Digital Games for Learning: A Systematic Review and Meta-Analysis

Executive Summary

DRAFT

August 2013

Developed by SRI International with funding from the Bill & Melinda Gates Foundation.

This report is based on research funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Suggested citation:
Clark, D., Tanner-Smith, E., Killingsworth, S ., Bellamy, S. (2013). *Digital Games for Learning: A Systematic Review and Meta-Analysis (Executive Summary)*. Menlo Park, CA: SRI International.

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
Phone: 650.859.2000

www.sri.com

Digital Games for Learning: A Systematic Review and Meta-Analysis

Executive Summary

Prepared By:

Douglas B. Clark
Emily E. Tanner-Smith
Stephen Killingsworth
Shara Bellamy

DRAFT

August 2013

Contents

Executive Summary

Background

In 2006, the Federation of American Scientists issued a widely publicized report stating their belief that games offer a powerful new tool to support education. The report encouraged private and governmental support for expanded research on complex gaming environments for learning. A special issue of *Science* in 2009 echoed and expanded this call (Hines, Jasny, & Mervis, 2009), as has the National Research Council's report on games and simulations for science learning (NRC, 2010). However, these reports also underscore that solid evidence of the contributions of games to learning is sparse.

Much of the early debate about digital games for education centered on whether games are good or bad for education. But that question is overly simplistic. The National Research Council's report on laboratory activities and simulations (NRC, 2005) makes clear that the design and not merely the medium of a physical or virtual learning activity determines its efficacy. Digital games are a medium with certain affordances and constraints, just as physical labs and virtual simulations are media with certain affordances and constraints. Simulations and digital games actually share many similarities in this regard. Although there are multiple definitions for games (cf., Salen & Zimmerman, 2004), the key characteristics differentiating games from simulations involve the explicit inclusion of (a) rules for engaging with the simulation, (b) goals for players to pursue, and (c) means for indicating players' progress toward those goals (Clark, Nelson, Sengupta, & D'Angelo, 2009; Klopfer, Osterweil, & Salen, 2009; Koster, 2004; McGonigal, 2011).

Properly designed, these features of games can provide powerful affordances for motivation and learning. Individual studies have shown, for example, that well-designed games can promote conceptual understanding and process skills (e.g., Annetta, Minogue, Holmes, & Cheng, 2009; Clark et al., 2011; Hickey, Ingram-Goble, & Jameson, 2009; Ketelhut, Dede, Clarke, & Nelson, 2006; Klopfer, Scheintaub, Huang, Wendal, & Roque, 2009; Moreno & Mayer, 2000, 2004), can foster a deeper epistemological understanding of the nature and processes through which science knowledge is developed (e.g., Barab et al., 2007; Neulight, Kafai, Kao, Foley, & Galas, 2007), and can produce gains in players' willingness and ability to engage in scientific practices and discourse (e.g., Barab et al., 2009; Galas, 2006; McQuiggan, Rowe, & Lester, 2008). Leveraging these affordances, however, appears to depend on careful design (Clark, Martinez-Garza, Nelson, D'Angelo, & Bellamy, submitted).

In the summer of 2012, the Bill & Melinda Gates Foundation, in cooperation with the MacArthur Foundation, made a significant investment to establish and support the Games Assessment and Innovation Lab (GlassLab), which includes top game developers, assessment experts, and researchers from multiple fields. The goal of GlassLab is to transform learning and formative assessment through digital games. During the planning stages of the investment, the program was divided into two teams — an investment in a program team (GLASSLab) and a second investment in a research team (GLASSLab-Research) — to mitigate conflict of interest and guarantee independent validation of assessments developed by the program. It was determined by all those associated with GlassLab that the GlassLab development team would design and develop state-of-the-art game-based assessments. Independently, GlassLab-Research would conduct research on the qualities, features, validity, reliability, and effectiveness of the games and assessments that are embedded within the gaming environments produced by GlassLab. The meta-analysis and systematic review of the simulation literature described in this report is part of the larger GlassLab-Research project, which also includes a parallel meta-analysis and systematic review of simulations and science learning (D'Angelo, Rutstein, Harris, Bernard, Borokhovski, & Haertel, 2013). The goals for the current study on digital games and learning are threefold:

1. To look specifically at digital games and learning across disciplines and learning outcome types

2. To analyze the impact of learning outcomes based on constituent design features as well as on the level of the game versus traditional instruction so that future development and research can build on that foundation

3. To more thoroughly cover eligible studies across fields so that the results represent this diversity and a large enough sample of studies can be collected to reliably explore specific questions of design.

Objectives

This meta-analysis synthesizes research on learning in digital games for students of K–12 age as well as students enrolled at postsecondary educational institutions. The studies were located in electronic bibliographic databases from engineering, computer science, medical, natural sciences, and social sciences fields. Learning is defined and categorized broadly in terms of the cognitive, intrapersonal, and interpersonal clusters of 21st century competencies outlined in the NRC's recent report *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (NRC, 2012). The research questions this study addresses are:

1. What are the effects of digital games on learning for K–16 students?

2. How do these effects vary by learning outcome type in alignment with the categorizations of the recent NRC *Education for Life and Work* report (NRC, 2012)?

3. How do these effects vary by game type?

Search Strategy

Meta-analysis database searches involve a balance between precision (retrieving only studies that are likely to meet the subsequent eligibility criteria) and sensitivity (retrieving all studies that might possibly meet subsequent eligibility criteria by not inadvertently missing studies at the database search stage). The latter approach creates more work at the level of title and abstract screening but is more complete. We wanted to maximize sensitivity in our search. Our database search term criteria simply specified that *game* or *games* needed to be in the abstract or title. All other potential terms were deemed likely to inadvertently cut out otherwise eligible studies.

Research on games for learning spans many fields. To ensure that we were maximally sensitive in our meta-analysis, we therefore searched the following fields: engineering, computer science, medical, natural sciences, and social sciences. To do so, we searched the following databases and subdatabases for the terms *game* or *games* in the title or abstract:

- ISI Web of Science (SSI, SSSI)
- Proquest (ERIC, PsycINFO, Soc Abstracts, Social Services Abstracts)
- PubMED
- Engineering Village (Inspec, Compendex)
- IEEE Xplore.

We also checked the reference lists in narrative reviews and meta-analyses, as well as in those studies identified as eligible for inclusion in the meta-analysis.

Article Selection Criteria

This systematic review and meta-analysis explores the effects of digital games on cognitive, affective, and other learning-related outcomes. For inclusion, studies must describe an eligible digital game program directed toward an eligible participant sample and report information on at least one eligible outcome variable that permits the calculation of an effect size. Each of the eligibility criteria we used are detailed below.

Digital Game

To be eligible, the journal article author or authors must explicitly designate the environment as a game. The study must focus on the effects of a digital game on an eligible outcome. Games need not to have been explicitly designed for learning.

Participants

All study participants must be in the K–12 age range of 6 to 18 years of age (whether or not the study was conducted in the context of a K–12 institution), be students in a K–12 institution, or be students enrolled in a postsecondary educational institution. We chose this age range to align with the preparation for and success in postsecondary education.

Research Designs

Although we originally intended to include all types of research designs, the large amount of literature identified made this impossible given time and resource constraints. Thus, to be eligible for the current meta-analysis, only randomized controlled trial and quasi-experimental designs were eligible for inclusion.

Learning Outcomes

Eligible studies must report information on at least one eligible outcome related to learning, aligned with the recent National Research Council's report *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (NRC, 2012). Eligible outcomes broadly include those related to cognitive competencies, intrapersonal competencies, or interpersonal competencies.

Date of Publication

Eligible studies should be relatively modern to reflect the current state of digital game design. Therefore, the date of publication or reporting of the study must be from 2000 to 2012.

Study Site and Language

The study must be published in English (but not necessarily conducted in English or an English- speaking country).

Effect Sizes

To be eligible for the meta-analysis, the study must report information sufficient for calculation of both posttest and pretest effect sizes comparing two groups. The pretest effect size is required to estimate the difference-in-difference effect size metric, which is used to control for baseline differences between groups. This is particularly important when including quasi-experimental studies that may have baseline differences between groups. The magnitudes of effect sizes can be categorized into different groups. For Cohen (1988), one way to think about categorizing effect sizes was that small effect sizes (.2 to .3) are those that are barely detectable by the naked eye, medium effect sizes (.4 to .6) are those that can be detected visually, and large effect sizes (greater than .7) are those that could not be missed by a casual observer. It is important to remember that effect sizes are dependent not just on the mean difference between two groups, but also the standard deviation of those groups. For example, there is an average height difference between 15- and 16-year old girls, but there is a lot of variation within each of those age groups, so this would correspond to a relatively small effect size. However, when comparing 13- and 18- year old girls, there is a much larger average height difference, and even with a similar amount of variation within each age group, this would correspond to a larger effect size.

Publication Status

Only peer-reviewed journals articles are eligible for inclusion, with the goal of concentrating on research that has been subjected to the scrutiny and quality requirements of the peer-review process.

Literature Search

All the literature searches were conducted in September 2012. Overall, the literature search yielded 61,887 net hits (after accounting for 7,476 duplicates that were initially identified in EndNote) (Table 1). Most of the articles were identified in ISI Web of Science (41,710) or PubMed (14,685), although Proquest, Engineering Village, and IEEE Xplore also yielded several thousand results.

The 61,887 articles were scanned for eligibility coding (Table 2). A majority were initially screened out at the title level (58,111). A total of 3,776 abstracts were next screened for eligibility, and 726 articles were screened and read in full text to determine final eligibility (4 "rogue" articles could not be obtained for the analysis beyond the abstract level). Most were ineligible for inclusion in the meta-analysis because of inadequate research designs (i.e., many were concept pieces that did not empirically examine the effect of a digital game). After the full-text screening, 80 articles based on analyses of 77 unique data samples ultimately met the eligibility criteria, including sufficient information to calculate effect sizes, and were included in the final meta-analysis. (See Figure 1 for the number of studies at each stage of the eligibility and coding process.) To clarify, some articles included more than one data sample, usually referred to as multiple "studies," but some articles reported on the same data sample as other articles. The number of articles and the number of distinct data samples are therefore not the same. Citations for the 80 reports from which effect sizes were calculated are in Appendix A of the full report.

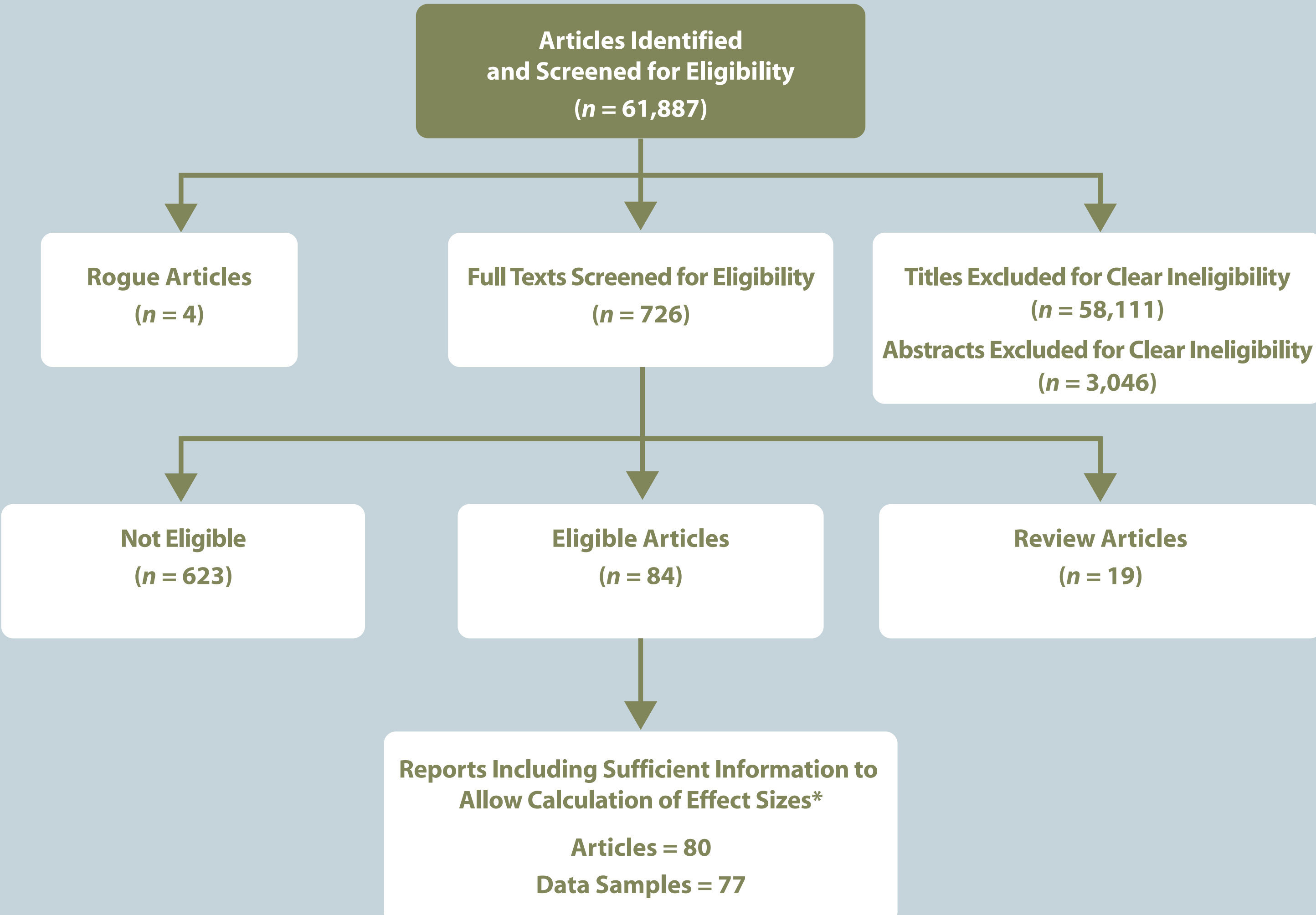
Table 1. Literature Search Results

Literature Database	Number of Articles
ISI Web of Science	41,710
Proquest	5,255
PubMED	14,685
Engineering Village	5,667
IEEE Xplore	2,038
Identified in reference lists	8
Gross total	69,363
Duplicates deleted in EndNote	-7,476
Net total from search	**61,887**

Table 2. Eligibility Screening Results

Screening Process	Number of Articles
Net total from database searches	**61,887**
Screened out at title level	**- 58,111**
Screened out at abstract level	**- 3,046**
Rogue articles (requested through interlibrary loan for inclusion in final report but unavailable for this report)	**-4**
Screened out at full-text level	**- 642**
Literature review articles	19
Ineligible research design	522
Ineligible participant population	41
Ineligible game/intervention	15
Ineligible outcome type	11
Ineligible language/publication year	8
Lack of statistical information for effect size calculations	26
Full-text articles coded	**84**
Coded articles including sufficient information for effect size calculations	**80**
Total distinct data samples included across the articles that included sufficient information for effect size calculations (Some articles included more than one data sample, usually referred to as multiple "studies," but some articles reported on the same data sample as other articles.)	**77**

Figure 1. Study Article Identification Process



*Some articles included more than one data sample, usually referred to as multiple "studies," but some articles reported on the same data sample as other articles. Thus, the number of reports and the number of data samples are not the same.

Study Characteristics

Most of the studies included in the meta-analysis posttested students an average of 4 weeks after the end of the game, although this ranged from 0 to 27 weeks (Table 3). On average, the digital games consisted of 14 sessions, played over the course of 40 days. Only 8% of the studies provided evidence of possible implementation problems in delivering the digital game (e.g., technical difficulties that led to shorter game duration). In terms of game players, most of the studies (82%) used games that were single-player and had no collaborative or competitive elements.

Almost all the games were played on computer platforms (92%), although a few were conducted on handheld, tablet, augmented reality, or virtual reality platforms (Table 4). Games were classified on the basis of the integration of the learning mechanic and the core game mechanic. The learning mechanic can be defined as the primary aspects and interactions within the game intended to support players in learning the target outcomes. The core game mechanic can be defined as the aspects and interactions within the game that are ostensibly designed to be the most interesting and engaging aspects. Very few games involved fully extrinsic integration of the learning mechanic and core mechanic, a controversial but simple design made famous by the Math Blaster® series of educational games where the learning mechanic of solving equations was completely separated from the game mechanic of blasting space trash, the core intended motivator. Most of the games were coded as "intrinsic by default" (62%), which meant that the learning mechanic was integrated directly into the core game mechanic but that there were no other elaborate game mechanics because of the simplicity of the game design. Tetris® would represent an example for this code, as would simple educational games where the player answers questions for points or rewards. Another 36% of the games represented intrinsic integration of learning mechanics and game mechanics in game designs involving more elaborate game mechanics.

Using the broad learning outcome domains from the NRC report on 21st century learning skills, 83% of outcomes measured were cognitive competencies, 16% were intrapersonal competencies, and less than 1% involved interpersonal competencies (Table 5). In terms of the narrow outcome domains from the NRC report, the majority of effect sizes were for learning outcomes that were measures of knowledge (66%) or cognitive processes/strategies (14%), followed by work ethic/conscientiousness (10%) and positive core self-evaluation (4%).

Table 3. Descriptive Statistics for Study, Participant, and Game Characteristics

	Mean	*N*	*SD*	Range
Study characteristics[a]				
Publication year	2009	77	2.80	2000–2012
Attrition (%)[b]	5	72	.10	0–.40
Randomized trial (%)	57	77		0–100
Cluster assignment (%)[c]	39	77		0–100
Location of study (%)				
North America	43	77		0–100
Asia	22	77		0–100
Europe, Middle East	30	77		0–100
South America	4	77		0–100
Australia	1	77		0–100
Participant characteristics[a]				
White (%)	40	23	.40	0–1
Nonwhite (%)	60	28	.42	0–1
Male (%)	55	54	.13	.2–.87
Average age (years)	11.08	46	5.34	4.2–34
Avg low end of grade range[d]	6.6	71	4.03	PreK–16
Avg high end of grade range[d]	7.67	70	5.26	PreK–16
General population sample (%)[d]	84	77		0–100
Game characteristics[e]				
Follow-up time (weeks)	4.16	393	7.77	0–27
Number of game sessions	14.49	385	13.24	1–50
Days game sessions are spread across	39.75	334	53.42	1–280
Possible implementation problems (%)	8	387		0–100
Game players (%)				
Single, no collaboration/competition	82	393		0–100
Single, competitive	5	393		0–100
Single, collaborative	1	393		0–100
Collaborative team competition	5	393		0–100
Multiplayer/MMO	7	393		0–100

[a]Variables measured at study level (n = 77).

[b]Attrition measured as the proportion of assigned study participants who were lost to follow-up/not included in the posttest measurement.

[c]Cluster assignment refers to assignment of participants at a cluster level (e.g., classroom, school).

[d]Low and high grade range refer to the mean low and grade levels included in each study (e.g., third- to fourth-grade students). General population sample refers to the percentage of studies that used general samples of students rather than isolating the sample to include only students from specific subdemographics (e.g., children with autism).

[e]Variables measured at effect size level (n = 393). Numbers close to but less than 393 or 77 are a function of some studies not reporting the targeted data. Percentages for categorical variables may not sum to 100 because of rounding.

Table 4. Descriptive Statistics for Game Characteristics as Percentage of 393 Effect Sizes[e]

Game Characteristics	Mean
Game platform (%)	
Computer	92
Handheld	5
Tablet	1
Augmented reality	1
Multiple non-augmented reality	2
Virtual reality space/equipment	1
Production values (%)	
Rough	31
Research grade/low budget	62
Indie	6
AAA	1
Technology type (%)	
2-D	88
3-D	12
Virtual reality	1
Augmented reality	1
Intrinsic/extrinsic type (%)	
Fully extrinsic	1
Mixed	2
Partial	1
Intrinsic	36
Intrinsic by default	62

[e]Percentages for categorical variables may not sum to 100 because of rounding.

Table 5. Descriptive Statistics for Outcome Characteristics as Percentage of 393 Effect Sizes[e]

Outcome Type	Mean
Broad outcome domain (%)	
Cognitive competencies	83
Intrapersonal competencies	16
Interpersonal competencies	1
Narrow outcome domain (%)	
Cognitive processes/strategies	14
Knowledge	66
Creativity	3
Intellectual openness	2
Work ethic/conscientiousness	10
Positive core self-evaluation	4
Teamwork/collaboration	1
Learning outcome discipline (%)	
Science	14
Math	15
Literacy	30
Social sciences	1
Engineering/computer science	1
Psychology	26
General knowledge	12
Assessment type (%)	
Preexisting instrument	58
Modification of existing instrument	9
Author-developed instrument	33

[e]Percentages for categorical variables may not sum to 100 because of rounding.

Meta-Analysis Comparisons

As stated, this meta-analysis addressed four core questions about digital games and learning:

1. What are the effects of digital games on learning for K–16 students?
2. How do these effects vary by learning outcome type in alignment with the categorizations of the recent National Research Council's *Education for Life and Work* report (NRC, 2012)? (See Table 7.)
3. How do these effects vary by game type? (See Tables 6 and 7.)

Overall, the largest body of literature we identified compared digital game interventions with other (nongame) instructional conditions, which are comparisons that may have the greatest relevance to educators (Table 6). Findings from these studies indicated that digital games were associated with significantly better cognitive competency outcomes among students relative to the other instruction comparison conditions. These beneficial effects were primarily based on knowledge outcome measures rather than cognitive processes/strategies outcome measures, of which there were fewer, or creativity outcome measures, of which there were none. Results indicated that game conditions integrating true simple games (i.e., game design involving more than simply draping school tasks with rudimentary game structures such as points and graphics) and those using interface enhancements (augmentations to the interface through which the player interacts with the game) showed the greatest beneficial effects on literacy and general knowledge measures.

Very few studies reported findings on intrapersonal competencies outcomes, but there was evidence that relative to other instructional conditions, digital games were associated with better intellectual openness and positive core self-evaluation outcomes. However, no studies provided information about learning outcomes within the interpersonal competencies domain, so evidence is insufficient to make any statements about the relative effectiveness of digital games for improving interpersonal competencies.

We then analyzed results from the 12 studies that compared digital game interventions with no-treatment control conditions, which indicated no beneficial effects of digital games on learning outcomes. This result was consistent across different outcome domains, subdomains, and disciplines. However, the failure to detect such effects could be due to low statistical power given the small ($n < 10$) number of studies that were available for any given analysis. Several additional factors may also have contributed to these findings: (1) most of these studies focused on psychological assessments involving students with autism or other disabilities, (2) most of the game conditions implemented in these studies were minimally described, and (3) these studies often appeared to involve game conditions with low production values. Given these issues, evidence is insufficient to make conclusions about the effectiveness or ineffectiveness of digital games on learning outcomes for students relative to no-treatment control conditions.

Finally, we analyzed several studies that compared digital games of different designs (Table 7). In many ways, we view these comparisons between designs as the most important in this study. Much of the debate in the field to date has centered on more simple questions about whether games are good or bad for learning. More productive questions address which designs and structures optimize which outcomes for whom and how. The National Research Council's reports on labs (NRC, 2005) and games and simulations (NRC, 2010) are much more useful when viewed through these lenses. Clearly, there are productive designs and unproductive designs of books, labs, movies, simulations, and games for specific goals and people. Nobody needs to be convinced that "bad" games, simulations, books, or labs are unproductive. From our perspective, the most important questions for future research are which design approaches are productive and what affordances are offered within a medium (c.f., Underwood, Banyard, & Davies, 2007).

Our initial findings make clear that there were significantly cognitive learning gains for the various enhanced game designs compared in the constituent studies. Our Phase II work will explore these relationships in greater detail. In terms of our initial findings, some evidence existed that game conditions using scaffolding enhancement (enhancements to the supports for the player within the game or aspects of the game that adapt to the needs or actions of the player) showed greater beneficial effects on cognitive processes/strategies and knowledge outcomes relative to those using interface enhancement (augmentations to the player-game interface) or player arrangement conditions (changes in the social arrangements between players, ranging from completely individual play to combinations of collaboration and competition). Again, however, there were relatively few (often $n < 10$) studies within any given analysis, so whether the lack of statistical significance for effects is due to low power or true null effects is unclear.

Findings from this meta-analysis should be interpreted in light of its limitations. The primary limitation is that most of the analyses were based on a small number of studies (often fewer than 10), and thus we cannot state whether the lack of observed effects in some instances are due to null effects or simply low statistical power to detect such effects. Although meta-analysis often increases statistical power to detect effects by pooling findings across multiple studies, it is nonetheless sensitive to the number of studies and estimated parameters in any given model. Furthermore, the exploratory moderator analyses used to examine whether effects varied across different types of game conditions were most likely severely underpowered given the small number of effect sizes within any given subgroup. Consequently, all subgroup analyses were considered exploratory and those results were presented descriptively rather than inferentially. Given these issues with statistical power and limited degrees of freedom, we also were unable to apply multivariable meta-regression models to examine whether other study, participant, methodological, or game characteristics were associated with effect size magnitude. In future analyses we plan to explore such multivariable models for those meta-analyses that included a large number of studies and at a minimum to explore for possible confounding effects among different study characteristics. However, the low statistical power to detect effects suggests that the effects measured in the statistically significant comparisons are substantial.

In summary, the findings from this meta-analysis indicate that compared with nongame instruction, digital games can enhance student learning as measured by cognitive competencies and some intrapersonal competencies. There was a noticeable lack of interpersonal competency outcomes reported in the literature, so evidence is insufficient at this time to make statements about digital game effects on those outcomes. There was also evidence that certain types of game structures may be more or less effective for certain types of outcomes, underscoring the importance of design beyond simple choice of medium when discussing the affordances of digital games for learning (just as researchers would assume for any other medium). Furthermore, there was no evidence in any of the analyses that digital games were associated with statistically significant adverse outcomes (i.e., worse learning outcomes).[1] Further analysis is required to investigate why the comparisons of games and no treatment showed a trend of no effect, whereas the comparisons of games and nongame treatments showed an effect.

1 A comparison is significant only if the confidence interval does not include 0 within its bounds (e.g., (.23,69) is significant while (-.36, .18) is not. Effect sizes associated with nonsignificant confidence intervals are not significant.

Table 6. Mean Effect Sizes for Digital Games vs. Other Instruction Comparison Conditions, by Broad Learning Outcome Domain

	Cognitive			Intrapersonal			Interpersonal		
	$\bar{g}$	95% *CI*	*n*	$\bar{g}$	95% *CI*	*n*	$\bar{g}$	95% *CI*	*n*
Type of Game									
All games	.32	(.19,.44)	38	.22	(-.04,.49)	8			0
Type of Game									0
Rudimentary game structure	.22	(.03,.40)	18	.33	(.03,.63)	4			0
Beyond rudimentary game structure	.40	(.24,.56)	20	.16	(-.36,67)	4			0
Type of Game									0
Rudimentary game structure	.22	(.03,.40)	18	.33	(.03,.63)	4			0
Integrating true simple games	.78	(.20,1.36)	3	.46	(-.11,1.02)	1			0
Situating in virtual context for exploration	.33	(.16,.50)	14	.28	(-.63,1.18)	2			0
Interface enhancement	.79	(.23,1.35)	1			0			0
Scaffolding enhancement	.14	(-.27,.56)	2	-.41	(-1.07,.24)	1			0

Note: A comparison is only significant if the confidence interval does not include "0" within its bounds (e.g., "(.23, 69)" is significant while "(-.36, .18)" is not. Effect sizes associated with non-significant confidence intervals are not significant.

Table 7. Mean Effect Sizes for Digital Games and Other Digital Game Conditions, by Broad Learning Outcome Domain

	Cognitive			Intrapersonal			Interpersonal		
Type of Focal Game	$\bar{g}$	95% *CI*	*n*	$\bar{g}$	95% *CI*	*n*	$\bar{g}$	95% *CI*	*n*
All focal games	.29	(.10,.48)	13	-.06	(-.29,.18)	9			0
Type of Focal Game									
Interface enhancement	-.01	(-.36,.34)	4			0			0
Scaffolding enhancement	.47	(.19,.75)	6	-.16	(-.45.14)	6			0
Player arrangement	.12	(-.25,.50)	3	.10	(-.27,.47)	3			0
Rich context			0			0			0

Notes: A comparison is significant only if the confidence interval does not include 0 within its bounds (e.g., (.23, 69) is significant while (-.36, .18) is not. Effect sizes associated with nonsignificant confidence intervals are not significant.

These tables retain the numbering from the full report to reduce confusion in comparing between the report and this executive summary. As a result, numbering of tables is not continuous here.

Implications and Plans for Phase II

Specific Questions Arising From Phase I

The Phase I results raise some issues for further exploration in Phase II. In particular, we want to explore (1) confounding variables between game design and learning outcome for the cognitive processes/strategies subdomain, and (2) differential outcomes for game design based on control group type.

Confounding Variables Between Game Design and Outcome Variable for Cognitive Processes/Strategies

Similarly, Phase I results noted potential confounding variables between game design and outcome variable for cognitive processes/strategies. The mean effect size for the cognitive processes/strategies subdomain was .31, which was not significantly different from zero and exhibited considerable heterogeneity. Most of these effect sizes were for games that were situated in a virtual context for exploration, however, so this might indicate confounding variables in terms of game type and outcome type. It also seems possible that this might be an issue of assessment methods because assessing cognitive processes/strategies is much trickier than measuring knowledge gains. Hence, these low effect sizes may be at least partially due to trying to use multiple-choice tests to measure more sophisticated forms of knowing (e.g., inquiry skills). In light of the substantial interest in leveraging immersive 3D worlds to engage students in inquiry and other disciplinary processes, this merits closer exploration in Phase II.

Different Outcomes for Game Design Based on Control Group

We also wish to explore the differential outcomes for game design based on the nature of the control group. The mean effect size for cognitive learning outcomes was .29, indicating a significant beneficial effect. Although results from a meta-regression provided no evidence that the type of focal game moderated this effect, results from subgroup analysis suggest that this positive effect was largely driven by studies in which the focal game used scaffolding enhancement (versus those that used interface enhancement or player arrangement). This contrasts with the findings from the game versus nongame comparisons, where interface enhancement showed the greatest impact. This may be a function of the quality of the scaffolding in the studies or its elaboration. Similarly, this may involve some constellation of the potential issues discussed in terms of psychology-focused studies versus the other studies as part of the first question. Thus, our exploration across all four questions from Phase I will address related issues during Phase II.

Study and Game Quality Issues

Our exploration of these questions during Phase II will also be connected to a systematic analysis of study and game quality. One of the critical themes in the preceding section relates to study and game quality—the quality of the comparisons made in a study, the quality of the game conditions used in these comparisons, the quality with which the methods were applied, and the quality of the assessments used. Our coding manuals have the functionality to support our examination of these issues, through our work during Phase I. The specific study quality variables to be evaluated in Phase II are internal validity (attrition rate and selection/regression to the mean bias), construct validity (instrument type and measurement reliability), statistical conclusion validity (sample size, sufficient reporting, appropriate application of statistical methods, and unit of assignment), and descriptive validity (equivalence and quality of comparison conditions and type of study design).

Application of More Sophisticated Statistical Techniques to Support Analysis Along These Fine-Grained Variables

In Phase II, we plan to conduct a more thorough examination of the heterogeneity in effect sizes, focusing on study quality characteristics as moderators of effects and variability across different learning constructs. To accomplish this goal, we intend to use newly developed statistical techniques of robust variance estimation in meta-regression that allows the inclusion of statistically dependent effect sizes within a given meta-analysis. Conventional meta-analytic methods, like those we used in Phase I, require that effect size estimates be statistically independent. For instance, in our analysis we used only one effect size estimate for each outcome construct for each group combination within a study. Nonetheless, our analysis still included multiple effect sizes from studies if a study included more than two group arms (e.g., two games shared the same comparison condition). Although sensitivity analysis indicated that this did not substantively alter our findings, we hope to improve on our analysis in the next phase by avoiding all loss of information associated with dropping effect sizes.

Cross-Check Search, Data, and Analysis Coding and Scripting

One of the greatest strengths of this meta-analysis relative to prior meta-analytic efforts in this field is the scope and comprehensiveness of the data collection. To ensure the highest quality possible in this regard, we intend to comb the field for any further studies that meet our eligibility criteria that were not located for any reason during Phase I. This will entail extensively cross-checking all search outcomes, eligibility coding, study coding, and meta-analytic scripting. These efforts will focus on

1. Rechecking all excluded titles from our literature search at the title level.

2. E-mailing across listserves of organizations typically involved in game-based research (e.g., the AECT, ISLS) with the list of currently included studies requesting authors to respond with citations of studies (theirs or others they know of) that might be eligible

3. Hand-checking the bibliographies of the studies currently identified as eligible as well as the bibliographies of review articles and other synthetic articles for any potentially eligible studies

4. Checking all the resulting abstracts of studies identified through Steps 1–3 as well as rechecking all abstracts excluded during Phase I.

After this extended cross-checking to ensure maximal identification of eligible studies, we will also have second independent coders recheck all full-article coding.

In addition, we will recheck the scripting of the meta-analytic comparisons run for Phase I as well as script all new comparisons resulting from our discussions with GLASS-Lab and resulting from the implications of Phase I comparisons and our extended analysis of issues related to study and game quality.

Through this process, our goal is to complete a meta-analysis of digital games for learning that has systematically incorporated all eligible studies across disciplines and that involves the careful cross-checking of all data and analyses to guarantee the highest level of reliability in the analyses and claims based on the corpus of studies. Our ultimate goal is to develop a meta-analytic review of the highest quality to be submitted to a first-tier academic journal so that it can support future research on the design of digital games for learning across disciplines and fields.

References

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers and Education, 53*(1), 74–85.

Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S., & Warrant, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology 18*, 305–320.

Barab, S. A., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E-J., Kouper, I., & Herring, S. C. (2007). Situationally embodied curriculum: Relating formalisms and contexts. *Science Education, 91*(5), 750–782.

Clark, D. B., Martinez-Garza, M., Nelson, B., D'Angelo, C. M., & Bellamy, S. (submitted). *Digital games and science learning: Research across the NRC strands of science proficiency.*

Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). *Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. Computers & Education, 57*(3), 2178–2195. doi:16/j.compedu.2011.05.007

Clark, D. B., Nelson, B., Sengupta, P., & D'Angelo, C. M. (2009). *Rethinking science learning through digital games and simulations: Genres, examples, and evidence.* Paper commissioned for the National Research Council Workshop on Games and Simulations, Washington, DC.

D'Angelo, C. M., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., & Haertel, G. (2013). *Simulations for STEM learning: Systematic review and meta-analysis (Vol. 1).* Menlo Park, CA: SRI International.

Federation of American Scientists. (2006). *Report: Summit on Educational Games: Harnessing the Power of Video Games for Learning.* Washington, DC: Federation of American Scientists.

Galas, C. (2006). Why *Whyville? Learning and Leading with Technology, 34*(6), 30–33.

Hickey, D., Ingram-Goble, A. ,& Jameson, E. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology, 18*(2), 187–208.

Hines, P. J., Jasny, B. R., & Mervis, J. (2009). Adding a T to the three R's. *Science, 323*, 53.

Ketelhut, D. J., Dede, C., Clarke J., & Nelson, B. (2006). A *multi-user virtual environment for building higher order inquiry skills in science.* Paper presented at the 2006 AERA Annual Meeting, San Francisco, CA. Available at http://muve.gse.harvard.edu/rivercityproject/documents/rivercitysympinq1.pdf

Klopfer, E., Osterweil, S., & Salen, K. (2009). *Moving Learning Games Forward*. Cambridge, MA: The Education Arcade.

Klopfer, E., Scheintaub, H., Huang, W., Wendal, D., & Roque, R. (2009). The simulation cycle: Combining games, simulations, engineering and science using StarLogo TNG. E-Learning, 6(1), 71–96.

Koster, R. (2004). *A theory of fun for game design (1st ed.).* Phoenix, AZ: Paraglyph Press.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. New York, NY: Penguin Press.

McQuiggan, S., Rowe, J., & Lester, J. (2008). The effects of empathetic virtual characters on presence in narrative-centered learning environments. In *Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems* (pp. 1511-1520), Florence, Italy.

Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology, 92*, 724–733.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96,* 165–173.

National Research Council (Honey, M. A., & Hilton, M., Eds.). (2010). *Learning science through computer games and simulations.* National Research Council. Washington, DC: National Academy Press.

National Research Council (Pellegrino, J. and Hilton, M., Eds.) (2012). *Education for Life and Work in the 21st Century. Committee on Defining Deeper Learning and 21st Century Skills,* Center for Education, Division of Behavioral and Social Sciences and Education, National Research Council. Washington D.C.: The National Academic Press.

National Research Council (Singer, S., Hilton, M. L., & Schweingruber, H. A., Eds.). (2005). *America's lab report: Investigations in high school science.* Washington, DC: National Academies Press.

Neulight, N., Kafai, Y. B., Kao, L., Foley, B., & Galas, C. (2007). Children's participation in a virtual epidemic in the science classroom: Making connections to natural infectious diseases. *Journal of Science Education and Technology, 16*(1), 47–58.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* West Sussex, England: John Wiley & Sons.

Salen, K., & Zimmerman, E. (2004). *Rules of play: game design fundamentals.* Cambridge, MA: The MIT Press.

Underwood, J. D. M., Banyard, P. E., & Davies, M. N. O. (2007). Students in digital worlds: Lost in Sin City or reaching Treasure Island? *BJEP Monograph Series II, Number 5 - Learning through Digital Technologies, 1,* 83–99.

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
Phone: 650.859.2000

www.sri.com

Developed by SRI International with funding from the Bill & Melinda Gates Foundation.

BILL & MELINDA GATES foundation