Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis

Executive Summary

SRI Education

March 2014

Developed by SRI Education 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 . (2014). *Digital Games, Design and Learning: A Systematic Review and Meta-Analysis (Executive Summary)*. Menlo Park, CA: SRI International.

SRI Education

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

www.sri.com/education

Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis

Executive Summary

Prepared By:

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

March 2014

Contents

Executive Summary

This meta-analysis systematically reviews research on digital games and learning for K-16 students in light of the recent NRC report on education for life and work in the 21st century (NRC, 2009). We synthesize comparisons of game conditions versus non-game conditions (i.e., media comparisons) as well as comparisons of augmented game designs versus equivalent standard game designs (i.e., value-added comparisons). We employed random-effects meta-analysis with robust variance estimates to summarize the overall effects of digital game interventions. Meta-regression models were used to assess the possible moderating effects of participant characteristics, game condition characteristics, and research quality characteristics. Findings from the media comparisons indicate that digital games significantly enhanced student learning relative to the non-game control conditions. Findings from value-added comparisons clearly demonstrate the importance of design beyond medium when evaluating the impact of digital games for learning. Media-comparison and value-added analyses underscore the importance of enhanced assessment techniques and research reporting going forward. An At-a-Glance Brief highlighting high-level findings of this meta-analysis is presented in Appendix A2: Games.

Introduction

In 2006, the Federation of American Scientists issued a widely publicized report stating that games as a medium offer a powerful new educational tool (FAS, 2006). The report encouraged private and governmental support for expanded research into complex gaming environments for learning. A special issue of Science in 2009 echoed and expanded this call (Hines, Jasny, & Mervis, 2009), as have reports by the National Research Council (Honey & Hilton, 2010; NRC, 2009). These reports acknowledged, however, the sparseness of systematic evidence for the efficacy of games as learning tools.

The present meta-analysis synthesized research on digital games to systematically examine their efficacy for learning. We focused on research published between 2000 and 2012 in light of the dramatic evolution of digital games for learning over the past decade. Our initial analyses attempted to replicate findings from three recent quantitative meta-analyses that compared digital games to non-game control conditions. As shown in Exhibit 1, however, the current study expanded upon those prior meta-analyses by focusing on an overlapping but distinct cross-section of the research literature[1], thus

---

1 Vogel et al. (2006) and Sitzmann (2011) included simulations, for example, but differences in eligibility criteria also resulted in the inclusion of less than 50% of Wouters et al. (2013) studies in the present meta-analysis (with publication date and our requirement of pretest-posttest designs accounting for most of the differences in this case).

Exhibit 1. Characteristics of recent meta-analyses on games for learning: Overlapping but distinct lenses

Authors	Learning environments	Scope	Study count	Years included	Demographics	Data
Vogel et al. (2006)*	computer games and interactive simulations	media-comparison	32	1986-2003	adult workforce trainees	pretest-posttest cognitive and attitudinal
Sitzmann (2011)**	simulation games	media-comparison	65	1976-2009	all age groups	pretest-posttest, posttest-only self-efficacy, declarative knowledge, procedural knowledge, retention
Wouters et al. (2013)***	serious games	media-comparison	39	1990-2012	all age groups	pretest-posttest, posttest-only knowledge, skills, retention, and motivation
Present Study	digital games	media-comparison and value-added	69	2000-2012	k-16 students	pretest-post cognitive, intrapersonal, and interpersonal

*Vogel, Vogel, Cannon-Bowers J., Bowers, Muse, & Wright (2006) *analyzed computer games and interactive simulations*. "A computer game is defined as such by the author, or inferred by the reader because the activity has goals, is interactive, and is rewarding (gives feedback). Interactive simulation activities must interact with the user by offering the options to choose or define parameters of the simulation then observe the newly created sequence rather than simply selecting a prerecorded simulation" (p. 231).

**Sitzmann (2011) analyzed *simulation games,* which they defined as "instruction delivered via personal computer that immerses trainees in a decision-making exercise in an artificial environment in order to learn the consequences of their decisions" (p. 492).

***Wouters, van Nimwegen, van Oostendorp, & van der Spek (2013) analyzed *serious games*. "We describe computer games in terms of being interactive (Prensky, 2001; Vogel et al., 2006), based on a set of agreed rules and constraints (Garris et al., 2002), and directed toward a clear goal that is often set by a challenge (Malone, 1981). In addition, games constantly provide feedback, either as a score or as changes in the game world, to enable players to monitor their progress toward the goal (Prensky, 2001). …. In speaking of a serious (computer) game, we mean that the objective of the computer game is not to entertain the player, which would be an added value, but to use the entertaining quality for training, education, health, public policy, and strategic communication objectives (Zyda, 2005)" (p 250).

providing an important extension and verification of the generalizability of earlier findings. Based on these prior meta-analyses, we predicted that students in digital game conditions will outperform *students in non-game conditions in terms of learning outcomes (Hypothesis 1)*.

Whereas prior meta-analyses on this topic have focused exclusively on comparisons of game conditions versus non-game control conditions (which Mayer, 2011, calls media comparisons), the present study also focused on value-added comparisons. Value-added comparisons measure the efficacy of a standard version of a game relative to an enhanced version of that game augmented in some manner to test a theoretical design proposition (Mayer, 2011). The present study thus moved beyond a sole focus on media comparisons to also assess the contribution of design to learning beyond medium. We predicted that *games with theoretically augmented designs for learning will outperform standard versions of those games (Hypothesis 2)*.

Beyond these two broad hypotheses, the present study explored the potential moderating effects in media comparisons of general study characteristics, game mechanics characteristics, and visual and narrative characteristics.

General study characteristics in media comparisons

Prior research suggests that the following general study characteristics may be related to learning outcomes in media comparisons—game duration, presence of non-game instruction in game conditions, and player grouping—and therefore we examined these characteristics as potential moderators in the current meta-analysis.

With regard to duration of game-play, Sitzmann (2011) found that media comparisons in which trainees had unlimited access to the game demonstrated significantly higher learning gains (relative to the non-game control conditions) than media comparisons in which the trainee had limited access to the game. Similarly, Wouters et al. (2013) found that comparisons in which participants engaged with the game for only one session did not result in significant learning gains relative to the non-game control condition, whereas comparisons in which participants interacted with the game for more than one session resulted in significant learning gains relative to participants in the non-game condition. Based on these findings, we predicted that *comparisons involving more than a single game-play session, as well as comparisons involving longer play durations, will be associated with better learning outcomes relative to non-game conditions. (Hypothesis 3a)*. Note that for this hypothesis, as well as for Hypotheses 3b-3c, 4a-4d, and 5a-5e below, the core comparisons in the included studies are between the game conditions and the non-game conditions. Our hypotheses therefore compare the effect sizes from one set of comparisons (e.g., games with additional instruction versus the non-game control conditions) to the effect sizes from a second set of comparisons (e.g., games without additional instruction versus the non-game control conditions) for significant relationships in associated effect sizes.

In terms of supplemental non-game instruction, Sitzmann (2011) and Wouters et al. (2013) both found that comparisons where game conditions included supplemental non-game instruction demonstrated better learning outcomes relative to comparisons in which the game conditions did not include non-game instruction. Given the importance of verbalization for learning (Wouters et al., 2008), and the positive effects of supplemental instruction on learning observed in prior meta-analyses, we predicted that *game conditions with non-game instruction will outperform game conditions without non-game instruction relative to non-game conditions (Hypothesis 3b)*.

In terms of player structuring in game conditions, Vogel et al. (2006) found significant learning outcomes for single-player as well as for collaborative conditions in comparisons to non-game conditions, and reported a trend toward larger effect sizes with solitary players, but did not report analyses comparing effect size magnitudes between the two player structures. Based on the trend in Vogel et al., and given ambiguity in prior research on the benefits of collaborative play (e.g., van der Meij, Albers, & Leemkuil, 2011; Schwartz, 1995), Wouters et al. (2013) hypothesized that single-user play would result in better learning than group play. Wouters and colleagues found, however, that learners who played serious games in a group learned more than learners who played alone. Based on this prior research, we therefore predicted that *collaborative game conditions will outperform single-player game conditions relative to non-game conditions (Hypothesis 3c)*.

Game mechanics characteristics in media comparisons

In addition to exploring general study characteristics, we also explored whether specific game mechanics were associated with larger effects in terms of learning outcomes. Previous meta-analyses on this topic have focused on whether the learning mechanics actively versus passively engaged the learner. Sitzmann (2011) found that trainees in the simulation/game group learned more, relative to the comparison, when the simulation/game conveyed course material actively rather than passively. Vogel et al

(2006) similarly found that traditional teaching methods outperformed simulations/games when the computer dictated the sequence of the simulations/games rather than the students. For the current study, we wished to explore the role of game and learning mechanics at a finer grain.

Specifically, we explored broad sophistication of game mechanics (simple gamification of academic tasks versus more elaborate game-mechanics), variety of player actions (focused games like Tetris versus games like SimCity where players engage in a wider variety of actions), intrinsic/extrinsic design properties (c.f., Kafai, 1996; Habgood & Ainsworth, 2011), and sophistication of scaffolding. We predicted that *more sophisticated game mechanics, increased variety of player actions, intrinsic integration of the game mechanic and learning mechanic, and more specific/detailed scaffolding will be related to larger effects on learning outcomes relative to non-game conditions (Hypotheses 4a-4d)*.

Visual and narrative game characteristics in media comparisons

Results from prior meta-analyses examining the effects of digital games on learning have yielded inconsistent and conflicting results in terms of the potential moderating effect of visual and narrative contextualization. For instance, Vogel et al. (2006) found no evidence that realism was associated with larger or smaller effects on learning outcomes, whereas Wouters et al. (2013) found that the least realistic (schematic) games were more effective for learning than cartoon or realistic games. Additionally, Wouters et al. (2013) found suggestive evidence ($p=.09$) that games without narratives outperformed games with narratives.

One problem with analyses of visual realism, narrative, or other contextual features of game environments is that these features are likely intercorrelated. The most visually realistic games, for example, typically incorporate a first-person perspective, have more anthropomorphic game entities, and generally have deeper story lines. We therefore expanded coding of visual features to include not only visual realism but also camera perspective and anthropomorphism. These latter two variables were not included in prior meta-analyses and are likely to be correlated with visual realism. The relevance of camera viewpoint for learning is included due to the numerous reports that have shown that playing first person perspective shooter games may lead to improvement on certain visual cognitive tasks (e.g., Feng, Spence, & Pratt, 2007; Green & Bavelier, 2006; Green & Bavelier, 2007). Anthropomorphism was included based on numerous findings suggesting that anthropomorphic features and motions affect a range of perceptual, cognitive, and social tasks across various age ranges (e.g., Heider & Simmel, 1944; Killingsworth, Levin, & Saylor, 2011; Levin, Killingsworth, Saylor, Gordon, & Kawamura, 2012; Mahajan & Woodward, 2009).

In addition to including these visual features in our analyses, we also examined the narrative features of each game condition. Overarching research on learning supports the inclusion of narrative context. Brown, Collins, and Duguid (1989), for example, highlighted the importance of learning in the context of a larger activity, while Bransford, Sherwood, Hasselbring, Kinzer, and Williams (1990) highlighted the power of anchoring instruction in narrative. Bransford, Brown, and Cocking (2000) synthesized a broad range of research supporting these points. In research on games for learning, Malone and Lepper (1987) hypothesized that the narrative should be closely integrated with the learning mechanics to support learning. The role of narrative in games for learning remains a central focus of the field (e.g., Dickey, 2006; Lim, 2008; Echeverria, Barrios, Nussbaum, Amestica, and Leclerc, 2012). We therefore examined the relevance of the narrative to the learning mechanic as well as the depth and richness of the narrative.

Based on the findings of prior meta-analyses in this field, we predicted that visual realism, anthropomorphism, camera perspective, story relevance, and story depth will be related to smaller effects on learning outcomes relative to non-game conditions (Hypothesis 5a-5e). As mentioned, we also investigated and controlled for the relationships between these features. In addition, given the trends observed in prior research, we predicted that *greater overall contextualization will be related to smaller effects on learning outcomes relative to non-game conditions (Hypotheses 5f)*.

Research quality characteristics in value-added and media comparisons

Beyond study and game characteristics, we also explored whether research quality was associated with larger or smaller effects in the media comparisons and value-added studies. Prior meta-analyses have noted issues with the methodological quality of the primary studies in the games literature (Vogel et al., 2006), and have noted that the beneficial effects of serious games may be attenuated in studies with random assignment versus quasi-experimental designs (Wouters et al., 2013). To explore research quality characteristics more deeply, we predicted that *comparison condition quality, sufficient condition reporting, sufficient reporting of methods and analyses, over-alignment of assessment with game, assessment type, and study design will be related to learning outcomes in value-added and media comparisons (Hypotheses 6a-6f)*.

Meta-Analytic Methods

A meta-analysis is the systematic synthesis of quantitative results from a collection of studies (Borenstein, et al., 2009) focused on a given topic. Part of the systematic approach in a meta-analysis is to document the decisions that are made regarding the collection of the articles and the steps of the analysis. In a meta-analysis, articles are included based on pre-defined criteria and not due to favorable results found in the article or familiarity with certain authors. This can help to remove some of the bias and subjectivity that would result from a less systematic review.

Meta-analysis quantifies results by using effect sizes. Effect sizes are a measure of the difference between two groups. In the case of an intervention, an effect size can be thought of as a measure of the (standardized) difference between the control group and the treatment group, thereby providing a measure of the effect of the intervention. Effect sizes are not the same as statistically significant differences that are typically reported and determined through the use of inferential statistics, such as t-tests or analysis of variance (ANOVAs). A research study, for example, could have a statistically significant finding, but the effect of that difference could be minimal. Thus the effect size allows researchers to determine the magnitude of the impact of an intervention, not just whether or not the intervention made a difference. For example, an effect size of 1.00 would be interpreted as a difference of one standard deviation between the two groups being compared. Another way of interpreting a one standard deviation effect size would be moving a student at the 50th percentile before the intervention to the 84th percentile after the intervention.

The current meta-analysis employs a recently developed statistical technique for robust variance estimation in meta-regression (Hedges, Tipton, & Johnson, 2010; Tipton, 2013). This technique permits the inclusion of multiple effect sizes from the same study sample within any given meta-analysis—a common occurrence in meta-analyses in the educational and social sciences (e.g., Tanner-Smith & Tipton, 2013; Tanner-Smith, Wilson, & Lipsey, 2013; Wilson, Tanner-Smith, Lipsey, Steinka-Fry, & Morrison, 2011). This approach avoids loss of information associated with dropping effect sizes (to ensure their statistical independence) and does not require information about the covariance structure of effect size estimates that would be necessary for the use of multivariate meta-analysis techniques (see the full report for more details and Tanner-Smith & Tipton, 2013, for a discussion).

Inclusion and Exclusion Criteria

Digital game

Eligible studies were required to include at least one comparison of a digital game versus a non-game condition and/or at least one comparison of an augmented game design versus an equivalent standard game design (but these two types of comparisons were always analyzed separately). To be eligible for inclusion, studies were required to explicitly designate the environment as a game, and the term game or games needed to appear in the abstract or title of the report. A digital game was defined for the purposes of the present meta-analysis as a digital experience in which: (1) the participants strive to achieve a set of non-trivial fictive goals within the constraints of a set of rules that are enforced by the software; (2) the participants receive feedback toward the completion of these goals (e.g., score, progress, advancement, win condition, narrative resolution); and (3) the participants are intended to find some recreational value in the game. Hybrid augmented reality games that used digital platforms to create games in physical space were eligible, but physical games without a digital platform were excluded (e.g., board games). Games that focused exclusively on teaching youth to create or program games were not included for the present analyses because these approaches were considered distinct (and potentially more powerful) in light of their closer alignment with design-based learning (e.g., Kafai, 2006).

Participants

Eligible participant samples included students in grades K-16, ages 6 to 25. Participants were required to be students in a K-12 institution, or students enrolled in a postsecondary educational institution. Studies of participants beyond the K-16 grade range were not eligible for inclusion. Studies focusing on samples from specific clinical populations of students were also excluded (e.g., autism spectrum).

Research designs

Because the goal of the meta-analysis focused on making causal inferences regarding the effects of digital games on learning, only those studies using randomized controlled trial and quasi-experimental research designs were eligible for inclusion.

Learning outcomes

Eligible studies were required to measure information on at least one eligible outcome related to "learning" aligned with the recent NRC report on Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century (Pellegrino & Hilton, 2012). This report categorized learning into three broad domains: the cognitive, intrapersonal, and interpersonal domains. The cognitive domain includes cognitive processes and strategies, knowledge, and creativity. The intrapersonal domain includes intellectual openness, work ethic and conscientiousness, and positive core self-evaluation. The interpersonal domain includes teamwork, collaboration, and leadership.

Publication type

To reflect the current state of digital game design, eligible studies were required to have been published between January 2000 and September 2012 in a peer-reviewed journal article. Restricting eligibility to publications in peer-reviewed journals was selected to provide consistent sampling across the diverse fields and databases covered in the literature search as outlined in the Search Strategies section below. Nonetheless, to be sensitive to any biases this may have created in our study set, we conducted extensive sensitivity analyses to assess for the possibility of publication bias, as outlined below in the Data Analysis section.

Study site and language

Eligible studies were those published in English (although eligible studies need not be conducted in English or in an English speaking country).

Effect sizes

Eligible studies were required to report sufficient information needed to calculate both pretest and posttest effect sizes on at least one measure of learning and the variables involved in the effect sizes had to have a known direction of scoring (i.e., whether high or low values represent favorable or less favorable results).

Search Strategies

Comprehensive literature 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 any studies at the database search stage). We wanted to maximize sensitivity in our search. Our database search criteria therefore simply specified that the term game or games needed to be included in the abstract or title. All other potential search terms were deemed likely to eliminate otherwise eligible studies. Because research on games for learning spans many fields including engineering, computer science, medical, natural sciences, and social sciences, we searched the following hosts/databases: ISI Web of Science (SSI, SSSI); Proquest (ERIC, PsycINFO, Soc Abstracts, Social Services Abstracts); PubMED; Engineering Village (Inspec, Compendex); and IEEE Xplore. We also hand checked the bibliographies in narrative reviews and meta-analyses.

Coding Procedures

Coding occurred in two stages: at the eligibility stage and the full-study stage. Eligibility coding first occurred at the title level, where two research assistants independently screened all titles identified in the literature search to eliminate clearly ineligible reports (e.g., reports in non-English languages) or publications that reported on games that were clearly irrelevant for the current study (e.g., discussion of the Olympic Games, sports injuries, or investigations of game theory in psychology and economics).

Eligibility coding next occurred at the abstract level, where remaining abstracts were independently screened by two research assistants. All research assistants were first trained on a randomly selected subset of 100 abstracts, which were discussed until 100% consensus was reached with the entire group. The remaining abstracts were screened independently by two research assistants, and any disagreements were resolved by one of the authors. If there was any ambiguity about the potential eligibility of a report based on the abstract (or title, when abstracts were not available), we erred on the side of inclusivity and retrieved the full text report for final eligibility coding.

The final stage of eligibility coding occurred at the full-text level, in which all reports previously identified as potentially eligible at the abstract level were screened for final eligibility. At least two research assistants conducted independent full-text screening of each article, and any questions about eligibility were resolved by consensus with one of the study authors. The reason for ineligibility was recorded for each report, using the criteria outlined above.

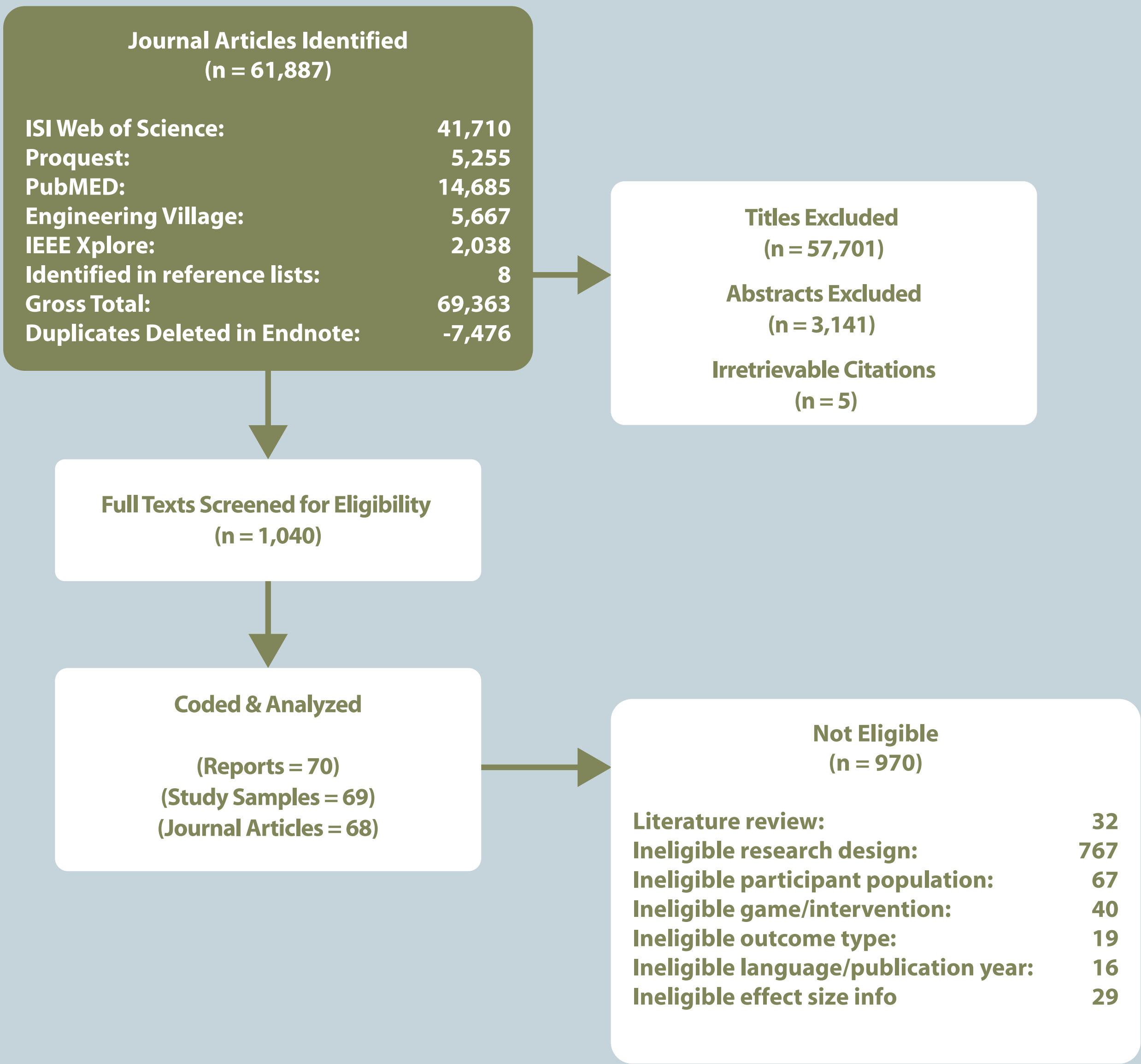
Studies that were deemed ineligible at the full text level were not coded further. Studies identified as eligible at the full text level progressed to full-study coding, in which two of the study authors coded all game and non-game condition characteristics while two research assistants independently extracted information about the studies, participants, research conditions, and effects sizes. Any discrepancies in the coding were discussed in person and resolved via consensus between coders and at least one of the study authors. For all coding procedures, coders entered data directly into a FileMaker Pro database using computer screens tailored to the coding items. Effect size calculations were built into the data entry screens, thereby minimizing the possibility of error in effect size calculations.

Results

Literature Search

All literature searches were conducted in September 2012. Figure 1 outlines the eligibility coding for the 61,887 reports identified in the literature search. A majority of reports were initially screened out at the title level (n = 57,701). We next screened the resulting 3,141 abstracts for eligibility for coding at the full report level. We then screened the resulting 1,040 reports in full text to determine final eligibility status. Most of the reports were ineligible for inclusion in the meta-analysis due to inadequate research designs (i.e., many were concept pieces that did not empirically examine the effect of a digital game or conduct comparisons across conditions). After screening the full text articles, 69 unique study samples included in 70 reports from 68 journal articles ultimately met the eligibility criteria and were included in the final meta-analysis (Figure 1). These 69 study samples provided information on a total of 6,868 unique participants. Citations for the 68 journal articles are included in a separate reference section.

Figure 1. Study Identification Flow Diagram



Overarching Media-Comparison and Value-Added Findings

All meta-analyses were estimated using robust variance estimates and could include multiple effect sizes from each study sample. Please see full report for complete analyses and statistics.

Hypothesis 1: Students in digital game conditions will outperform students in non-game conditions in terms of cognitive, intrapersonal, and interpersonal learning outcomes. Fifty-seven studies included comparisons of digital game interventions versus other non-game instructional conditions (i.e., media comparisons). Overall, results indicated that digital games were associated with a .33 standard deviation improvement relative to control conditions, even after adjusting for baseline differences in achievement between groups. Thus, the analyses show that digital games conditions were on average more effective than the non-game instructional conditions included in those comparisons.

These findings generally confirm and parallel the overall findings of Vogel (2006), Sitzmann (2011), and Wouters et al. (2013) in terms of learning outcomes. The present meta-analysis does diverge with the Wouters et al. (2013) finding that game conditions and non-game instructional conditions did not significantly differ in terms of motivation outcomes. In the current meta-analysis, motivation was incorporated within the intrapersonal learning outcome domain (as outlined in the NRC report on education for life and work in the 21st century), but intrapersonal outcomes also included intellectual openness, work ethic and conscientiousness, and positive core self-evaluation. Thus the findings of the current meta-analysis do not disconfirm the motivation findings of Wouters et al., but do demonstrate that game conditions support improvements in intrapersonal learning outcomes relative to non-game instructional conditions.

Hypothesis 2: Games with theoretically augmented designs for learning will outperform standard versions of those games. Twenty studies included comparisons of augmented versions versus standard versions of digital games (i.e., value-added comparisons). Overall, results indicated that augmented game designs were associated with a .37 standard deviation improvement in learning relative to standard versions, even after adjusting for baseline differences in achievement between groups. This finding highlights the importance of design in learning outcomes.

Furthermore, the largely overlapping confidence intervals around the mean effect sizes from the media-comparison and value-added analyses suggest that the findings for the media-comparison and value-added analyses of an augmented versus a standard version were similar in magnitude. This suggests that the design of an intervention is associated with as large an effect as the medium of an intervention. Although this finding may appear common-sense, the role of design is often de-emphasized in debates over whether digital games are "better" or "worse" than traditional instruction. It is critical to consider this finding when interpreting the media-comparison analyses.

General Study Characteristics in Media Comparisons

Hypothesis 3a: Comparisons involving more than a single game-play session, as well as comparisons involving longer play durations, will be associated with better learning outcomes relative to non-game conditions. Consistent with the results reported by Wouters et al. (2013), we found that (a) game conditions involving multiple game-play sessions demonstrated significantly better learning outcomes than non-game control conditions and (b) game conditions involving single game-play sessions did not demonstrate significantly different learning outcomes than non-game control conditions. We then conducted analyses of game-play duration (treating total gameplay duration as a continuous moderator variable), but we found no evidence of a consistent correlation between total duration and effects on learning outcomes.

Taken together, these findings may reflect a memory benefit of spaced learning as compared to massed learning in game contexts. Longer play durations may thus enhance learning, but only when sessions are adequately spaced. Alternatively, the distinction in findings may simply suggest that (a) a single session is not sufficient but (b) the multiple-session game conditions were being played longer than needed to achieve relevant learning outcomes in most studies. Games were played for an average of 347 minutes (or almost 6 hours). It may be that students in the studies tended to learn the key ideas relatively quickly from the perspective of the assessments utilized in those studies. From this perspective, future research would be desirable to determine whether longer durations of game play are associated with better performance that would be detected by deeper assessments of student learning.

The relationships to Sitzmann's (2011) findings are more difficult to interpret because Sitzmann focused on a different contrast. Sitzmann compared unlimited access to the simulation/game versus restricted access to the simulation/game. Gains relative to control conditions were significantly greater for the group with unlimited access. Sitzmann's finding may connect more directly to research on increased learner control, which has proven important in prior meta-analyses and broader research on motivation to learn (c.f., Pintrich, 2003).

Hypothesis 3b: Game conditions with non-game instruction will outperform game conditions without non-game instruction relative to non-game conditions. Additional non-game instruction was not significantly associated with larger or smaller effects for game conditions in media comparisons. These findings diverge from Sitzmann (2011) and Wouters et al. (2013), who found that supplemental non-game instruction supported learning. One possible explanation for this difference involves how "additional instruction" was coded across meta-analyses. Both Sitzmann (2011) and Wouters et al. (2013) may have used a much more stringent definition for "additional instruction". In the present meta-analysis, game conditions were coded as including any additional non-game instruction (whether integrated or not) if players were exposed to a learning context that was likely to provide them with additional topic-relevant information (such as spending days in typical classroom instruction). Based on Wouters and colleagues' example of "inclusive" studies (i.e., those including additional non-game instruction), it is possible that only studies that explicitly stated that players received additional domain-relevant instruction were coded as such. This might suggest that additional teaching or activities specifically designed to supplement game content as part of an integrated experience can increase learning, but that unintegrated supplemental teaching on a topic is unlikely to contribute to larger gains.

Hypothesis 3c: Collaborative game conditions will outperform single-player game conditions relative to non-game conditions. When controlling for game characteristics, gains from single-player games without competition and gains from collaborative team competition games exceeded those from single-player games with competition. These findings partly parallel the findings of Wouters et al. (2013), but may elaborate upon their findings. Wouters et al. found that collaborative play was generally more effective than individual play. Our findings suggest that collaborative games may not be more effective for learning than single player games, but instead suggest that games with competitive single-player structures are least effective.

This explanation would align with research on motivation. The motivational support of self-efficacy for certain students in a single-player competitive structure is necessarily a failure to support other students (because one student's gain necessitates another student's loss). Several models of motivation suggest that self-efficacy is a critical variable in motivation to learn (e.g., Bandura, 1997; Pintrich, 2003; Schunk, 1991). This explanation, however, does not account for the higher learning gains from single-player games without competition over those observed for multiplayer/massively multiplayer game conditions. Another explanation may simply be that the other features of the games and assessments were not equivalent across the player structure categories. Caution is critical given the small number of studies involved in all of the categories except the category of single-player without competition.

Game Mechanic Characteristics in Media Comparisons

We predicted that more sophisticated game mechanics, increased variety of player actions, intrinsic integration of the game mechanic and learning mechanic, and more specific/detailed scaffolding will be related to larger effects on learning outcomes relative to non-game conditions (Hypotheses 4a-4d). The comparison of broad design sophistication (Hypothesis 4a) demonstrated that simple gamification as well as more sophisticated game mechanics can prove effective. Future research and analyses should explore whether or not the "simple gamification" studies (e.g., games that simply add contingent points and badges to learning activities) more frequently focus on lower-order learning outcomes as compared to studies with more sophisticated game mechanics. Regardless, these results support the proposal that simple gamification can prove effective for improving certain types of learning outcomes (as suggested by Charles, Bustard, & Black, 2011; Lee & Hammer, 2011; Sheldon, 2011). These findings parallel those regarding variety of game actions (Hypothesis 4b), in which significant learning outcomes relative to non-game control conditions were observed for all levels of action variety with no significant differences between them.

The present meta-analysis is largely silent with regard to intrinsic versus extrinsic design (Hypothesis 4c) due to the fact that only one study involved a fully extrinsic condition. Regarding the nature of scaffolding (Hypothesis 4d), each category of scaffolding demonstrated significant effects on learning relative to non-game control conditions, and higher levels of scaffolding were associated with higher relative learning outcomes than lower levels of scaffolding. Enhanced scaffolding also showed significant effects on learning outcomes in the value-added analyses. Taken together, these findings suggest that current approaches and designs to enhance scaffolding have proven successful, providing a productive foundation for ongoing work on enhancing scaffolding in games (e.g., Barzilai & Blau, 2014).

Visual and Narrative Game Characteristics in Media Comparisons

Based on the findings of prior meta-analyses in this field, we predicted that visual realism, anthropomorphism, camera perspective, story relevance, and story depth will be related to smaller effects on learning outcomes relative to non-game conditions (Hypothesis 5a-5e) as will overall aggregate contextualization (Hypothesis 5f). The influence of individual visual and narrative game characteristics (Hypotheses 5a-5e) proved highly intercorrelated. An aggregate contextualization variable created from these game features (Hypothesis 5f) demonstrated a small but significant negative relationship with learning gains overall. Our findings parallel the findings of Wouters et al. (2013), (1) showing that schematic games were more effective than cartoon-like or realistic serious games and (2) supporting the non-significant trend observed in their study suggesting that games with no narrative might be more effective than games with narratives. More tangentially, our findings also suggest a possible parallel to findings from Sitzmann (2011) showing that the entertainment value of the simulations and games did not significantly affect learning outcomes.

On the surface, these contextualization findings conflict with research and theory highlighting the value of situating learning in context (e.g., Bransford, Sherwood, Hasselbring, Kinzer, and Williams, 1990; Brown, Collins, & Duguid, 1989). Mitigating this interpretation, however, is the fact that almost all of the studies included in the present meta-analysis involved immediate posttests. The arguments for situating learning in context focus on supporting students in developing deep, durable, integrated understanding that students can apply across contexts (essentially the opposite of an immediate focused posttest). We therefore interpret our findings as potentially highlighting (a) the challenges of designing rich contexts that support learning without obscuring key relationships and (b) the importance in future research of assessments designed to measure the deeper understanding highlighted in research on situated cognition and learning (e.g., Bransford, Brown, & Cocking, 2000). This shift in assessment is especially important in light of theoretical propositions suggesting that the greatest strengths of digital games as a medium involve their affordances for supporting higher-order cognitive, intrapersonal, and interpersonal outcomes (e.g., Gee, 2007; Squire, 2011).

Research Quality Characteristics in Value-Added and Media Comparisons

To explore research quality characteristics more deeply, we predicted that comparison condition quality, sufficient condition reporting, sufficient reporting of methods and analyses, over-alignment of assessment with game, assessment type, and study design will be related to learning outcomes in value-added and media comparisons (Hypotheses 6a-6f). Few studies met all four study-design-independent quality variables for Hypotheses 6a-6d, supporting claims that overall rigor needs to be increased in research on games for learning. That said, results from moderator analyses indicated that few study quality variables (design-independent or design-dependent) were highly correlated individually with the effects of digital games on learning outcomes in the media-comparison or value-added analyses (Hypotheses 6a-6f). This provides additional confidence in our effect estimates and suggests that findings were not unduly biased by individual study quality variables. Further discussion is merited, however, for one design-independent variable (control condition quality) and both design-dependent variables (assessment type and research design).

Control condition quality

Sitzmann (2011) found that trainees learned less from simulations/games relative to control groups when the control groups received an active learning experience. Wouters et al. (2013) found the opposite. Wouters and colleagues found (a) that serious games did not improve learning more than passive instruction and (b) that the benefits of serious games were significantly larger relative to mixed instruction control conditions as opposed to passive instruction control conditions. Most of the comparison conditions in the present meta-analysis were active or mixed, and we found a significant negative relationship between non-game comparison condition quality and learning gains. Restricting the meta-analysis to only those studies with medium or better comparison condition quality (thus weeding out "straw-man" comparisons) reduced the effect size from .33 to .28 (although remaining significant).

These findings further underscore the importance of design (and careful reporting of that design) for both game and non-game conditions. Media-comparison research often highlights medium while placing less emphasis on the design of the game and control conditions. Many of the media comparison studies in the present meta-analysis, for example, underspecified (or didn't specify) the nature of the game and/or control interventions. As research on games shifts toward closer analysis of role of design, researchers will need to provide "thicker" descriptions of game and control conditions, including screenshots, URLs, and detailed descriptions of the underlying theoretical bases for the chosen designs. This will require many journals to increase page limits.

Assessment type

There are trade-offs between research questions of interest and the availability of pre-existing normed instruments. While pre-existing assessments can clearly enhance confidence in research quality, these instruments exist only for certain outcomes. However, the present meta-analysis found no evidence of a relationship between assessment type (i.e., pre-existing normed instrument, modification of a pre-existing instrument, or author-developed instrument) and effect size. The present meta-analysis also found no evidence of a relationship between effect sizes and potential over-alignment of assessments. Given the aforementioned trade-offs and our null result concerning the impact of normed instruments on effect sizes, we propose that requiring research to rely exclusively on pre-existing normed instruments would unnecessarily limit digital games research. This issue is particularly relevant for the outcome types most desirable from a 21st century-skills and preparedness perspective, such as inquiry and leadership, for which normed assessments are scarce. Furthermore, while over-alignment of assessments and tasks is not desirable, misalignment is even more damaging. Researchers should not be discouraged from choose appropriate assessments based on learning goals, but should be held accountable for reporting reliability and validity information for author created or modified instruments. At present, such reporting is rare.

Research design

Randomized designs are often considered more rigorous than quasi-experimental designs, but randomized designs preclude many research questions and populations. Although there were no significant differences in average effects across randomized and quasi-experimental designs in the present meta-analysis, observed effects were notably smaller in the studies using randomized designs. Follow-up analyses showed, though, that differences in game characteristics between games in studies using randomized versus quasi-experimental designs might partially account for effect size differences across study designs. The tradeoffs between randomized and quasi-experimental designs remain unclear. We therefore argue that researchers should carefully weigh the benefits of experimental designs in light of fundamental issues of ecological validity, authenticity, and specific requirements of the research questions under exploration[2] In studies where quasi-experimental designs are implemented, however, researchers must (a) provide more substantial information about the attributes of the groups being compared and (b) account for those attributes in analyses.

2 The authors wish to emphasize that this discussion should in no way be taken as denying the critical role or relative importance of qualitative research designs and other quantitative research designs. We are focusing this discussion in terms of quantitative meta-analyses.

Caveats and Limitations

This section raises three issues for consideration. The first issue is commensurability, which should be considered when interpreting this (or any) meta-analysis. Meta-analyses assume that the included pairwise comparisons (effect sizes) represent relatively standardized or homogenous conditions. In actuality, this is not the case even in settings that might appear highly standardized. Even meta-analyses of medical research, for example, involve difficult balancing acts in terms of commensurability. Jüni, Witshci, Bloch, and Egger (1999) described these hazards in great detail in their Journal of the American Medical Association article. Commensurability issues pose even greater challenges when aggregating studies of learning and education, where variations across contexts, interventions, and approaches are more extreme.

The present meta-analysis explored only the specific game designs and methodological choices included in the constituent studies. Furthermore, as outlined in Table 1, this meta-analysis included a distinct cross-section of studies relative to other recent meta-analyses. Thus, while meta-analyses aggregate research conditions into categories that sound highly generalizable, the included research conditions do not fill or equally represent the entire domain suggested by the categories. Neither this nor any other meta-analysis can thus account for all possible design approaches or the implementation quality of those approaches.

We therefore do not suggest that future research and design should focus only on the characteristics and mechanics that outperformed others in this meta-analysis. Instead, if designs around a specific characteristic demonstrated lower learning outcomes, then other designs should be investigated if that characteristic is considered critical. We argue that this implication is particularly salient regarding our findings for visual and narrative contextualization, where overarching research supports situating cognition in terms of transfer and deeper understanding (c.f., Bransford, Brown, & Cocking, 2000), but the findings of this meta-analysis underscore challenges in terms of design implementation.

In addition to commensurability of game conditions, there are also commensurability issues for non-game comparison conditions. The studies in the present meta-analysis generally compared targeted game interventions to traditional or typical instructional approaches rather than to optimized learning activities. Thus, the findings of the media-comparison analyses should not be interpreted as suggesting that game-based instruction is superior to all learning experiences that could be designed within traditional media; rather, the findings suggest that the game-based experiences analyzed in these studies were superior to the traditional non-game approaches implemented in the included studies. All forms of media have particular affordances and constraints that must be considered in the design of high-quality instruction. Digital games and traditional instruction necessarily vary on many dimensions. We therefore argue against simplistic quotations of findings suggesting that games universally outperform non-game learning approaches. The results and comparisons are more complex and should be acknowledged as such.

The second caveat concerns sample size. Although meta-analyses can increase statistical power to detect effects by pooling findings across multiple studies, meta-analyses are nonetheless sensitive to the number of studies and estimated parameters in any given model. Analyses involving small numbers of studies should therefore be interpreted in light of the limitations to their statistical power.

The final caveat concerns assessments. Generally speaking, higher-order cognitive, intrapersonal, and interpersonal processes and skills prove more challenging to accurately and reliably measure than lower-order cognitive processes, skills, and rote knowledge. These differences manifest themselves in the heavy skewing of the research that has been conducted toward lower-order cognitive skills and immediate posttests. The NRC report on education for life and work in the 21st century, however, emphasizes a more distributed focus across outcomes, if not a complete reversal in emphasis. Furthermore, proponents of digital games for learning (e.g., Gee, 2007; Squire, 2011) propose that the greatest strengths of digital games as a medium involve their affordances for supporting higher-order cognitive, intrapersonal, and interpersonal outcomes. For all of these reasons, ongoing development should focus on accurate and reliable assessment of higher-order outcomes in order to shift the emphasis and increase the value of future research.

Role of Design and Final Thoughts

Much of the research to date on digital games has focused on proof-of-concept studies and media comparisons. The present meta-analysis highlights the importance of questions that ask not if but how games can support learning. More specifically, the results of the present meta-analysis parallel those of the recent NRC report on laboratory and inquiry activities (Singer, Hilton, & Schweingruber, 2005). Design, rather than medium alone, predicts learning outcomes. Research on games and game-based learning should thus shift emphasis from proof-of concept-studies ("can games support learning?") and media-comparison analyses ("are games better or worse than other media for learning?") to value-added comparisons and cognitive-consequences studies exploring how theoretically-driven design decisions influence learning outcomes for the broad diversity of learners within and beyond our classrooms.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. (2009). *Introduction to meta-analysis.* Chichester, UK: Wiley.

Bransford, J. D., A. L. Brown, and Cocking, R.R., eds. (2000). *How People Learn*. Washington, D.C., National Academy Press.

Bransford, J. D., Sherwood, R. D., Hasselbring, T. S., Kinzer, C. K., & Williams, S. M. (1990). Anchored instruction: Why we need it and how technology can help. Cognition, education, and multimedia: Exploring ideas in high technology, 115-141.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.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher, 18*(1), 32-42.

Charles, T., Bustard, D., & Black, M. (2011). Experiences of promoting student engagement through game-enhanced learning. In *Serious Games and Edutainment Applications* (pp. 425-445). Springer London.

Dickey, M. D. (2006). Game design narrative for learning: appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. *Educational Technology Research and Development, 54*(3), 245–263.

Duval, S., & Tweedie, R. (2000). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*, 89-98.

Echeverria, A., Barrios, E., Nussbaum, M., Amestica, M., & Leclerc, S. (2012). The atomic intrinsic integration approach: A structured methodology for the design of games for the conceptual understanding of physics. *Computers & Education, 59*(2), 806-816.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634.

Federation of American Scientists. (2006). R*eport: Summit on Educational Games: Harnessing the Power of Video Games for Learning.* Washington, D.C.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science, 18(*10), 850-855.

Gee, J. P. (2007). *What Video Games Have to Teach Us About Learning and Literacy. Second Edition: Revised and Updated Edition (2nd ed.*). Palgrave Macmillan.

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. *The handbook of research synthesis and meta-analysis,* 357-376.

Green, C. S. & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*(1), 88-94.

Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of experimental psychology: Human perception and performance, 3*2(6), 1465.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*, 169–206.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107-128.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32,* 341-370.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education, 22*, 1-15.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*, 39-65.

Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243-259.

Higgins, J. P. T., Deeks, J. J., & Altman, D. G. (2008). Special topics in statistics. Pg. 481-529 in J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic review of interventions.* Wiley: Hoboken, NJ.

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

Honey, M. A., & Hilton, M. (Eds.). (2010). *Learning Science through Computer Games and Simulations.* National Research Council. Washington, DC: National Academy Press.

Jüni, P., Witshci, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association, 282*(11), 1054–1060.

Kafai, Y. B. (1996). Learning design by making games: Children's development of strategies in the creation of a complex computational artifact. In Y. B. Kafai & M. Resnick (Eds.), *Constructionism in practice: Designing, thinking and learning in a digital world* (pp. 71–96). Mahwah, NJ: Erlbaum.

Kafai, Y. B. (2006). Playing and making games for learning: Instructionist and constructionist perspectives for game studies. *Games and Culture, 1*(1), 36-40.

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effect of modern computer games on mathematics achievement and class motivation. *Computers & Education, 55*, 427–443.

Killingsworth, S. S., Levin, D. T., & Saylor, M. M. (2011). Analyzing action for agents with varying cognitive capacities. *Social Cognition, 29*(1), 56-73.

Lee, J. J. & Hammer, J. (2011). Gamification in Education: What, How, Why Bother? *Academic Exchange Quarterly, 15*(2), 1-5.

Levin, D. T., Killingsworth, S. S., Saylor, M. M., Gordon, S. M. & Kawamura, K. (2012). Tests of concepts about different kinds of minds: Predictions about the behavior of computers, robots, and people. *Human Computer Interaction, 28*(2), 161-191.

Lim, C. (2008). Global citizenship education, school curriculum and games: Learning Mathematics, English and Science as a global citizen. *Computers & Education, 51* (3), 1073-1093.

López-López, J. A., Lipsey, M. W., Wilson, S. J., Tanner-Smith, E. E., & Van den Noortgate, W. (submitted). *Assessing the influence of moderators in meta-analysis with dependent effect sizes: A Monte Carlo simulation.*

López-López, J. A., Viechtbauer, W., Sánchez-Meca, J., & Marín-Martínez, F. (2010, July). Comparing the performance of alternative statistical tests for moderators in mixed-effects meta-regression models. *Paper presented at the 5th Annual SRSM Meeting, Cartagena (Spain).*

Mahajan, N., & Woodward, A. L. (2009). Seven-month-old infants selectively reproduce the goals of animate but not inanimate agents. *Infancy, 14*(6), 667-679.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: a taxonomy of intrinsic motivations for learning. In R. E. Snow, & M. J. Farr (Eds.), Aptitude, learning, and instruction. *Conative and affective process analyses, Vol. 3 (*pp. 223–253). Hillsdale, NJ: Lawrence Erlbaum.

Mayer, R. E. (2011). Multimedia learning and games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 281–305). Charlotte, NC: Information Age.

National Research Council. (2009). *National Research Council Workshop on Games and Simulations.* October 6-7, 2009, Washington, D.C.

Pellegrino, J., & Hilton, M. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century.* Washington D.C.: National Academy of Sciences.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of educational Psychology, 95*(4), 667.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis.* (pp. 295-315). New York: Russell Sage Foundation.

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

Schunk, D.H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*, 207–231.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences, 4*(3), 321-354.

Sheldon, L. (2011). *The multiplayer classroom: Designing coursework as a game.* Boston, MA: Cengage.

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

Sitzmann T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology 64,* 489–528.

Squire, K. (2011). *Video Games and Learning: Teaching and Participatory Culture in the Digital Age.* New York: Teachers College Press.

Tanner-Smith, E. E., & Tipton, E. (2013). Robust variance estimation with dependent effect sizes: Practical considerations and a software tutorial in Stata and SPSS. *Research Synthesis Methods,* in press. doi:10.1002/jrsm.1091

Tanner-Smith, E. E, Wilson, S. J., & Lipsey, M. W. (2013). The comparative effectiveness of outpatient treatment for adolescent substance abuse: A meta-analysis. *Journal of Substance Abuse Treatment, 44,* 145-158. doi:10.1016/j.jsat.2012.05.006

Tipton, E. (2013). Robust variance estimation in meta-regression with bi*nary dependent effects.* Research Synthesis Methods, 4, 169-187.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

van der Meij, H. Albers, E., & Leemkuil, H. (2011). Learning from games: Does collaboration help? *British Journal of Educational Technology, 42*(4), 655-664.

Vogel J.J., Vogel D.S., Cannon-Bowers J., Bowers C.A., Muse K. & Wright M. (2006) Computer gaming and interactive simulations for learning: a meta-analysis. *Journal of Educational Computing Research 34*, 229–243.

Wilson, S. J., Tanner-Smith, E. E., Lipsey, M. W., Steinka-Fry, K., & Morrison, J. (2011). Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. *Campbell Systematic Reviews, 8*.

Wouters, P., Paas, F., & van Merriënboer, J. J. M. (2008). How to optimize learning from animated models: A review of guidelines base on cognitive load. *Review of Educational Research, 78*, 645–675. doi:10.3102/0034654308320320

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013, February 4). *A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. Journal of Educational Psychology, 105*(2), 249-265.

Appendix: Citations of Articles Included in the Meta-Analysis

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*(1), 235-249. doi:10.1037/a0025595

Akpinar, Y., & Turan, M. (2012). Designing a collaborative learning game: Its validation with a turn-taking control scheme in a primary science unit. E*gitim ve Bilim-Education and Science, 37*(163), 254-267.

Amon, K. L., & Campbell, A. (2008). Can children with AD/HD learn relaxation and breathing techniques through biofeedback video games? *Australian Journal of Educational & Developmental Psychology, 8,* 72-84.

Anderson, J., & Barnett, M. (2011). Using video games to support pre-service elementary teachers learning of basic physics principles. *Journal of Science Education and Technology, 20*(4), 347-362. doi:10.1007/s10956-010-9257-0

Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology, 43*, 993–1003. doi:10.1111/j.1467-8535.2011.01269.x

Barab, S., Pettyjohn, P., Gresalfi, M., Volk, C., & Solomou, M. (2012). Game-based curriculum and transformational play: Designing to meaningfully positioning person, content, and context. *Computers & Education, 58*(1), 518-533. doi:10.1016/j.compedu.2011.08.001

Bartholomew, L.K., Gold, R.S., Parcel, G.S., Czyzewski, D.I., Sockrider, M.M., Fernandez, M., Shegog, R., & Swank, P. (2000). Watch, Discover, Think, and Act: Evaluation of computer-assisted instruction to improve asthma self-management in inner-city children. *Patient Education and Counseling, (39),* 269-280. http://dx.doi.org/10.1016/S0738-3991(99)00046-4

Chang, K-E., Sung, Y.-T., & Lin, S.-Y. (2007). Developing geometry thinking through multimedia learning activities. *Computers in Human Behavior, (23)*, 2212-2229. http://dx.doi.org/10.1016/j.chb.2006.03.007

Chang, K.-E., Wu, L.-J., Weng, S.-E., & Sung, Y.-T. (2012). Embedding game-based problem-solving phase into problem-posing system for mathematics learning. *Computers & Education, 58*(2), 775-786. doi:10.1016/J.Compedu.2011.10.002

Chen, Y.-H., Looi, C.-K., Lin, C.-P., Shao, Y.-J., & Chan, T.-W. (2012). Utilizing a collaborative cross number puzzle game to develop the computing ability of addition and subtraction. *Educational Technology & Society, 15*(1), 354-366.

Cherney, I. D. (2008). Mom, let me play more computer games: They improve my mental rotation skills. Sex Roles, 59(11-12), 776-786. doi:10.2466/pms.98.4.1269-1282

Chimeno, J. S., Wulfsberg, G. P., Sanger, M. J., & Melton, T. J. (2006). The rainbow wheel and rainbow matrix: Two effective tools for learning ionic nomenclature. *Journal of Chemical Education, 83*(4), 651-654. doi: 10.1021/ed083p651

Eagle, M., & Barnes, T. (2009). Experimental evaluation of an educational game for improved learning in introductory computing. *ACM SIGCSE Bulletin, 41*(1), 321-325. doi: 10.1145/1539024.1508980

Echeverría, A., Améstica, M., Gil, F., Nussbaum, M., Barrios, E., & Leclerc, S. (2012). Exploring different technological platforms for supporting co-located collaborative games in the classroom. *Computers in Human Behavior, 28*(4), 1170-1177. doi:10.1016/j.chb.2012.01.027

Echeverría, A., Barrios, E., Nussbaum, M., Améstica, M., & Leclerc, S. (2012). The atomic intrinsic integration approach: A structured methodology for the design of games for the conceptual understanding of physics. *Computers & Education, 59*(2), 806-816. doi:10.1016/j.compedu.2012.03.025

Eseryel, D., Ge, X., Ifenthaler, D., & Law, V. (2011). Dynamic modeling as a cognitive regulation scaffold for developing complex problem-solving skills in an educational massively multiplayer online game environment. *Journal of Educational Computing Research, 45*(3), 265-286. doi:10.2190/EC.45.3.a (Note: report contributes two studies)

Fontana, L., & Beckerman, A. (2004). Childhood violence prevention education using video games. *Information Technology in Childhood Education Annual, 2004(1)*, 49-62.

Habgood, M. P., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*(2), 169-206. doi:10.1080/10508406.2010.508029

Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity, 7*(2), 93-100. http://dx.doi.org/10.1016/j.tsc.2012.03.006

Hauptman, H. (2010). Enhancement of spatial thinking with Virtual Spaces 1.0. *Computers & Education, 54,* 123-135. doi:10.1016/j.compedu.2009.07.013

Helms, D., & Sawtelle, S. M. (2007). A study of the effectiveness of cognitive skill therapy delivered in a video-game format. *Optometry and Vision Development, 38*(1), 19-26.

Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology, 18*(2), 187-208. doi:10.1007/S10956-008-9143-1

Hsu, C.-Y., Tsai, C.-C., & Wang,-H. Y. (2012). Facilitating third graders' acquisition of scientific concepts through digital game-based learning: The effects of self-explanation principles. *Asia-Pacific Education Researcher, 21*(1), 71-82.

Huang, C.-C., Yeh, T.-K., Li, T.-Y., & Chang, C.-Y. (2010). The Idea Storming Cube: Evaluating the effects of using game and computer agent to support divergent thinking. *Educational Technology & Society, 13*(4), 180-191.

Hung, P.-H., Hwang, G.-J., Lee, Y.-H., & Su, I-H. (2012). A cognitive component analysis approach for developing game-based spatial learning tools. *Computers & Education, 59*(2), 762-773. doi:10.1016/j.compedu.2012.03.018

Hwang, G.-J., Sung, H.-Y., Hung, C.-M., Huang, I., & Tsai, C.-C. (2012). Development of a personalized educational computer game based on students' learning styles. *Educational Technology Research and Development, 60*(4), 623-638. doi:10.1007/s11423-012-9241-x

Hwang, G.-J., Wu, P.-H., & Chen, C.-C. (2012). An online game approach for improving students' learning performance in web-based problem-solving activities. *Computers & Education, 59*(4), 1246-1256. doi:10.1016/j.compedu.2012.05.009

Jong, M. S. Y., Shang, J. J., Lee, F. L., Lee, J. H. M., & Law, H. Y. (2006). Learning online: A comparative study of a situated game-based approach and a traditional web-based approach. *Technologies for E-Learning and Digital Entertainment,* 3942, 541-551. doi:10.1007/11736639_65

Kablan, Z. (2010). The effect of using exercise-based computer games during the process of learning on academic achievement among education majors. *Educational Sciences: Theory & Practice, 10*(1), 351-364.

Ke, F. (2008). Alternative goal structures for computer game-based learning. *International Journal of Computer-Supported Collaborative Learning, 3*(4), 429-445. doi:10.1007/s11412-008-9048-2

Ke, F. F. (2008). Computer games application within alternative classroom goal structures: Cognitive, metacognitive, and affective evaluation. *Educational Technology Research and Development, 56*(5-6), 539-556. doi:10.1007/S11423-008-9086-5

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55*(2), 427-443. doi:10.1016/j.compedu.2010.02.007

Lach, T., & Sakshaug, L. (2004). The role of playing games in developing algebraic reasoning, spatial sense, and problem solving. *Focus on Learning Problems in Mathematics, 26*(1), 34-42.

Liu, T.-Y., & Chu, Y.-L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education, 55*(2), 630-643. doi:10.1016/J.Compedu.2010.02.023

Main, S., & O'Rourke, J. (2011). 'New directions for traditional lessons': Can handheld game consoles enhance mental mathematics skills? *Australian Journal of Teacher Education, 36*(2), 43-55.

Martin-Gutiérrez, J., Saorín, J. L., Martín-Dorta, N., & Contero, M. (2009). Do video games improve spatial abilities of engineering students? *International Journal of Engineering Education, 25*(6), 1194-1204.

Miller, D. J., & Robertson, D. P. (2010). Using a games console in the primary classroom: Effects of "Brain Training" programme on computation and self-esteem. *British Journal of Educational Technology, 41*(2), 242-255. doi: 10.1111/j.1467-8535.2008.00918x

Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. British *Journal of Educational Technology, 42*(5), 850-864. DOI: 10.1111/j.1467-8535.2010.01114.x

Mitchell, M.J., & Fox, B.J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction, 40*(4), 315-332. http://dx.doi.org/10.1080/19388070109558353

Moreno, R., & Durán, R. (2004). Do multiple representations need explanations? The role of verbal guidance and individual differences in multimedia mathematics learning. *Journal of Educational Psychology, 96*(3), 492-503. doi:10.1037/0022-0663.96.3.492

Moshirnia, A., & Israel, M. (2010). The educational efficacy of distinct information delivery systems in modified video games. *Journal of Interactive Learning Research, 21*(3), 383-405.

Norton-Barker, C., Corbit, M., Bernstein, R., & Greene, E. (2009). Leveraging game-play in a 3D world: A comparative study in a biology classroom. *International Journal of Gaming and Computer-Mediated Simulations, 1*(2), 17-31. doi: 10.4018/jgcms.2009040102

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. doi:10.1016/j.compedu.2008.06.004

Park, H. (2012). Relationship between motivation and student's activity on educational game. *International Journal of Grid and Distributed Computing, 5*(1), 101-114.

Peng, W. (2009). Design and evaluation of a computer game to promote a healthy diet for young adults. *Health Communication, 24,* 115-127. doi:10.1080/10410230802676490

Rasanen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development, 24(*4), 450-472. doi:10.1016/J.Cogdev.2009.09.003

Rieber, L. P., & Noah, D. (2008). Games, simulations, and visual metaphors in education: Antagonism between enjoyment and learning. *Educational Media International, 45*(2), 77-92. doi:10.1080/09523980802107096

Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Ling Wong, W. (2009). Multimodality and interactivity: Connecting properties of serious games with educational outcomes. *CyberPsychology & Behavior, 12*(6), 691-697. doi:10.1089/cpb.2009.0099

Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior, 28*(2), 648-658. doi:10.1126/science.1736359173635910.112

Sedig, K. (2007). Toward operationalization of 'flow' in mathematics learnware. *Computers in Human Behavior, 23*(4), 2064-2092. doi:10.1016/j.chb.2006.11.001

Segers, E., & Verhoeven, L. (2003). Effects of vocabulary training by computer in kindergarten. *Journal of Computer Assisted Learning, 19*(4), 557-566. doi:10.1046/j.0266-4909.2003.00058.x

Segers, E., & Verhoeven, L. (2005). Long-term effects of computer training of phonological awareness in kindergarten. *Journal of Computer Assisted Learning, 21*(1), 17-27. doi:10.1111/j.1365-2729.2005.00107.x

Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology, 43*(4), 540-560. doi:10.1111/j.1467-8535.2011.01197.x

Sindre, G., Natvig, L., & Jahre, M. (2009). Experimental validation of the learning effect for a pedagogical game on computer fundamentals. *IEEE Transactions on Education, 52*(1), 10-18. doi:10.1109/TE.2007.914944

Smith, S., & Ericson, E. (2009). Using immersive game-based virtual reality to teach fire-safety skills to children. Virtual Reality, 13(2), 87-99. doi:10.1007/S10055-009-0113-6

St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology, 30*(2), 203-219. doi:10.1080/01443410903509259

Suh, S., Kim, S. W., & Kim, N. J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *Journal of Computer Assisted Learning, 26*(5), 370-378. doi:10.1111/j.1365-2729.2010.00353.x

Sung, Y.-T., Chang, K.-E., & Lee, M.-D. (2008). Designing multimedia games for young children's taxonomic concept development. *Computers & Education, 50*(3), 1037-1051. doi:10.1016/j.compedu.2006.07.011

Tüysüz, C. (2009). Effect of the computer based game on pre-service teachers' achievement, attitudes, metacognition and motivation in chemistry. *Scientific Research and Essay, 4*(8), 780-790.

Van Eck, R. (2006). The effect of contextual pedagogical advisement and competition on middle-school students' attitude toward mathematics and mathematics instruction using a computer-based simulation game. *Journal of Computers in Mathematics and Science Teaching, 25*(2), 165-195.

van Meegen, A., & Limpens, I. (2010). How serious do we need to be? Improving information literacy skills through gaming and interactive elements. *LIBER Quarterly, 20*(2), 270-288.

Virvou, M., Katsionis, G., & Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness. *Educational Technology & Society, 8*(2), 54-65. (Note: report contributes two studies)

Vogel, J. J., Greenwood-Ericksen, A., Cannon-Bowers, J., & Bowers, C. A. (2006). Using virtual reality with and without gaming attributes for academic achievement. *Journal of Research on Technology in Education, 39*(1), 105-118.

Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education, 55(*1), 178-187. doi:10.1016/j.compedu.2010.01.003

Yang, J. C., Chen, C. H., & Jeng, M. C. (2010). Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. *Computers & Education, 55*(3), 1346-1356. doi:10.1016/j.compedu.2010.06.005

Yang, Y.-T. C. (2012). Building virtual cities, inspiring intelligent citizens: Digital games for developing students' problem solving and learning motivation. Computers & Education, 59(2), 365-377. doi:10.1016/j.compedu.2012.01.012

Yien, J.-M., Hung, C.-M., Hwang, G.-J., & Lin, Y.-C. (2011). A game-based learning approach to improving students' learning achievements in a nutrition course. *Turkish Online Journal of Educational Technology, 10*(2), 1-10.

Yip, F. W. M., & Kwan, A. C. M. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International, 43(*3), 233-249. doi:10.1080/09523980600641445

SRI International

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

www.sri.com

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

BILL & MELINDA GATES *foundation*