digital games as tools for learning remains limited and very diverse in focus (Clark et al., 2013; Connolly et al., 2009; Connolly et al., 2012; Ke, 2009). The most recent meta-analysis of this literature (Clark et al., 2013) identified 77 experimental and quasi-experimental impact studies that met their review criteria and had been published since 2000. These studies encompassed grades K through 12 and a wide range of content areas and targeted outcomes. Their analyses found persistent positive effects of playing digital games for learning on K–12 learning outcomes. This evidence is strongest for science, and is weighted toward the upper grades, with the largest number of studies conducted in the 9–12- and 6–8-grade bands.

Connolly et al.'s (2012) research synthesis, which examined both games and simulations and included only studies involving students ages 14–18, identified 70 high-quality studies. Of these, only 19 looked at games and their impact on content knowledge; many others focused on simulations or on other kinds of outcomes. They judged the sample of high-quality articles to be too limited and too diverse to support a meta-analysis. Based on a systematic, descriptive review of the studies, they report that outcomes were mixed, but consistently showed that "how games are integrated into the learning experience…[was] key to the success of the games-based approach."

Each of these syntheses makes clear that the evidence is spread across many different types of games, instructional designs, grade levels (including post-secondary), and curricular domains. Both reviews demonstrate that while a very broad range of articles are published about the impact of computer games and simulations on learning, few employ rigorous designs for testing impact, many test simulations rather than games, and very few look rigorously at impact on student learning within core curricular domains.

Both Connolly et al.'s (2012) and Clark et al.'s (2013) reviews also suggest that scaffolding of the in-game experience is important to the effectiveness of the game as a learning tool. While neither paper describes the types of scaffolding used in various studies in any detail, both demonstrate that games that provide scaffolds to help students recognize and reflect on what they are experiencing, whether through in-game supports or by making connections to other instructional experiences, are more likely to have positive effects on student outcomes. Similarly, these reviews suggest that effective games are played out over significant periods of time, and provide pedagogical context for play—suggesting that teachers should take an active role in leveraging students' experiences playing the games, relative to the desired learning outcomes.

1.1.2 Research on gameplay in the classroom context. Papastergiou (2009) notes the paucity of research that tests the efficacy of games for learning in the context of the social complexities of real classrooms. Many games for learning begin from a theoretical focus on the interaction between the student and the computer as the key opportunity for learning. But a long tradition of research drawing on social cognition frameworks has demonstrated that student interactions with educational technological tools are deeply embedded in, and influenced by, the ongoing social life of the individual student and the classroom as a larger social system (Koschmann, Hall, & Miyake, 2013; Sheingold, Hawkins, & Char, 1984). There are several important bodies of research on games for learning that have explored this broader context (Dede, 2009; Ketelhut & Schifter, 2011; Thomas, Barab, & Tuzun, 2009). But these studies have generally focused on whether and how games can support metacognitive and social skills rather than practices related to core curricular content. Squire, Barnett, Grant, & Higginbotham (2004) conducted one of the few studies of digital games that documented the role that teacher-led, in-classroom discussion of game features and the gameplay experience played in students' learning of core curricular concepts.

Researchers working in other domains have emphasized the importance of attending to potential mediating and moderating variables in experimental trials, such as features of the social or instructional context (Wayne, Yoon, Zhu, Cronen, & Garet, 2008). Similar choices are likely to be productive for research on games for learning. For example, the critical role the teacher plays in guiding the students' sense-making, based on interactions with technology-rich interventions, has been well documented through design-based research and implementation studies (Roschelle, Knudsen, & Hegedus, 2010). But there is very little evidence yet available that examines whether and how the quality or character of teachers' instructional practice might mediate or moderate the impact of a digital game on students' learning.

1.1.3 Pre-instructional experiences: Preparation for future learning and analogical reasoning. The digital-game–based intervention discussed in this article was designed in accordance with Bransford and Schwartz's (1999) proposed instructional model called "preparation for future learning." This model suggests that engaging students in structurally relevant direct experiences that are then followed by instruction should increase the likelihood of transfer. Preinstructional activities set the stage for learning from subsequent instruction by providing students with experiences from which they can draw to make sense of subsequent material.

Schwartz and Martin (2004) write, "When preparing students to learn, the instructional challenge is to help students transfer in the right knowledge" (p. 132). In order to increase the likelihood that learners forge productive connections between a preparatory experience—in this case, playing a digital game—and the targeted concepts, teachers must be prepared to help their students by clarifying which features the preparatory experience and the targeted concept share, and discussing how the relevant processes in both are alike (Cameron, 2002; Venville, 2008).

Providing a shared, structurally relevant source for analogical thinking is particularly important when addressing difficult scientific concepts that are often the subject of misconceptions. There is a broad literature on the status and structure of scientific misconceptions that has been documented and critically reviewed elsewhere (see Duit, 2009, and Vosniadou, 2008). This project primarily follows the work of Chi, Slotta, and others who have argued that a fundamental feature of persistent scientific misconceptions is the absence of a pre-existing conceptual category that students could use to ground their exploration of novel scientific information, experience, or evidence (Chi, 2008; Chi, Roscoe, Slotta, Roy, & Chase, 2011; Slotta & Chi, 2006; Slotta, Chi, & Joram, 1995). For alternative approaches to the issue, see, for example, Gupta, Hammer, & Redish, 2011; Hammer, Gupta, & Redish, 2011; Smith, diSessa, & Roschelle, 1994; or Vosniadou, 2012.

According to Chi and Slotta's work, students may be better prepared to begin to accommodate and persist with a new, more accurate scientific concept when they have an accessible, familiar, and analogous mental model in hand prior to exposure to the new concept. The familiar analogical ground becomes a tool they can use to begin to make sense of this new explanation of a given phenomenon.

Gentner (2010) and her colleagues have extensively researched another paradigm for understanding the role pre-instructional experiences can play in supporting conceptual learning. Their work on structure mapping theory has demonstrated that specific approaches to structuring and rehearsing analogical reasoning with students can support student understanding of abstract concepts. Reese (2007, 2009) has developed digital games grounded in this theoretical approach, creating games for learning whose features are designed to map to conceptual features in a target domain. The game features then become metaphors for abstract concepts learners will encounter in their classes.

1.2 The current study

This article reports on a test of the impact of *Exploring Photosynthesis* on student learning. *Exploring Photosynthesis* is one of four supplementary modules developed as part of a larger project. Each of the four modules focuses on one difficult-to-teach topic that is often the subject of scientific misconceptions, and each includes a digital game and a series of related in-class, non-digital activities that can be integrated into regular instruction.

The study reported on here is part of a larger sequence of research and development activities conducted over a four-year period. Other reports, in preparation, present detailed findings about classroom implementation of early versions of the module. In addition to these exploratory, descriptive studies, we elected to conduct this early-stage, rigorous test of the impact of one module in order to contribute to the literature on the impact of digital games on learning in the ways described in Section 1, and to determine whether early evidence of promise might be forthcoming from such a rigorous test of the game and its relationship to student learning.

This study poses three research questions.

1. *Impact on student outcomes*. In classrooms where teachers implement the *Exploring Photosynthesis* module, do students demonstrate a significantly better understanding of how photosynthesis occurs and how mass and energy are

conserved during chemical changes than do students in classrooms where teachers do not use the module?

2. *Fidelity of implementation*. Do teachers in the treatment group implement the intervention with a high level of fidelity?
3. *Influence of quality of instruction on student outcomes.* Does the quality of teachers' instruction moderate the impact of the module on student understanding of the target concepts?

2. Method

This blocked, cluster randomized controlled study compares student performance on a photosynthesis knowledge assessment for middle-school students whose teachers taught their photosynthesis unit with or without the *Exploring Photosynthesis* intervention (*Exploring Photosynthesis* vs. business-as-usual groups). A cluster randomized design was used because the nature of this classroom-level intervention precludes randomization of individual students. Instead, randomization occurs at the teacher level, and hierarchical linear modeling techniques are used to estimate effects of the intervention on students, allowing us to account for within-group commonalities among students who share the same teacher. The outcome measure was a researcher-designed, multiple-choice assessment that was aligned to standard curricular expectations for coverage of photosynthesis in the middle grades and had been previously pilot-tested and revised based on psychometric analyses. To minimize the burden on students and teachers, and to minimize error introduced into the assessment measure, the assessment was administered only after the intervention. Other data sources, detailed below, were used to check equivalence of the two samples. Implementation of the intervention was staggered throughout the 2011–12 school year because science teachers cover photosynthesis at various points during the school year.

2.1 Description of the intervention

Exploring Photosynthesis is part of a larger research and development project that sought to test the potential of portable, digital games as a way to provide a preparatory, pre-instructional experience. It draws on Bransford and Schwartz's "preparation for future learning" instructional model (1999), and positions digital gameplay as an activity that engages middle-grade students in repeated interactions, through core game mechanics, that are structurally analogous but nominally unrelated to the target concepts (like sandbox play or open-ended construction activities for younger children). These repeated, shared in-game experiences become a source for grounded analogical reasoning during later instruction. In this approach, digital gameplay becomes a necessary—but insufficient on its own—first step in an instructional process that includes multiple forms of engagement with the target concepts.

As implemented for this study, *Exploring Photosynthesis* included five sequenced activities that were integrated into teachers' normal instruction about photosynthesis. Students were assigned the game as homework and asked to play it for a minimum of 30 minutes, using Nintendo DSs provided as part of the intervention. Teachers then integrated visuals drawn from the game into their discussion of the structure of glucose and the photosynthetic process. Students participated in two active, hands-on activities that reinforced the process of breaking apart carbon dioxide and water and recombining the elements to create glucose and oxygen. They then participated in a consolidation activity in which they drew on their knowledge of photosynthesis to evaluate scientific claims made by journalists in a fictional tabloid.

2.2 About the technology platform

The *Possible Worlds* games were developed for the Nintendo DS (shifting to the DSi when it was released in 2009) to support investigation of the role that portable devices could play in responding to the chronic limitations of in-school technology infrastructures. This platform was chosen in 2007, when tablet computers were not yet widely available, smartphones were just gaining traction with adult consumers, and adolescents had very limited access to either type of device. Using the DS was a low-cost way to ensure that all students could play the games, regardless of their level of technology access at home or at school. The Nintendo DS also was a compelling choice for this project because it can deliver games that look and feel familiar and entertaining to many students and is designed to withstand being transported by children.

2.3 Random assignment

Forty-two teachers from 25 schools in New York State were recruited for the study. We randomly assigned teachers to the treatment or control group using a blocked randomization plan that grouped teachers by the percent of students in their school who qualify for free or reduced-price lunch (FRPL) prior to random assignment. One group taught in schools where 40% or fewer students received FRPL, and the other taught in schools where more than 40% of the students received FRPL. Given the diversity of participating teachers' school characteristics (see Table 1), simple random assignment could have resulted in differences between the treatment and control groups, which would have threatened the baseline equivalence of the study.

Table 1

Characteristics of participating teachers' schools

School characteristic (*n* = 25)	Mean (SD)	Median	Min	Max
Percent of students eligible for free or reduced-price lunch	32.2 (24.7)	25.8	0	84
Total school enrollment	758.7 (263.8)	735	198	1153
Average grade-level enrollment	240.5 (100.3)	253	64	384
Race/ethnicity of student population [a]				
Percent black/non-Hispanic	11.3 (11.0)	7.0	0	38
Percent Hispanic	17.0 (17.4)	11.0	0	62
Percent white/non-Hispanic	65.6 (25.2)	71.5	15	98

[a] To maintain the confidentiality of participating schools, we do not report the representation of race/ethnic group categories that are low-incidence in New York State. Race/ethnic group categories included in the Common Core but not reported in this table include American Indian/Alaskan Native; Asian or Native Hawaiian/Pacific Islander; and "two or more races."

The sample size ensured that the study would have the statistical power to detect a mean detectable effect size in student assessment scores between treatment and control groups of .24 of a standard deviation, which is a common metric for describing the magnitude of an educational intervention in statistical terms. One of the teachers in the treatment group withdrew from the study in January, prior to teaching photosynthesis.

2.4 Sample

2.4.1 Teacher sample. The final sample of 41 teachers taught in a total of 25 primarily middle-grades schools. Three of the schools spanned grades K–8. As Table 1 (above) shows, the schools in which teachers worked varied considerably in terms of school size, the racial/ethnic group composition of the school, and the students' socio-economic status as measured by the percent qualifying for free or reduced-price lunch. Table 2 (below) presents descriptive statistics for the teacher sample. On average, participating teachers had over 14 years' total teaching experience (SD = 7.1) and 12 years of teaching middle-school science (SD = 7.2). Over 90% had earned their Masters' degree (with most of these having focused their graduate study on science or science education) and were certified in New York State to teach biology.

Table 2

Characteristics of participating teachers

Characteristic	Full Sample Mean (SD)	Treatment Mean (SD)	Control Mean (SD)
Number of teachers	41	21	20
Years of experience overall	14.12 (7.1)	12.1 (7.23)	16.25 (6.48)
Years of experience teaching middle-school science	12.08 (6.8)	10.47 (7.26)	13.60 (6.19)
	N (%)	N (%)	N (%)
Master's degree earned	38 (92.7)	20 (95.2)	18 (90.0)
Major field of study for highest degree completed			
Biology, other science, or science education	28 (68.3)	15 (71.4)	13 (65.0)
Humanities, social science, or education (not	12	5	7

science-specific)	(29.3)	(23.8)	(35.0)
Certified in New York State to teach Biology	38 (92.7)	20 (95.2)	18 (90.0)

Note: In cases where participants chose not to respond, the sum of teachers is less than 41 and percentages will not add up to 100. This is also true in cases where the number of teachers in any particular cell would be fewer than five, as we do not report these statistics in order to ensure the confidentiality of teachers' participation.

There are no statistically significant differences between treatment and control groups.

2.4.2 Student sample. In order to reduce the data collection burden on teachers and students, for each participating teacher we randomly selected one class to be the "focal" participating class for the study and collected all study data from this classroom. Only general education classrooms were eligible to be the focal class. Classes that were targeted either to "accelerated" students or students with learning difficulties were excluded. A total of 914 students participated in the study. There were no statistically significant differences in demographic characteristics, TOSRA scores or state standardized test scores between the students in treatment and control classrooms. Table 3 presents descriptive statistics for the student sample as a whole and for the treatment and control groups. See Section 2.5.2 for detail on the measures used to collect these data.

Table 3

Student characteristics

Characteristic	Full sample N (%)	Treatment N (%)	Control N (%)
Race/ethnicity (n = 912)			
White/non-Hispanic	571 (63%)	302 (63%)	269 (62%)

Black/non-Hispanic	121 (13%)	60 (13%)	61 (14%)
Hispanic or Latino	178 (20%)	90 (19%)	88 (20%)
Minority [a]	341 (37%)	175 (37%)	166 (38%)
Male (n = 914)	452 (50%)	238 (50%)	214 (49%)
Has an IEP (*n* = 870)	107 (12%)	63 (15%)	44 (10%)
Classified as ELL (*n* = 893)	56 (6%)	17 (4%)	39 (9%)
	Mean (SD)	Mean (SD)	Mean (SD)
Age in years [b] (*n* = 849)	12.7 (.5)	12.6 (.5)	12.7 (.5)
Mean centered state mathematics score (*n* = 868) [c]	-0.7 (26.2)	-1.2 (24.9)	-.1 (27.6)
Mean centered state language arts score (*n* = 865) [d]	1.0 (15.0)	-0.3 (14.0)	2.4 (16.0)
TOSRA score	3.3 (.6)	3.4 (.5)	3.3 (.6)

Note. This table reports the descriptive statistics for the student sample prior to multiple imputation. IEP = Individualized education plan; ELL = English language learner; TOSRA = Test of science. There are no statistically significant differences between treatment and control groups.

[a] The minority classification is the sum of black/non-Hispanic; Hispanic or Latino; Asian, Pacific Islander, or Alaska Native; and multiracial students. Because several of the race/ethnic group categories had very little variability, and therefore could not be included in the analysis, we report the percent of students in a minority race/ethnic group as this is the variable included in the analysis.

[b] Age is calculated as the difference between each students' birthdate and the mid-point of the 2011–12 academic year.

[c] Students' state mathematics scores are mean-centered using the state average mathematics score for the grade during which students took the exam.

[d] Students' state language arts scores are mean-centered using the state average language arts score for the grade during which students took the exam.

2.5 Measures

Data were collected for four purposes: to measure student outcomes; to test the equivalence of the two randomized samples of classrooms on variables potentially relevant to student outcomes; to provide basic documentation of implementation fidelity in the treatment classrooms; and to measure the quality of teachers' instructional practice.

2.5.1 Measuring outcomes. To assess students' understanding of photosynthesis and chemical change, we administered a 33-item, paper-and-pencil assessment during a regular classroom period (i.e.., approximately 40 minutes) at the conclusion of the unit on photosynthesis. A composite total score was calculated for all students completing more than half of the items. The internal reliability for the assessment, as measured by Cronbach's alpha, is .86.

Members of the study team with expertise in middle-grades science and assessment development created, pilot-tested, and revised the assessment. A pilot version of the assessment consisted of 46 items that addressed the photosynthesis module's content and were available as released items from multiple states' science assessments, published formative assessment probes, and questions developed by the study team. This version of the assessment was pilot tested with 484 students in four public middle schools during field tests of this module in urban and suburban areas of the Northeast during the 2010–2011 school year. The assessment was then revised and shortened to include 33 questions, based on psychometric analysis of the pilot data.

2.5.2 Testing baseline equivalence on relevant variables. As reported in Section 2.4.2, these data were collected to establish the baseline equivalency of treatment and control group teachers' focal classrooms on potentially relevant variables.

2.5.2.1 *Student state standardized mathematics and language arts test scores*. In lieu of administering a pre-test that could be used as a covariate in the data analysis, the study collected the previous year's state standardized mathematics and language arts test scores for all students in each of the teachers' participating classrooms. This approach is used to minimize the data collection burden on teachers and students, study costs, and the introduction of bias into the outcome assessment that could be associated with use of a baseline assessment more closely aligned with the outcome (Bloom, Richburg-Hayes, & Black, 2007). Several methodological reports have established the use of state test scores at baseline as an acceptable alternative to the use of more closely aligned assessments (Bloom et al., 2008; Deke, Dragoset, & Moore, 2010). For seventh-grade participants, we collected sixth-grade standardized test scores, and for eighth-grade participants, we collected seventh-grade scores. Because sixth and seventh graders take different standardized tests and the scoring for each test varies, they were transformed into a common metric.

2.5.2.2 *Student demographic characteristics*. The study collected data describing the demographic characteristics of students in the participating classrooms from school administrative records. These variables include gender, race/ethnicity, whether the student had an Individual Education Plan (IEP), and whether the student was classified as an English Language Learner (ELL). We use these variables to describe the sample, establish the baseline equivalence of the treatment and control groups, and as covariates in the data analysis.

2.5.2.3 *Student attitudes toward science*. Students responded to a series of 13 items designed to measure their attitudes toward science. The items were drawn from the

Test of Science Related Attitudes (or TOSRA, see Fraser, 1981), a questionnaire originally developed to measure seven aspects of student attitudes towards science. For this study, we selected 13 items from three of the original subscales: *leisure interest in science*, *enjoyment of science lessons*, and *attitude toward scientific inquiry*. We computed an overall TOSRA score for each student by calculating their average response across all 13 items for students who answered at least two-thirds of the items (9 of 13 items). Table 3 (above) reports the average TOSRA scores for the full sample and for the treatment and control groups. The internal reliability for the scale as measured by Cronbach's alpha is .73.

2.5.3 Documenting fidelity of implementation.

2.5.3.1 *Student gameplay.* The study collected data about the duration of students' gameplay by issuing each student a uniquely identified game cartridge. The cartridges collected the length of time that a student spent playing each level of the game. We summed the amount of time to create a total gameplay variable.

2.5.3.2 *Teachers' implementation of the* Exploring Photosynthesis *module.* Teachers in the treatment condition completed a log detailing aspects of the intervention that were implemented. Our use of self-report logs drew on prior work that has documented high levels of agreement between teacher logs and observer ratings (Mayer, 1999; Mullens & Gayler, 1999). This approach maximizes the amount of detail we were able to collect while minimizing the cost associated with collecting the data, as extended observations in all treatment classrooms were not possible. We asked teachers to complete the relevant sections of the log on the day they implemented those components of the *Exploring Photosynthesis* module, and we collected the logs once they completed the intervention. For each item, teachers indicated whether or not they implemented that aspect of the intervention. We created four composite scales to measure the *content coverage* (12 items), *making links between the game and science content* (28 items), *student engagement* (4 items), and *technical difficulties* (11 items). For the first two scales, teachers' scores on each scale indicate the number of intervention components teachers reported implementing.

2.5.4 Measuring instructional quality. We used the Classroom Assessment Scoring System-Secondary Edition (CLASS-S) observation framework to measure the quality of instruction provided by each of the participating teachers during a typical day of science instruction (Pianta, Hamre, Hayes, & Mintz, 2011). The procedures for use of this instrument as described here are consistent with standards used in other peer-reviewed studies and are recommended by members of the instrument development team (Malmberg & Hagger, 2009; Pianta & Hamre, 2009; M. Stuhlman, personal communication, October 11, 2011). These studies have documented that even very limited snapshots of teacher practice, when collected using the CLASS-Secondary, are highly reliable when collected by certified observers and are predictive of student performance on high-stakes standardized tests—a debatable, but standard, benchmark of the validity of measures of instructional quality. More specifically, Pianta and Hamre (2009) have demonstrated that small numbers of observations prove to hold a high level of predictive validity, and can effectively discriminate among teachers, because between-teacher variation is, on average, much broader than within-teacher variation over time.

Study team members who conducted the CLASS-Secondary observations were trained and certified in the use of the instrument. Before each teacher began teaching the unit that included photosynthesis, trained researchers observed the participating classroom on one occasion for 40 minutes. This allowed for two 15-minute observation intervals, each followed by 5 minutes for coding, as recommended by the CLASS-S manual.

Observers rated teachers on 12 dimensions covering emotional support, classroom organization, instructional support, and student engagement. Scores on each the 12 dimensions can range from low (1 or 2) to mid (3, 4, or 5) to high (6 or 7). We created the overall CLASS-S score for each teacher by computing the average rating across the 12 scores. The average CLASS-S score across all teachers in the study was 4.28 (SD = 0.83) and the internal consistency as measured by Cronbach's alpha was .91. There were no statistically significant differences between treatment and control teachers on the composite CLASS-S score or any of the sub-scales.

2.6 Data analysis approach

2.6.1 Missing data strategy. As can happen in a study that collects data from participants over time, we did not have complete data for all students in each teacher's focal classroom. Simply deleting cases with missing data can produce estimates that are biased or unreliable, especially as there is evidence that students with missing data may be systematically different from students without missing data (Peugh & Enders, 2004). To address this issue, we used a multiple imputation strategy that created multiple versions of the data set in which all of the missing values are predicted using the existing values for other variables (Enders, 2010; Song & Herman, 2010).

2.6.2 Data analysis. We conducted two types of analyses to answer the research questions for this study. To answer Research Questions 1 and 3, which investigate whether the photosynthesis module resulted in improved student learning and whether the impact of the intervention was moderated by instructional quality, we conducted two-level regression analyses using HLM 6 software (Raudenbush, Bryk, & Congdon, 2004). The two-level regression models take into account that students are "nested" in classrooms and therefore are not statistically independent from each other. To ignore the nested structure of the student data could result in underestimating the size of the standard errors for the treatment effect and overestimating the impact of the intervention on students' assessment scores. The HLM model conducted for this study includes student-level data in the Level 1 equation and teacher-level data in the Level 2 equation. The models are explained in detail in the results section. We investigated Research Question 2, which asks about the implementation of the *Exploring Photosynthesis* module, using descriptive statistics.

3. Results

3.1 Impact of the *Exploring Photosynthesis* intervention on student outcomes

The photosynthesis assessment scores of students whose teachers were randomly assigned to use the *Exploring Photosynthesis* module were not significantly different from the scores of students whose teachers were assigned to the control group. Table 4 presents regression coefficients, standard errors, and p-values for each of the Level 1 and Level 2 predictors.

As described above, we used a multi-level regression to estimate the impact of the intervention on students' photosynthesis assessment scores. This analysis accounts for the fact that students are nested within teachers. The Level 1 (student-level) model was:

$$Assessment_{ij} = \beta_{0j} + \beta_{1j}(StateMath_{ij}) + \beta_{2j}(StateLangArts_{ij}) + \beta_{3j}(TOSRA_{ij}) + \beta_{4j}(Male_{ij}) + \beta_{5j}(Age_{ij}) + \beta_{6j}(Minority_{ij}) + \beta_{7j}(IEP_{ij}) + \beta_{8j}(ELL_{ij}) + \varepsilon_{ij}$$

where $Assessment_{ij}$ is the score on the photosynthesis assessment score for each student *i* of teacher *j* at the end of the photosynthesis unit. The remaining variables in the Level 1 model are covariates we included in order to statistically adjust for pre-existing differences in students and thereby increase the estimate of the impact of the intervention. $StateMath_{ij}$ and $StateLangArts_{ij}$ are each student's state mean-centered standardized assessment scores from the previous year, $TOSRA_{ij}$ is each student's grand mean-centered TOSRA score, $Male_{ij}$ is each student's gender (0 = female, 1 = male), $Minority_{ij}$ indicates whether the student is a member of a race or ethnic minority group (0 = white/non-Hispanic, 1 = member of a race or ethnic minority group), IEP_{ij} indicates whether a student has an individualized education plan and qualifies for special education services, and ELL_{ij} indicates whether a student is classified by the school as being an English language learner.

Because this is a teacher-level intervention and the study randomized teachers to the treatment and control groups, the test of whether the *Exploring Photosynthesis* module intervention had an impact on students' assessment scores is specified in the Level 2 (teacher-level) model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Treatment_j) + \gamma_{02}(PercentFRPLj) + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$
$$\beta_{5j} = \gamma_{50}$$
$$\beta_{6j} = \gamma_{60}$$
$$\beta_{7j} = \gamma_{70}$$
$$\beta_{8j} = \gamma_{80}$$

Specifically, $Treatment_j$ indicates whether or not a teacher was in the treatment group (0 = control group, 1 = treatment group) and γ_{01} captures the difference in average assessment scores for treatment and control group classrooms. The Level 2 model also included the percent of students in each teacher's school who qualified for free or reduced price lunch, as this was the variable we used to group teachers prior to conducting random assignment ($PercentFRPL_j$).

Table 4

Coefficients (and standard errors) for all variables included in the multi-level regression models conducted to answer Research Question 1 and Research Question 3

	Model estimating the impact of the intervention	Model testing the interaction between condition and CLASS-S scores
Intercept	21.82*** (.57)	21.82*** (.54)
Level 2 (teacher) covariates		
Condition (treatment or control)	-1.01 (.79)	-1.00 (.76)
Percent of students in the school eligible for free or reduced-price lunch	-.05* (.02)	-.05* (.02)
CLASS-S		-.93

		(.69)
Condition x CLASS-S		1.72†
interaction		(.92)
Level 1 (student) covariates		
Male	-1.10*	-1.12**
	(.44)	(.37)
Age	-.40	-.42
	(.45)	(.43)
Minority	-.45	-.42
	(.56)	(.44)
IEP	-.04	-.02
	(.79)	(.74)
ELL	-1.82	-1.73
	(1.13)	(.90)
State math test score	.08***	.08***
	(.01)	(.01)
State language arts test	.11***	.11***
score	(.02)	(.02)
TOSRA	1.11**	1.10**
	(.35)	(.34)

Note. CLASS-S = Classroom Assessment Scoring System-Secondary Edition; IEP = Individualized education plan; ELL = English language learner; TOSRA = Test of science related attitudes
*** $p \leq .001$
** $p \leq .01$
* $p \leq .05$
† $p \leq .05$

3.2 Implementation of the *Exploring Photosynthesis* intervention

3.2.1 Time spent on gameplay as homework. We determined whether and how much students played the digital game as homework by computing the mean and standard deviation for the total gameplay variable. Non-zero data on time spent playing the game were available for 77.4% of students in the treatment group. Missing gameplay data is due either to a student not playing the game at all, or to the game chip that records a student's game activity being missing or damaged. Prior field tests of *Exploring Photosynthesis* had demonstrated that students did sometimes encounter faulty chips that allowed them to play the games but did not record time played. This most likely did occur and was responsible for some of the missing gameplay times, but others are surely indicators that students did not play the games. Therefore, at most 22.6% of the student sample did not play the games on their assigned DS machines. For the students with non-zero gameplay data available, the average gameplay time was 47.4 minutes (SD = 31.5), and ranged from 1 to 168 minutes. This average time was well above the assigned benchmark of 30 minutes of gameplay time that students were expected to complete as homework.

3.2.2 Teachers' fidelity to the instructional sequence of the module. To describe the extent of teachers' fidelity to the intended implementation sequence for the module, we computed the means and standard deviations for each of the four scales included in the implementation log. Teachers' responses on the *content coverage* subscale showed that on average they implemented 11.55 of the 12 items (SD = 0.8), indicating that treatment teachers covered essentially all of the topics covered by the intervention. The *student engagement* composite consisted of four questions and yielded a mean of 3.2 (SD = 0.8) with a range of 2–4, indicating that teachers judged students to be mostly engaged in module activities. The *technical difficulties* composite consisted of 11 questions and yielded a mean of 3.2 (SD = 1.51; range 1 to 6), indicating that there were few technical difficulties present to impede the implementation.

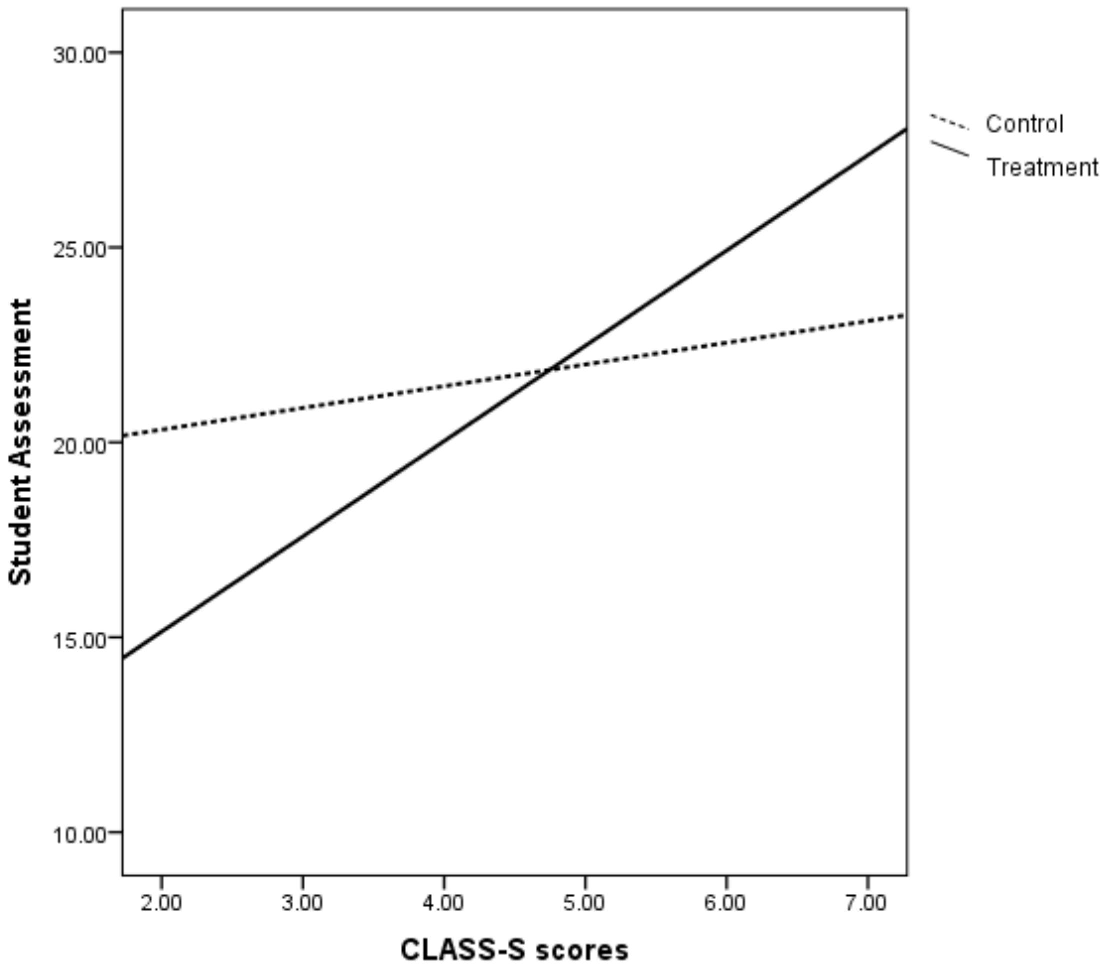
The *making links between science content and the game* subscale asked teachers to indicate whether or not they made direct reference to the game during their instruction. Teachers' responses to items on this subscale indicate that teachers implemented these aspects of the *Exploring Photosynthesis* module less consistently. Of the 28 opportunities to link the game to science content, as outlined in the implementation log, teachers reported making 21.1 of the links on average (SD = 4.0, range 14 to 28). A closer inspection of this data indicates that only a few teachers made certain of the linkages between science content and the game. For example, only 11% of teachers reported that they asked additional questions that linked the "Molecules in Motion" activity to the DSi game.

3.3 Quality of teachers' instruction as a moderator of student outcomes

We conducted a second, exploratory multi-level regression analysis to examine the possibility that the impact of the intervention on students' assessment scores varied depending on the quality of instruction provided by students' classroom teachers, as measured by the CLASS-S. This required adding two additional variables to the Level 2 model—each teacher's CLASS-S score collected during the classroom observation, and a variable representing the *Treatment x CLASS-S* interaction.

These analyses indicated that CLASS-S scores were not associated with students' assessment scores across the study sample as a whole. However, there was a trend toward statistical significance for the *Treatment X CLASS-S* interaction term (p = .07). This finding is best illustrated by looking at a scatterplot of teachers' CLASS-S scores by their students' average assessment scores. In Figure 1, the solid line represents the association between treatment teachers' CLASS-S score and their students' average score on the photosynthesis assessment, and the dotted line represents the association between control teachers' CLASS-S scores and their students' scores. The figure suggests that intervention was more strongly associated with students' assessment scores in the classrooms where there was higher quality teaching than in the classrooms where the teaching quality received lower ratings.

Figure 1



4. Discussion and Conclusions

This trial demonstrated that *Exploring Photosynthesis* did not have a significant impact on student understanding of the photosynthetic process. Below we briefly discuss several features of the intervention as implemented and limitations of the study design that may have contributed to this outcome.

4.1 Logistics of implementation

Exploring Photosynthesis was designed to relate digital gameplay to typical middle-grade science instruction in a way that caused minimal disruption to teachers' normal instructional sequences and classroom practices. The limited data available suggest that the intervention was logistically and technically feasible for teachers to implement. Teachers in the treatment group were asked to integrate the *Exploring Photosynthesis* resources, including the digital game and several in-class activities, into their normal coverage of the photosynthetic process. Teachers were willing and able to manage the process of sending portable electronic devices home with students, and reported that they encountered minimal technical challenges during the intervention. No digital devices were lost or stolen by study participants, and parents and administrators did not raise any concerns about the use of the devices, or about assigning digital games as homework. The study had a very low attrition rate, with only one teacher leaving the study after randomization.

Student compliance with the expectation that they play the digital game as homework was consistent with other evidence about student homework completion. 77.4% of students in the treatment group did play the digital game as homework. This rate of completion is consistent with an overall homework completion rate of 77% reported by students in a nationally representative survey (Markow, Kim, & Liebman, 2007).

4.2 Instructional challenges of implementation

While implementation of the intervention appeared to be logistically manageable for teachers, data on fidelity of implementation suggests that providing effective instructional support for the intervention was more difficult. Fidelity log reports indicate that most treatment teachers did not make explicit connections between the game and the target concepts that were recommended in the instructional sequence. For example, during in-class discussion of the structure of glucose, teachers were asked to use an illustration of the glucose model that was drawn from *Ruby Realm*, which students would have just played. They were asked to prompt students to remember where they had seen the image before, and what they did in the course of the game to create the structure (which was not labeled as glucose in the game itself). The majority of treatment teachers reported that they did not make this or other similar explicit connections between the game and the target concepts. While there is likely no one single cause for teachers' lack of fidelity in this step of the intervention, anecdotal evidence from this study and evidence from earlier field tests of the module suggest that teachers had not played the game extensively themselves, and may not have been comfortable using the game as a point of reference.

Providing explicit scaffolding to help students build analogical relationships between the game and the target concepts appears to require considerable instructional skill, as well as comfort and familiarity with the digital game itself. Exploratory analyses of the CLASS-S data, described above, also suggest that in the sample as a whole the more-skillful teachers explained, supported, and structured students' overall instructional experience in ways that had a critical impact on what students learned.

4.3 Limitations of the study

This randomized control trial tested the impact of a specific intervention on student outcomes, and its results cannot be generalized to other specific digital games. The study intentionally focused on the potential impact of the intervention in the context of broader classroom environments and practices, and most, though not all, aspects of the intervention were implemented with a high level of fidelity. However, the study design and the scope of the data collected do not allow us to investigate potential associations between specific implementation contexts and student outcomes in any detail. Conclusions about the quality of implementation in the treatment group are based on limited data sources. Future publications about this project will investigate these issues in more depth.

The structure of the intervention led, in many cases, to teachers in the intervention group spending several more instructional periods on photosynthesis than they normally did, and longer than did teachers in the control group. Extended instructional time for the treatment group is generally viewed as a confounding factor that should be avoided in impact evaluations. However, given the existing research base on the persistence of the misconceptions students were likely to hold regarding photosynthesis, simply providing more instruction per se was unlikely to have any impact on the quality of student outcomes. As prior research has demonstrated, change in conceptual understanding requires effective intervention in and gradual displacement of prior beliefs. Extending exposure to ineffective methods is unlikely to change student understanding.

This project sought to shift student understanding about a topic that has been documented to be the subject of persistent, lifelong misconceptions. The student assessment was psychometrically sound and effectively discriminated among students with demonstrably different levels of understanding. These assessment data allowed us to draw rigorous conclusions about the impact of the intervention in the context of typical classroom cycles of instruction and assessment. However, it is possible that a different research approach, such as a longitudinal study of students' emerging conceptual understanding, or the use of a concept inventory (Hestenes, Wells, & Swackhamer, 1992) rather than an assessment of students' accurate knowledge, might have captured more nuanced evidence of student learning about the target concepts.

4.4 Conclusions

This study demonstrates that *Exploring Photosynthesis* did not have an impact on student outcomes as measured by an objective assessment closely aligned to the goals of the intervention. The findings should be relevant to others in the games for learning community for several reasons. The study findings suggest that teacher instructional quality played a meaningful moderating role in determining student outcomes for the treatment group. This finding should be of interest to other developers of games for learning who are seeking to provide informal, pre-instructional learning experiences for students. Both prior research (Gentner, 2010; Richland, Zur, & Holyoak, 2007) and the limited fidelity of implementation data collected in this study suggest that, in order for pre-instructional gameplay to support targeted learning goals, it may be particularly important (though not necessarily sufficient) for teachers to articulate and map explicit analogical relationships between relevant features of the game and the targeted concepts during instruction. In the absence of skilled facilitation of this process, these findings suggest that students did not benefit from the tacit analogical grounds made available by this particular digital game. Further research should investigate in more detail what supports teachers need in order to make explicit analogical connections between features of digital gameplay and target concepts, and whether those connections then lead to more productive outcomes.

References

Alderman, M. K. (2013). *Motivation for achievement: Possibilities for teaching and learning*. New York, NY: Routledge.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*, 30–59.

Bloom, H., Zhu, P., Jacob, R., Raudenbush, S., Martinez, A., & Lin, F. (2008). *Empirical issues in the design of group-randomized studies to measure the effects of interventions for children*. (MDRC Working Papers on Research Methodology). New York: MDRC.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education: 24* (pp. 61–100). Washington, DC: American Educational Research Association.

Brophy, J. E. (2010). *Motivating students to learn*. New York, NY: Routledge.

Buckingham, D. (2009). Beyond technology: Rethinking learning in the age of digital culture. *Youth Media Democracy: Perceptions of New Literacies. Proceedings of the Youth Media Democracy Conference*, pp. 37-43. Dublin, Ireland: Center for Social and Educational Research.

Cameron, L. (2002). Metaphors in the learning of science: A discourse focus. *British Educational Research Journal*, *28*(5), 673–688.

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of* *research on conceptual change* (pp. 61–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2011). Misconceived causal explanations for emergent processes. *Cognitive Science*, *36*, 1–61.

Clark, D. B., Tanner-Smith, E., & Killingsworth, S. (2013). *Digital games for learning: A systematic review and meta-analysis* (Preliminary Executive Summary and Brief). Retrieved from SRI International website: http://www.sri.com/work/projects/glasslab-research.

Connolly, T. M., Stansfield, M., & Hainey, T. (2009). Towards the development of a games-based learning evaluation framework. In T. Connolly, M. Stansfield, & L. Boyle (Eds.), *Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices* (pp. 251–273). Hershey, PA: IGI Global.

Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T., Boyle, J.M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers and Education, 59*(2), 661–686.

de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers and Education*, *46*(3), 249–264.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, *323*(5910), 66–69.

Deke, J., Dragoset, L., and Moore, R. (2010). *Precision gains from publically available* *school proficiency measures compared to study-collected test scores in education cluster-randomized trials* (NCEE 2010-4003). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1993). *Making sense of secondary science.* London, UK: Routledge.

Duit, R. (2009). *Bibliography—STCSE: Students' and teachers' conceptions and science education.* Retrieved from IPN—Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematic an der Universitat Kiel website: http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Fraser, B. J. (1981). *TOSRA test of science-related attitudes Handbook*. Hawthorn, VIC, Australia: Australian Council for Educational Research.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775.

Gupta, A., Hammer, D., & Redish, E. F. (2011). The case for dynamic models of learners' ontologies in physics. *Journal of the Learning Sciences*, *19*(3), 285–321.

Hammer, D., Gupta, A., & Redish, E. F. (2011). On static and dynamic intuitive ontologies. *Journal of the Learning Sciences*, *20*(1), 163–168.

Helsper, E., & Eynon, R. (2009). Digital natives: Where is the evidence? *British Educational Research Journal*, *36*(3), 503–520.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141.

International Telecommunications Union (2013). *Measuring the information society* (ITU Publication No. 254). Geneva, Switzerland: Author. Retrieved from International Telecommunications Union website: http://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2013/MIS2013_without_Annex_4.pdf.

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (Vol. 1, pp. 1–32). Hersey, PA: Information Science Reference.

Ketelhut, D. J. & Schifter, C. C. (2011). Teachers and game-based learning: Improving understanding of how to increase efficacy of adoption. *Computers and Education, 56*(2), 539–546.

Koschmann, T., Hall, R. P., & Miyake, N. (Eds.). (2013). *CSCL 2*. New York, NY: Routledge.

Lenhart, A., Kahne, J., Middaugh, E., Macgill, A. R., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics*. Retrieved from Pew Internet & American Life Project website: http://www.pewinternet.org/Reports/2008/Teens-Video-Games-and-Civics.aspx

Malmberg, L. E., & Hagger, H. (2009). Changes in student teachers' agency beliefs during a teacher education year, and relationships with observed classroom quality, and day-to-day experiences. *British Journal of Educational Psychology, 79*(4), 677–694.

Markow, D., Kim, A., & Liebman, M. (2007). *MetLife survey of the american teacher: The homework experience. A survey of students, teachers and parents*. Retrieved from http://eric.ed.gov/?id=ED500012 .

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), 29–45.

Millstone, J. (2012). *Teacher attitudes about digital games in the classroom*. New York: The Joan Ganz Cooney Center at Sesame Workshop in collaboration with BrainPOP.

Mullens, J. E., & Gayler, K. (1999). *Measuring classroom instructional processes: Using survey and case study field test results to improve item construction* (NCES 1999-08). Washington, DC: National Center for Education Statistics.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers and Education, 52*(1), 1–12.

PBS & Grunwald Associates, LLC (2011). *Deepening connections: Teachers increasingly rely on media and technology*. Washington, DC: PBS. Retrieved from http://www.pbs.org/teachers/research.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525–556.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.

Pianta, R. C., Hamre, B. K., Hayes, N., & Mintz, S. (2011). *Classroom assessment scoring system–secondary* (CLASS-S). Charlottesville, VA: University of Virginia.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*(6), 459–470.

Prensky, M. (2003). Digital game-based learning. *ACM Computers in Entertainment*, *1*(1), 1-4

Raudenbush, S. W., Bryk, A. S, & Congdon, R. (2004). *HLM 6 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.

Reese, D. D. (2007). First steps and beyond: Serious games as preparation for future learning. *Journal of Educational Multimedia and Hypermedia,16*(3), 283–300.

Reese, D. D. (2009). Structure mapping theory as a formalism for instructional game design and assessment. In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), *New frontiers in analogy research: Proceedings of the 2nd International Conference on Analogy* (pp. 394–403). Sofia, Bulgaria: New Bulgarian University Press.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science, 316*(5828), 1128.

Rideout, V.J., Foehr, U.G., & Roberts, D.F. (2010). *Generation M2: Media in the lives of 8- to 18-year-olds*. Menlo Park, CA: Henry J. Kaiser Family Foundation. Retrieved from http://kff.org/other/report/generation-m2-media-in-the-lives-of-8-to-18-year-olds/

Roschelle, J., Knudsen, J., & Hegedus, S. (2010). From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In *Designs for learning environments of the future* (pp. 233–262). New York, NY: Springer.

Schneps, M. H., Sadler, P. M., Woll, S., & Crouse, L. (1989). *A private universe* [Motion picture]. Annenberg Foundation/Corporation for Public Broadcasting Math and Science Project. Cambridge, MA: Harvard Smithsonian Center for Astrophysics.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.

Sheingold, K., Hawkins, J., & Char, C. (1984). "I'm the thinkist, you're the typist": The interaction of technology and the social life of classrooms. *Journal of Social Issues*, *40*(3), 49–61.

Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction, 24*(2), 261–289.

Slotta, J. D., Chi, M. T. H., & Joram, E. (1995). Assessing students' misclassification of physics concepts: An ontological basis for conceptual change. *Cognition and Instruction, 13*(3), 373–400.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, *3*(2), 115–163.

Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the What Works Clearinghouse (Phase I). *Educational Evaluation and Policy Analysis*, *32*(3), 351–371.

Squire, K., Barnett, M., Grant, J. M., & Higginbotham, T. (2004). Electromagnetism supercharged!: Learning physics with digital simulation games. In *Proceedings of* *the 6th International Conference on Learning Sciences* (pp. 513–520). International Society of the Learning Sciences.

Thomas, M. K., Barab, S. A., & Tuzun, H. (2009). Developing critical implementations of technology-rich innovations: A cross-case study of the implementation of Quest Atlantis. *Journal of Educational Computing Research, 41*(2), 125–153.

Venville, G. (2008). Knowledge acquisition as conceptual change: The case of a theory of biology. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on science and technology in early childhood education* (pp. 41–63). Greenwich, CT: Information Age Publishing.

Vosniadou, S. (2008). *International handbook of research on conceptual change.* New York, NY: Taylor & Francis.

Vosniadou, S. (2012). Reframing the classical approach to conceptual change: Preconceptions, misconceptions and synthetic models. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (Vol. 1., pp. 119–130). New York, NY: Springer.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher, 37*(8), 469–479.