

# Testing the Impact of a Pre-instructional Digital Game on Middle-Grade Students' Understanding of Photosynthesis

Katherine McMillan Culp · Wendy Martin · Margaret Clements · Ashley Lewis Presser

Published online: 17 August 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** Rigorous studies of the impact of digital games on student learning remain relatively rare, as do studies of games as supports for learning difficult, core curricular concepts in the context of normal classroom practices. This study uses a blocked, cluster randomized controlled trial design to test the impact of a digital game, played as homework prior to instruction, and associated supplemental instructional activities, on middle grade students' understanding of the process of photosynthesis. The role of the teacher as a potential moderator of the game's impact on student outcomes was also investigated, using Classroom Assessment Scoring System-Secondary Edition (CLASS-S) observations as a measure of instructional quality. Study findings demonstrate that the intervention did not have a significant impact on student understanding of photosynthesis. The interaction of treatment teachers' CLASS-S scores and students' average photosynthesis assessment scores approached significance. This study suggests that when digital games are used as a step in the process of learning difficult conceptual material, teachers may need support and guidance to make productive connections between in-game experiences and the target concepts.

**Keywords** Applications in subject areas · Pedagogical issues · Teaching/learning strategies · Photosynthesis · Digital games

---

K. M. Culp (✉) · W. Martin · M. Clements · A. Lewis Presser  
Center for Children and Technology, Education Development Center, Inc., 96 Morton St., 7th Floor,  
New York, NY 10014, USA  
e-mail: kculp@edc.org

W. Martin  
e-mail: wmartin@edc.org

M. Clements  
e-mail: pclements@edc.org

A. Lewis Presser  
e-mail: alewis@edc.org

## 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 et al. 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. 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 and 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 et al. 2013). Others maintain that the investment required to develop and deploy digital games justifies a desire for strong evidence of effectiveness (Connolly et al. 2009, 2012; de Freitas and 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 tests the impact of a digital game and associated classroom activities on middle-grade students’ understanding of the photosynthetic process. For the purposes of this study, we draw on Salen and Zimmerman’s classic definition of a game as “a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome” (2003, p. 96). This definition’s emphasis on the artificial conflict at the heart of gameplay helps us to distinguish digital games from simulations, which are designed to engage their users with veridical models of real systems, and which may be deeply concerned with balancing or managing systems but do not inherently involve the idea of conflict. We acknowledge that definitions and categorical boundaries around games and simulations, particularly as they pertain to learning, are contested and negotiable. See Juul (2003) and Young et al. (2012), for useful discussions of the definitions of games, simulations, and games for learning.

This study 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 et al. 1993; Schneps et al. 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.

## 1.1 Theoretical Background

### 1.1.1 *Prior Research on the Impact of Digital Games on Learning*

The evidence base regarding the general effectiveness of digital games as tools for learning remains limited and very diverse in focus (Clark et al. 2013; Connolly et al. 2009, 2012; Ke 2009; Young et al. 2012). 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.

Connolly et al.'s (2012), Clark et al.'s (2013), Ke's (2009) and Young et al.'s (2012) reviews all suggest that scaffolding of and reflection on the in-game experience are important to the effectiveness of the game as a learning tool. While these papers do not describe the types of scaffolding used in various studies in great detail, they 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*

Young et al. (2012) and Papastergiou (2009) note 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 et al. 2013; Sheingold et al. 1984). There are several important bodies of research on games for learning that have explored this broader context (Dede 2009; Ketelhut and Schifter 2011; Thomas et al. 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 et al. (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 et al. 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 et al. 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; Vosniadou 2008). This project primarily follows the work of Slotta and Chi (2006), 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 et al. 2011; Slotta et al. 1995). For alternative approaches to the issue, see, for example, Gupta et al. (2011), Hammer et al. (2011), Smith et al. (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.

## 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. Each includes a digital game and a series of related in-class, non-digital activities that can be integrated into regular instruction. This study poses three research questions.

1. In classrooms where teachers implement the *Exploring Photosynthesis* module (treatment group), 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 (control group)?
2. Do teachers in the treatment group implement the intervention with a high level of fidelity?
3. Does the quality of teachers' instruction moderate the impact of the *Exploring Photosynthesis* module on student understanding of the target concepts?

## 2 Methods

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 the *Exploring Photosynthesis* intervention to students whose teachers taught photosynthesis without the *Exploring Photosynthesis* intervention (the treatment and control groups). The study used a cluster randomized design because the nature of this classroom-level intervention precludes randomization of individual students. Instead, randomization occurred at the teacher level, and the study employed hierarchical linear modeling techniques to estimate effects of the intervention on students, allowing us to account for within-group commonalities among students who share the same teacher. Implementation of the intervention was staggered throughout the 2011–2012 school year because the time of year when each teacher taught photosynthesis varied across teachers.

### 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. Detailed information about the intervention, and all intervention materials, including the digital game, are available for review at <http://possibleworlds.edu.org>. The structure of this intervention draws on Bransford and Schwartz's "preparation for future learning" (1999) instructional model, 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. 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. Prior to instruction, treatment teachers took part in a 1-day professional development workshop that introduced them to the module and helped them plan for integrating the module components into their usual approach to teaching about photosynthesis. At the beginning of the teacher's normal unit on photosynthesis, students were assigned the game, *The Ruby Realm*, as homework, and were asked to play it for a minimum of 30 min, using Nintendo DSs provided as part of the intervention. *The Ruby Realm* is a 20-level adventure/maze game. Players navigate a vast cavern in search of missing friends, but discover that they have entered a hidden, treasure-filled world. Players are guided through the caves by Biobot Bob, a robot powered by artificial photosynthesis. As players progress, Bob helps them fend off hungry bats and angry vampires. Players must find light sources where Bob can generate the glucose he needs for power. Players use the light beams to break apart carbon dioxide and water molecules, and recombine the atoms to form glucose.

Teachers then integrated visuals drawn from the game into their discussion of the structure of glucose and the photosynthetic process during their normal instruction about photosynthesis. They also led students 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. Finally, students worked in groups on a consolidation activity in which they drew on their knowledge of photosynthesis to evaluate scientific claims made by journalists in a fictional tabloid.

The *Exploring Photosynthesis* game was 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. 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.

## 2.2 Random Assignment

The study recruited 42 teachers from 25 schools in New York State. The intervention was delivered by teachers to the students in their classrooms. Therefore, teachers are the unit of analysis and we randomly assigned teachers (and each teachers' students) to either the treatment group or the control group. Because the sample of participating schools varied significantly in terms of students' socio-economic background (as measured by the percentage of students eligible for free or reduced-price lunch), a variable known to significantly predict students' achievement, the study used a blocked randomization procedure that grouped teachers by the percentage of students in their school who qualified for free or reduced-price lunch (FRPL) prior to random assignment, then randomly assigned teachers within each block to either the treatment and control group. 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. Below we present descriptive statistics that

demonstrate the random assignment created equivalent groups of teachers and their students (see Sect. 2.4, which describes the teacher and student samples).

### 2.3 Measures

The study collected four types of data: (1) demographic and background data about schools, teachers, and students to describe the sample, to establish baseline equivalence between the treatment and control groups, and to use as covariates in the data analysis; (2) student outcome data to measure the impact of the *Exploring Photosynthesis* intervention; (3) measures of the implementation fidelity to describe how treatment teachers and students used the *Exploring Photosynthesis* intervention; and (4) data describing the quality of all teachers' instructional practice to examine whether instructional quality moderated the impact of the intervention.

#### 2.3.1 Demographic and Background Characteristics of Schools, Teachers, and Students

To describe the characteristics of participating schools, we downloaded publicly available data about each school from the New York State Department of Education website, including student enrollment, the number of students eligible for free or reduced-price lunch, the number of male and female students, and the number of students from different race/ethnic groups. We collected data on teachers' demographic and background characteristics using a survey that all participating teachers completed prior to teaching photosynthesis.

We collected data on students' background and demographic characteristics from three sources. In lieu of administering a pre-test to measure students' academic ability prior to the intervention, 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. 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 et al. 2010). This approach minimizes 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 et al. 2007). 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, we transformed scores into a common metric by mean centering each score using the state-level average score for the appropriate grade (sixth or seventh) and test (mathematics or language arts). The study obtained these data from students' administrative records. Other student data obtained from administrative records were gender, race/ethnicity, whether the student had an Individual Education Plan (IEP), and whether the student was classified as an English language learner (ELL).

To measure students' attitudes toward science, students also completed a survey that included 13 items drawn from the Test of Science Related Attitudes (or TOSRA, see Fraser 1981). For this study, we selected items from three of the original seven subscales: *leisure interest in science*, *enjoyment of science lessons*, and *attitude toward scientific inquiry*. We computed an overall TOSRA score for students by calculating their average response across all 13 items, if they answered at least two-thirds of the items (9 of 13 items). The internal reliability of students' scores, as measured by Cronbach's alpha, is 0.73.

### 2.3.2 Student Knowledge of Photosynthesis

To assess students' understanding of photosynthesis and chemical change, we administered a 33-item, paper-and-pencil assessment during a regular classroom period at the conclusion of the photosynthesis unit. We computed a total score for all students who completed at least half of the items by summing the number of correct responses. The internal reliability for the assessment, as measured by Cronbach's alpha, is 0.86.

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 during the 2010–2011 school year. Based on psychometric analysis of the pilot data, we then revised and shortened the assessment to include the final 33 questions.

### 2.3.3 Fidelity of Implementation

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.

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 and Gayler 1999). This approach maximized 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 after teachers completed the intervention. For each item, teachers indicated whether or not they implemented that aspect of the intervention. The items formed four composite scales. Two of the scales indicate the number of intervention components teachers reported implementing: *content coverage* (12 items) and *making links between the game and science content* (28 items). The other two scales represent *student engagement* (4 items) and *technical difficulties* (11 items).

### 2.3.4 Instructional Quality

We used the Classroom Assessment Scoring System-Secondary Edition (CLASS-S) observation framework to measure the quality of instruction provided by all participating teachers during a typical day of science instruction (Pianta et al. 2011). Observers rated teachers on 12 dimensions covering emotional support, classroom organization, instructional support, and student engagement. Scores on each of the 12 dimensions can range from low (1 or 2) to mid (3, 4, or 5) to high (6 or 7).

The procedures used to collect and score the CLASS-S observations for this study follow the recommendations of the researchers who developed the measure (Malmberg and Hagger 2009; Pianta and Hamre 2009; M. Stuhlman, personal communication, October 11, 2011). Study team members who conducted the CLASS-S observations were trained and



certified in the use of the instrument. Existing studies using the CLASS-S observation protocol document that the scores are highly reliable when collected by certified observers. Before each teacher began teaching the unit that included photosynthesis, trained researchers observed the participating classroom on one occasion for 40 min. This allowed for two 15-min observation intervals, each followed by 5 min for coding. Pianta and Hamre (2009) have demonstrated that relatively few observations can effectively discriminate among teachers, because between-teacher variation is typically much greater than within-teacher variation over time.

We created an overall CLASS-S score for each teacher by computing the average rating across teachers' scores on each of the 12 dimensions. 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 0.91.

## 2.4 Sample

### 2.4.1 Teacher Sample

One of the 42 recruited teachers withdrew from the study prior to teaching photosynthesis. Of the final sample of 41 teachers, 21 were randomly assigned to the treatment group and 20 were randomly assigned to the control group. Participating teachers taught in a total of 25 primarily middle-grades schools, with three of the schools spanning grades K–8. The schools in which teachers worked varied considerably in terms the percentage of students eligible for free or reduced-price lunch ( $M = 32.2$ ,  $SD = 24.7$ , range 0–84) and student enrollment ( $M = 758.7$ ,  $SD = 263.8$ , range 198–1,153 students). The racial/ethnic group composition of the schools also varied ( $M_{White/non-Hispanic} = 65.6$ ,  $SD_{White/non-Hispanic} = 25.2$ , percentage range 15–98;  $M_{Black/non-Hispanic} = 11.3$ ,  $SD_{Black/non-Hispanic} = 11.0$ , percentage range 0–30;  $M_{Hispanic} = 17.0$ ,  $SD_{Hispanic} = 17.4$ , percentage range 0–62).

Participating teachers had an average of 14.1 years' total teaching experience ( $SD = 7.1$ ) and 12.1 years of teaching middle-school science ( $SD = 7.2$ ). Over 90 % had earned a Masters' degree—with most of teachers having focused their graduate study on science or science education—and were certified in New York State to teach biology. Table 1 presents descriptive statistics for the teacher sample overall, as well as by experimental group (treatment or control). There were no significant differences between teachers in the treatment and control groups on any of the measured demographic or background variables.

### 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 targeted either “accelerated” students or students with learning difficulties were excluded. A total of 914 students participated in the study. Table 2 presents descriptive statistics for the student sample as a whole and separately for the treatment and control groups. There were no statistically significant differences between the students in treatment and control classrooms on demographic characteristics or state standardized test scores.

**Table 1** Characteristics of participating teachers for full sample, treatment group, and control group

Characteristic	Full sample Mean ( <i>SD</i> )	Treatment Mean ( <i>SD</i> )	Control Mean ( <i>SD</i> )
Years of experience overall	14.1 (7.1)	12.1 (7.2)	16.3 (6.5)
Years of experience teaching middle-school science	12.1 (6.8)	10.5 (7.3)	13.6 (6.2)
	Number (%)	Number (%)	Number (%)
Master's degree earned	38 (93)	20 (95)	18 (90)
Major field of study for highest degree completed			
Biology, other science, or science education	28 (68)	15 (71)	13 (65)
Humanities, social science, or education (not science-specific)	12 (29)	5 (24)	7 (35)
Certified in New York State to teach biology	38 (92)	20 (95)	18 (90)

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

**Table 2** Student characteristics for full sample, treatment group, and control group

Characteristic	Full sample Mean (SD)	Treatment Mean (SD)	Control Mean (SD)
Age in years <sup>a</sup>	12.7 (0.5)	12.6 (0.5)	12.7 (0.5)
Mean-centered state mathematics test score <sup>b</sup>	-0.7 (26.2)	-1.2 (24.9)	-0.1 (27.6)
Mean-centered state language arts test score <sup>c</sup>	1.0 (15.0)	-0.3 (14.0)	2.4 (16.0)
TOSRA score	3.3 (0.6)	3.4 (0.5)	3.3 (0.6)
	Number (%)	Number (%)	Number (%)
Race/ethnicity ( $n = 912$ ) <sup>d</sup>			
White/non-Hispanic	571 (63 %)	302 (63 %)	269 (62 %)
Black/non-Hispanic	121 (13 %)	60 (13 %)	61 (14 %)
Hispanic	178 (20 %)	90 (19 %)	88 (20 %)
Male ( $n = 914$ ) <sup>e</sup>	452 (50 %)	238 (50 %)	214 (49 %)
Has an IEP ( $n = 870$ ) <sup>f</sup>	107 (12 %)	63 (15 %)	44 (10 %)
Classified as ELL ( $n = 893$ ) <sup>g</sup>	56 (6 %)	17 (4 %)	39 (9 %)

This table reports the descriptive statistics for the student sample prior to multiple imputation. There are no statistically significant differences between treatment and control groups

IEP Individualized education plan, ELL English language learner, TOSRA test of science related attitudes

<sup>a</sup> Age is calculated as the difference between each students' birthdate and the mid-point of the 2011–12 academic year

<sup>b</sup> Students' state mathematics scores are mean-centered using the state average mathematics score for the grade during which students took the exam

<sup>c</sup> Students' state language arts scores are mean-centered using the state average language arts score for the grade during which students took the exam

<sup>d</sup> The  $n$  reported in the parenthesis refers to the number of students for whom we have race/ethnic group data. The  $n$  and percentage reported in the table body represent the number of students in each race/ethnic group. To protect the confidentiality of participating schools and students, in cases where the number of students of a race or ethnic group is relatively small, we do not report the  $n$  or percentage

<sup>e</sup> The  $n$  reported in the parenthesis refers to the number of students for whom we have data on gender. The  $n$  and percentage reported here represent males only

<sup>f</sup> The  $n$  reported in the parenthesis refers to the number of students for whom we have data on IEP status. The  $n$  and percentage reported here represent students who have an IEP

<sup>g</sup> The  $n$  reported in the parenthesis refers to the number of students for whom we have data on ELL classification. The  $n$  and percentage reported here represent students who are classified as ELL

## 2.5 Statistical Power

Given the final sample of teachers ( $N = 41$ ) and their students ( $N = 914$ , an average of approximately 22 students per teacher), the study had the statistical power to detect a mean effect size of 0.24 of a standard deviation. In other words, the study's sample was large enough to find a significant impact of the *Exploring Photosynthesis* intervention if the difference in outcomes scores of students in treatment teachers' classrooms was at least 0.24 of a standard deviation larger than the outcome scores of students in control teachers classrooms. Reporting the difference in outcome scores in terms of a standard deviation is a common metric for describing the magnitude of an educational intervention in statistical terms.

### 3 Results

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 et al. 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.

#### 3.1 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, as students with missing data may be systematically different from students without missing data (Peugh and Enders 2004). To address this issue, we used a multiple imputation strategy to create ten versions of the data set in which all of the missing values are predicted using the existing values for other variables (Enders 2010; Song and Herman 2010). We conducted the data analysis to investigate whether teachers’ use of the *Exploring Photosynthesis* module resulted in higher students assessment scores (research question 1) using all ten data sets and then averaged the results, which are reported here. We also used the imputed data to examine whether instructional quality moderated the impact of the intervention (research question 3).

#### 3.2 Impact of the *Exploring Photosynthesis* Module on Student Knowledge of Photosynthesis Assessment Scores

As described above, to answer the first research question—do students of teachers in the treatment group demonstrate a significantly better understanding of how photosynthesis occurs and how mass and energy are conserved during chemical changes than students of teachers in the control group—we used a multi-level regression to estimate the impact of the intervention on students’ photosynthesis assessment scores. The Level 1 (student-level) model was:

$$\begin{aligned} \text{Assessment}_{ij} = & \beta_{0j} + \beta_{1j}(\text{StateMath}_{ij}) + \beta_{2j}(\text{StateLangArts}_{ij}) + \beta_{3j}(\text{TOSRA}_{ij}) \\ & + \beta_{4j}(\text{Male}_{ij}) + \beta_{5j}(\text{Age}_{ij}) + \beta_{6j}(\text{Minority}_{ij}) + \beta_{7j}(\text{IEP}_{ij}) + \beta_{8j}(\text{ELL}_{ij}) + \varepsilon_{ij} \end{aligned}$$

where  $\text{Assessment}_{ij}$  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 precision of the estimate of the intervention’s impact.  $\text{StateMath}_{ij}$  and  $\text{StateLangArts}_{ij}$  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 ELL.

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:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}(Treatment_j) + \gamma_{02}(PercentFRPL_j) + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \\ \beta_3 &= \gamma_{30} \\ \beta_{4j} &= \gamma_{40} \\ \beta_{5j} &= \gamma_{50} \\ \beta_{6j} &= \gamma_{60} \\ \beta_{7j} &= \gamma_{70} \\ \beta_{8j} &= \gamma_{80}\end{aligned}$$

Specifically,  $Treatment_j$  indicates whether or not a teacher was in the treatment group (0 = control group, 1 = treatment group) and  $\gamma_{01}$  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 ( $PercentFRPL_j$ ).

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,  $\gamma_{01} = -1.01$ ,  $p = 0.21$ . Table 3 presents the coefficients, standard errors, and  $p$  values for each of the Level 1 and Level 2 predictors.

### 3.3 Implementation of the *Exploring Photosynthesis* Intervention

#### 3.3.1 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, 63 % played for 30 min or more. The average gameplay time was 47.4 min ( $SD = 31.5$ ), and ranged from 1 to 168 min. This average time was well above the assigned benchmark of 30 min of gameplay time that students were expected to complete as homework. Table 4 presents the breakdown of student gameplay by minutes.

### 3.3.2 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 11.6 of the 12 items ( $SD = 0.8$ , range 9–12),

**Table 3** Coefficients, standard errors, and  $p$  values for all variables included in the multi-level regression model estimating the impact of the intervention

	Coefficient	SE	$p$
Intercept	21.82	0.57	<0.001
Level 2 (teacher) covariates			
Condition (treatment or control)	−1.01	0.79	0.206
Percentage of students in the school eligible for free or reduced-price lunch	−0.05	0.02	0.044
Level 1 (student) covariates			
Male	−1.10	0.44	0.012
Age	−0.40	0.45	0.370
Minority	−0.45	0.56	0.421
IEP	−0.04	0.79	0.958
ELL	−1.82	1.13	0.106
Mean-centered state mathematics test score	0.08	0.01	<0.001
Mean-centered state language arts test score	0.11	0.02	<0.001
TOSRA	1.11	0.35	0.002

*IEP* individualized education plan, *ELL* English language learner, *TOSRA* test of science related attitudes

**Table 4** Number and percentage of students by minutes of gameplay categories

Minutes of gameplay categories	Number	Percentage
Missing gameplay data	90	18.9
0	21	4.4
>0–10	38	8.0
>10–20	44	9.2
>20–30	42	8.8
>30–40	32	6.7
>40–50	46	9.6
>50–60	45	9.4
>60–70	39	8.2
>70–80	24	5.0
>80–90	16	3.4
>90	40	8.4

**Table 5** Number of items, means, ranges and standard deviations for subscales of teacher fidelity logs

Subscale	Number of items in subscale	Mean	Min/max	SD
Content coverage	12	11.6	9–12	0.8
Student engagement	4	3.2	2–4	0.8
Technical difficulties (reverse scale)	11	3.2	1–6	1.5
Making links between science content and the game	28	21.1	14–28	4.0

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 ( $SD = 0.8$ , range 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 1–6), 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 14–28). 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. Table 5 presents the mean, standard deviation, and range for each subscale.

### 3.4 Quality of Teachers' Instruction as a Moderator of Student Outcomes

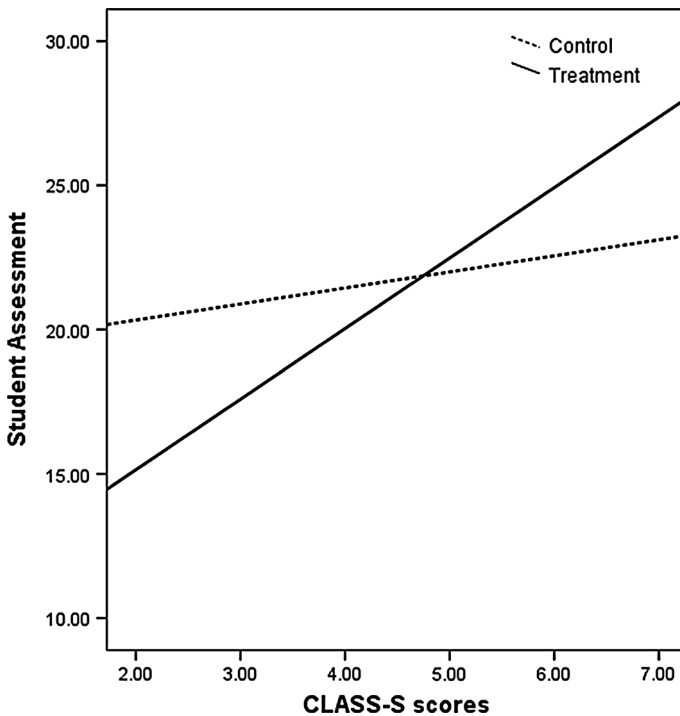
We conducted a second, exploratory multi-level regression analysis to answer research question 3, which asks whether the impact of the intervention on students' assessment scores varied depending on the quality of instruction provided by students' classroom teachers. This required adding two additional variables to the Level 2 model: each teacher's CLASS-S score collected during the classroom observation and a variable representing the *Treatment*  $\times$  *CLASS-S* interaction.

This analysis 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*  $\times$  *CLASS-S* interaction term,  $\gamma_{04} = 1.72$ ,  $p = 0.07$ . Table 6 presents the coefficients, standard errors, and  $p$  values for the model. One way to illustrate this finding is to look at a scatterplot of teachers' CLASS-S scores by their students' average assessment scores separately for the treatment and control groups. In Fig. 1, the solid line represents the association between treatment teachers' CLASS-S scores 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 the teachers' instructional quality was more strongly associated with students' assessment scores in the classrooms where teachers used the *Exploring Photosynthesis* module than in the classrooms where teachers did not use the module.

**Table 6** Coefficients, standard errors, and *p* values for all variables included in the multi-level regression model testing whether the impact of the intervention is moderated by instructional quality

	Coefficient	SE	<i>p</i>
Intercept	21.82	0.54	<0.001
Level 2 (teacher) covariates			
Condition (treatment or control)	-1.01	0.76	0.191
Percentage of students in the school eligible for free or reduced-price lunch	-0.05	0.02	0.020
CLASS-S	-0.93	0.69	0.186
Condition × CLASS-S interaction	1.72	0.92	0.070
Level 1 (student) covariates			
Male	-1.12	0.37	0.002
Age	-0.42	0.43	0.340
Minority	-0.42	0.44	0.343
IEP	0.02	0.74	0.977
ELL	-1.73	0.90	0.057
State math test score	0.08	0.01	<0.001
State language arts test score	0.11	0.02	<0.001
TOSRA	1.09	0.34	0.002

*CLASS-S* Classroom Assessment Scoring System-Secondary Edition, *IEP* individualized education plan, *ELL* English language learner, *TOSRA* test of science related attitudes

**Fig. 1** Association of CLASS-S scores and mean student assessment scores by teacher, for treatment and control groups



## 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 features and limitations of the intervention and the study design that may have contributed to this outcome.

As Connolly et al. (2012) have argued, the efficacy of individual digital games, like other digital tools for learning, is highly dependent on the pedagogical and material environment in which they are used. This study was designed to expand our knowledge base about the intersection of a particular digital game and the context of learning, by conducting the study in naturalistic classroom conditions, embedding the digital game in a longer sequence of linked curricular activities, and conducting an exploratory analysis of the moderating impact of teachers' instructional quality on the student outcome of interest.

Unlike many prior approaches to studying digital games and learning (Gee 2003), this study tested an intervention that did not presume that learning would occur exclusively or even primarily through gameplay itself. Rather, the design and structure of the *Exploring Photosynthesis* module was intended to position gameplay as a first step in a multi-step instructional process. This approach followed on Bransford and Schwartz's (1999) "preparation for future learning" instructional model, but also more broadly on research on conceptual learning and social cognition that demonstrates that students master and retain conceptual knowledge when they are provided with multiple and varied opportunities to articulate, negotiate and rehearse new knowledge over time (Palincsar 1998).

At the same time, *Exploring Photosynthesis* was designed to relate digital gameplay to middle-grade science instruction in a way that would coordinate with and not disrupt teachers' normal instructional sequences and classroom practices. The results of this study suggest that this design choice was unrealistic. First, prior research on instructional interventions that follow the "preparation for future learning" model have not explored whether the teachers' familiarity with the preparatory experience was relevant to student outcomes. The results of this study do not test the relevance of this factor directly, but suggest that teacher familiarity is relevant, as a precursor to necessary, skilled linking of the preparatory experience to the target concept. Second, research with other content domains and other age groups has demonstrated that explicit mapping techniques are needed to support robust analogical reasoning (Richland et al. 2007). 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.

Treatment teachers' self-reports about their fidelity of implementation suggest that providing effective instructional support for the intervention was difficult, or possibly did not seem relevant to these teachers. 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. The majority of treatment teachers reported that they did not make this or other similar explicit connections between the game and the target concepts. Anecdotal evidence suggests that teachers may not have felt that their students' prior in-game experience of breaking apart water and carbon dioxide molecules, and building the glucose molecule from their component parts, was relevant to

their goals, such as having students simply memorize the chemical equation for photosynthesis. Anecdotal evidence also suggests that another possible reason for teachers not making these types of connections is that teachers were unlikely to have played the game extensively themselves, and may not have been comfortable using the game as a point of reference.

Exploratory analyses of the CLASS-S data build on this descriptive evidence. As described above, there was no main effect between teachers' CLASS-S scores and their mean, class-aggregate assessment scores—that is, student performance on the photosynthesis assessment was not related to teachers' instructional quality, as measured by the CLASS-S, across the treatment and control groups. However, when CLASS-S scores were treated as a moderating variable for the treatment and control groups separately, they show that while CLASS-S continues to play no predictive role for student outcomes in the control group, there is a clear interaction between CLASS-S scores and student outcomes in the treatment group, with higher-scoring teachers having higher-performing students. In other words, teachers who both had high levels of instructional skill *and* who used *Exploring Photosynthesis* with their students did produce better student outcomes than teachers with similar levels of instructional skill but who did not use *Exploring Photosynthesis*. This could suggest that the more skillful teachers within the treatment group may have been able to guide and support students' analogical reasoning about the game and the target concepts in ways that had a critical impact on what students learned.

This conclusion is consistent with findings from many studies of other kinds of digital tools to support student learning in middle grade science. Linn et al. (2004), for example, have demonstrated in detail that carefully designed digital tools to support inquiry learning can only succeed when their use is facilitated by teachers who are prepared to guide students in the kinds of thinking and exploration the tools are designed to support. Similarly, the *Exploring Photosynthesis* digital game sought to support students in developing a robust understanding of specific aspects of the photosynthetic process, but the intervention as a whole did not invite teachers to examine whether and how to align those goals with the goals of their existing coverage of photosynthesis.

Creating more effective supports for teachers who want to use *The Ruby Realm* as a pre-instructional support for learning will require addressing two types of teacher needs. First, teachers will need efficient, accessible ways to become familiar with the core mechanics and key features of the game and to identify the points of connection between the game and their own goals for student learning. Second, teachers will need help developing a repertoire of instructional moves that can help students articulate and explore connections between their in-game experiences and the target concepts. Regarding becoming familiar with the game itself, we know anecdotally from teachers who participated in this study and in prior field tests that preparing for using digital games in the classroom is time-consuming and sometimes intimidating. Playing through a game for long enough to become familiar with it requires more time investment than, for example, selecting an illustration to show to the class. This is particularly true for the many participating teachers who did not consider themselves to be “gamers” and expressed trepidation about exploring the game at all. We view both additional materials and new strategies for tapping students' expertise as potentially effective approaches to this challenge. For example, brief Flash animations that duplicate the core mechanic of *The Ruby Realm* are already available for use during instruction. This same animation could be used to support a tutorial for teachers, which would focus less on successful gameplay and more on mapping out the relationships between the design and sequencing of the game and the target concepts it is intended to support. The study team has also repeatedly observed students acting as expert consultants

to teachers, explaining how the game works and how to progress through it. Teachers could be encouraged to formalize this advisory role for students, looking to them for expertise in understanding both how to play the game and how to make sense of it as a tool for learning.

Regarding instructional moves to support effective analogical mapping between games and target concepts, it will be necessary to confront the challenge of helping teachers focus their instruction on sustained exploration of analogies that are tightly aligned to learning targets. Several lines of research, including detailed work by Roth (2013), have demonstrated that science teachers use analogical language frequently, and often informally, as they try to help students make sense of novel concepts. Therefore the challenge is not so much about encouraging teachers to use analogies in their teaching, but helping them to select potentially powerful and broadly accessible analogies and to articulate and unpack them in more structured and routinized ways, so that they can become shared, explicit tools to support students' emergent understanding. In a current study that builds on the work reported here, the authors are drawing on work in mathematics teaching by Richland et al. (2007), and collaborating with middle school science teachers, to identify the analogical mapping practices that might be most effective for supporting middle grade science learning.

#### 4.1 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, conclusions about the quality of implementation in the treatment group are based on limited data sources.

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.

#### 4.2 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 (Richland et al. 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. 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.

**Acknowledgments** We gratefully acknowledge support for this work from the Institute of Education Sciences, U.S. Department of Education, Grant No. R305C080022. The research team would like to thank all of the teachers and students who participated in this study, as well as our advisory board members for their ongoing guidance and insight.

## 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(1), 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, NY: 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. In J. Pettersen (Ed.), *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), 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. <http://www.sri.com/work/projects/glasslab-research>.
- 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 & Education*, 59(2), 661–686.
- 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.
- de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, 46(3), 249–264.
- Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, 323(5910), 66–69.
- Deke, J., Dragoset, L., & Moore, R. (2010). *Precision gains from publicly 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: Routledge.
- Duit, R. (2009). *Bibliography—STCSE: Students' and teachers' conceptions and science education*. <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: Australian Council for Educational Research.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave Macmillan.
- 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.
- International Telecommunications Union. (2013). *Measuring the information society*. ITU Publication No. 254. [http://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2013/MIS2013\\_without\\_Annex\\_4.pdf](http://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2013/MIS2013_without_Annex_4.pdf).
- Juul, J. (2003). The game, the player, the world: Looking for a heart of gameness. In M. Copier & J. Raessens (Eds.), *Level up: Digital games research conference proceedings* (pp. 30–45). Utrecht: Utrecht University.
- 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). Hershey, 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 & Education*, 56(2), 539–546.
- Koschmann, T., Hall, R. P., & Miyake, N. (Eds.). (2013). *CSSL 2*. New York, NY: Routledge.
- Lenhart, A., Kahne, J., Middaugh, E., Macgill, A. R., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics*. <http://www.pewinternet.org/Reports/2008/Teens-Video-Games-and-Civics.aspx>.
- Linn, M. C., Davis, E. A., & Bell, P. (Eds.). (2004). *Internet environments for science education*. London: Routledge.
- 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.
- 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 Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, 49, 345–375.
- Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1), 1–12.
- PBS & Grunwald Associates, LLC. (2011). *Deepening connections: Teachers increasingly rely on media and technology*. <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.
- 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*. <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 M. Jacobson & P. Reimann (Eds.), *Designs for learning environments of the future* (pp. 233–262). New York, NY: Springer.
- Roth, W. M. (2013). *On meaning and mental representation: A pragmatic approach*. Rotterdam: Sense Publishers.
- Salen, K., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.
- 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* (Vol. 6, pp. 513–520).
- 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.
- Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., et al. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, 82(1), 61–89.