This manuscript has been published and is available in
Studies in Science Education, volume 56, February 11th 2020, at
https://www.tandfonline.com/doi/full/10.1080/03057267.2019.1722420

Impact of Serious Games on Science Learning Achievement Compared with More Conventional Instruction: An Overview and a Meta-Analysis

Martin Riopel[1], Lucian Nenciovici, Patrice Potvin, Pierre Chastenay, Patrick Charland, Jérémie Blanchette Sarrasin, Steve Masson

Département de didactique, Université du Québec à Montréal, Montréal, QC, Canada

Abstract

Serious games have become increasingly available to educators. Empirical studies and meta-analyses have examined their impact on learning achievement. However, natural sciences could have a special relation to serious games by their systematic use of quantitative and predictive models that can generate microworlds and simulations. Since no known meta-analysis on serious games observed a significant impact in the specific context of science learning, the present meta-analysis synthesized results from 79 empirical studies that compared the impact on science learning achievement of instruction using serious games versus instruction using more conventional methods. Consistent with theory and past meta-analyses not specifically related to science learning, post-instruction learning achievement was weakly to moderately higher for declarative knowledge, knowledge retention and procedural knowledge for students taught with serious games. Furthermore, findings of the present work suggest that five moderator variables produced significant effects on the relationship between playing serious games and learning outcomes, and three showed consistent variations in mean effect size that could lead to significance, with more studies and larger samples. These findings are discussed in connection with previous meta-analyses' findings, potential pedagogical implications and possible future research.

Keywords

Science education; Learning achievement; Serious game; Conventional instruction; Meta-analysis; Overview

[1] CONTACT Martin Riopel riopel.martin@uqam.ca

Introduction

Serious Games for Science Learning

It has been widely argued that more active instructional methods must be implemented in science classrooms (Aziz, Z., Nor, S. H. M. & Rahmat, R., 2011; Millar, 2011; Avvisati, 2011; Wieman, 2012). To define these more active instructional methods, one can refer to the general proposition of Bonwell (1991, p. *iii*) for students to be engaged in usual tasks such as listening, reading, writing, discussing or solving problems but also, and most importantly, to be engaged in higher-order thinking tasks such as analysis, synthesis, or evaluation and for which an inquiry-based approach to science produces good examples. The National Research Council (2011, p. 22) has also pointed out that:

> *A growing body of research indicates that engaging students in science processes (inquiry) can motivate and support science learning. However, because inquiry approaches can be difficult for students, teachers, and schools, they are rarely implemented.*

In this particular context, computer simulation and serious games can play a special role because the quantitative and predictive models of science can be used to generate interactive microworlds and simulations that can be freely experienced. "Computer simulations and games have great potential to catalyze and support inquiry-based approaches to science instruction, overcoming curricular and logistical barriers.” (National Research Council, 2011, p. 22).

Serious games are generally defined as digital software the primary purpose of which is learning rather than entertainment (Klopfer, Osterweil & Salen, 2009). They could be a beneficial alternative to other instructional methods (Griffiths, 2002; Munienge & Muhandji, 2012; Scanlon, 2002) and could transform education (Shaffer, Squire, Halverson & Gee, 2004) because: 1) simulation and video games let players participate in worlds otherwise inaccessible to them and thus develop new situated understanding; 2) video games make it possible for players to participate in very large scale communities of practice and to learn by doing the ways of thinking that organize those practices. With regard to the role of serious games in education, some scholars (e.g. Prensky, 2001) have even argued that developing digital-based educational games is a “moral imperative,” as learners of the new generation do not respond as effectively to more conventional instruction. In a widely publicized report following its 2006 Summit on educational games, the Federation of American Scientists (FAS) echoed these opinions by writing that “Given the digital natives’ affinity for digital technologies, digital games for learning could be potentially powerful tools for teaching” (p. 17). Of course, one must recognize the complexity of the “digital natives” notion and raise questions about the empirical evidence supporting it (Helsper & Eynon, 2010), consider the social inequalities related to Internet and digital technology access (Camerini, Schulz & Jeannet, 2018), recognize the complexity of changing policies and practices (Coburn, 2004), and the profound challenges related to implementing constructivist instruction (Windschitl , 2002), but there still appears to be enough potential to justify at least some private and government-sector investment in educational game research.

Despite the potential and the popularity of serious games, there is currently no consensus with regard to their impact on science learning. At times, empirical evidence supports higher science learning achieved by students subjected to serious games in comparison with more conventional instructional methods (e.g. Cameron, 2003; Huppert, Lomask & Lazarowitz, 2002; Kolloffel & de Jong, 2013; Myneni, Narayanan, Rebello, Rouinfar & Pumtambekar, 2013; Pyatt & Sims, 2007, 2012; Zacharia, Olympiou & Papaevripidou, 2008). In this context, more conventional instruction can be considered the opposite of a more active approach (previously defined) and refers mostly to lectures, discussions, textbook readings, exercises and problem solving (McLaren & Kenny, 2015; Waldrop, 2015) or, more generally, of less involvement and fewer higher-order thinking tasks. Of course, one has to acknowledge that the opposition between active and more conventional approaches can be misleading because quality of both active games and conventional instruction have changed over the years. At other times, no difference in science learning achievement is found between serious games in comparison with more conventional instructional methods (Corter, Esche, Chassapis, Ma & Nickerson, 2011; Lang, 2012; Zacharia & Olympiou, 2011; Renken & Nunez, 2013; Wiesner & Lan, 2004). The National Research Council (2011, p. 54) concluded that “Evidence for the effectiveness of games for supporting science learning is emerging, but is currently inconclusive.” They observed that, even if the research on simulations is stronger than the research on games, both have not yet been studied enough to reach a definitive conclusion (Clark *et al.*, 2009). This can be explained partially by the rapid changes in technology and the related difficulty in focusing the research. Another problem is the poor or missing description of the variables describing the context or the students that could also influence learning. Some methodological issues were also identified, such as small sample size, ecological biases related to nested groups, wide range of theoretical perspectives, and variability of instruments to measure learning outcomes. Globally, dissentious views are present among both practitioners and researchers (Brinson, 2015; National Science Teachers Association, 2013; O’Neil, Wainess & Baker, 2005; Vogel *et al*., 2006). Young *et al*. (2012, p. 70) similarly noted that:

> *Despite a decade of research emphasis on STEM education, there has been little peer-reviewed literature published in game-based learning for science, and that which already exists, like that for mathematics, is not consistent in terms of activities being monitored, learning outcomes assessed, or types of science-based gaming being used as the treatment variable.*

More recently, Giessen (2015, p. 2242), after conducting an overview of the field of digital game-based learning, concluded that despite the fact that “it seems language learning is a sphere where computer-based games are quite convincing, […] we are still far away from a general result on whether computer-based gaming or *Serious Games*, for that matter, are successful or not.” Wouters et al. (2013, p. 258) similarly observed that serious games significantly improve learning in all domains except in science.

To fill in the gaps and strengthen the overall quality of the research, the National Research Council (2011) proposed a list of possible actions for researchers such as: explicitly specifying the desired science learning outcomes and studying the role of a simulation or game in advancing these; studying the effects of a simulation and game both in formal and informal contexts; using the simulations and games to assess and support individualized learning. Other actions were also proposed for developers and institutions. All these proposed actions were meant to coordinate and provide guidance to players in the field. Another possible approach to address this currently inconclusive state of the art concerning the impact of serious games on science learning achievement is to conduct a quantitative review of previously published empirical studies, called a *meta-analysis*. This approach provides several benefits, such as the ability to improve the power by combining small or inconclusive studies to answer important questions on a topic, the capacity to identify sources of diversity across various types of studies, and the ability to reveal how heterogeneity among populations, settings and methodologies affects the educational interventions (Fagard, Staessen & Thijs, 1996; Ioannidis & Lau, 1999). Another advantage of a meta-analytical approach is to allow for a comparison of studies that differ in experimental rigor and other methodological factors (Lipsey, 2003). In brief, meta-analyses can increase the precision with which the treatment effect of an intervention, such as serious games, can be estimated (Bartolucci & Hillegass, 2010).

Because it can be argued that natural sciences could have a special relation to serious games and because of the current inconclusiveness of scientific literature with regard to the impact of serious games on science learning achievement, a meta-analysis was conducted in the present work to answer two research questions:

(1) What is the impact of serious games on science learning achievement when compared with more conventional instructional methods?

(2) Which moderator variables influence the relationship between playing serious games and science learning achievement? A moderator variable (i.e. hereinafter referred to as *moderator*) is a continuous (e.g. age, school marks) or discrete (e.g. gender, ethnicity) variable that affects the strength or direction of the relationship between an independent or predictor variable and a dependent or criterion variable (Baron & Kenny, 1986[2,3]).

Answering these questions should provide the best up-to-date high-level description that characterizes the significant impact of serious games on science learning. It is important to note that, because of its general stance related to the available data in the previous studies, the present work could not satisfactorily address some questions that might prove useful to educators or designers and be related to the effectiveness of games at the instructional level, such as: “Why is a given game better than another one?”, “What is the best way to use a given game in school?”, or more generally, “What are the best underlying educational purposes, pedagogical models, or other forms of typologies for games?” It is also important to note that, even if the diversity and limited quality of learning outcomes in science must be acknowledged, the observed significant differences,

---

[2] http://psych.wisc.edu/henriques/mediator.html.
[3] Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

if any, can still be interpreted at a high level with appropriate caution. The focus of the present work therefore is to determine cautiously if the previously studied games, in the specific context of natural sciences and for a short list of moderator variables, had significant impact on measured learning. It is meant to give the best possible answer with available data from literature and serve as one possibly useful starting point for future studies.

Previous Meta-Analyses on Serious Games

Because the present meta-analysis aims to fill the lack of knowledge regarding the impact of serious games on science learning achievement, it is important to look first at relevant past meta-analyses that examined the impact of serious games on learning achievement in general, with an emphasis on how these past meta-analyses can inform us with respect to the definition of serious games and the present research questions. As this overview will point out, the main limitation of these past meta-analyses is that the question of the impact of serious games or simulations in the specific context of science learning achievement was addressed only weakly. Another scope of this overview is to establish the hypotheses of the present work with regard to the research questions. Table 1, at the end of this text, summarizes all the relevant meta-analyses found in the literature. Although this list goes back to 1981 for completeness, only results for the studies produced after the year 2000 are discussed briefly below.

The meta-analysis conducted by Vogel *et al*. (2006) analyzed results from 32 studies from 1986 to 2003. Vogel *et al*.'s meta-analysis focused on pretest/post-test comparisons of learning achievement outcomes for *simulations*, *games* or or an activity combining features of both, versus more conventional instruction. Simulation was defined as an "activity which 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 pre-recorded simulation" (p. 231). A computer game was defined as being "defined as such by the author, or inferred by the reader because the activity has *goals*, is *interactive*, and is *rewarding* (gives feedback)" (p. 231). Comparison conditions ranged from no training to more conventional instructional methods, such as lectures, tutorials and discussions. Learners in the studies included ranged from preschoolers to adults. The main learning outcome considered was *cognitive gain*, which was not explicitly defined, but which seems to consist of both knowledge (e.g. course material) and skill (e.g. flight navigation) acquisition. Moderator variables were analyzed and consisted of (1) gender, (2) level of learner control during navigation (i.e. game controlled by student vs. teacher/computer), (3) type of activity (i.e. simulation, game, combination), (4) age (i.e. from preschool to adult), (5) image realism (i.e. unrealistic/low-quality images, cartoon-like, photo-realistic) and (6) user grouping (i.e. individual vs. group). Vogel *et al*.'s meta-analysis did not focus on a particular subject area, and included studies from various knowledge domains such as psychology, flight navigation, computer science, mathematics, general cognitive skills, science, etc. Overall, Vogel *et al*. found that significantly higher cognitive learning gains were observed in subjects using simulations, games or a combination versus traditional instructional methods (z = 6.05, p < .0001, N = 8549). More specifically, the highest effect size was found for simulations (z = 9.147, p < .0001, N[4] = 2179), with a lower effect size found for games (z = 3.706, p = .0001, N = 2165) and combination (z = 3.209, p = .0007, N = 4205), although Vogel *et al*. noted a low reliability for these results. With respect to other moderators, Vogel *et al*.'s findings suggested a lack of significant differences across genders (z = .9910, p = .1594, N = 394).

[4] In Vogel *et al*.'s meta-analysis, *N* represents the cumulative number of participants from primary studies on which overall effect sizes were computed.

For age, the combination of preschool, elementary, middle, and high school children showed significant results (z = 4.111, p < .0001, N = 6138) favoring simulations and games, while slightly more beneficial, albeit not significantly different, results were obtained for the combination of college students and adult populations (z = 7.434, p < .0001, N = 2336). Likewise, image realism did not have a significant moderator effect, with both the unrealistic simulations and games (z = 5.447, p < .0001, N = 11481148) and the photo-realistic ones (z = 4.105, p < .0001, N = 842) being more beneficial than more conventional instruction. However, significant differences were found with regard to learner control and user grouping. Compared with more conventional instruction, significantly higher cognitive learning gains were observed for studies in which learners controlled their navigation through the system (z = 7.038, p < .0001, N = 3656), and lower, albeit not significantly different, cognitive learning gains for studies where learners were automatically navigated through the system (z = -2.099, p = .018, N = 94). Cognitive learning gains were also significantly higher for studies involving learners individually navigating through the simulation or game (z = 7.352, p < .0001, N = 3413) versus studies in which group navigation was performed (z = 2.222, p = .0131, N = 931), with both forms of navigation being more beneficial than more conventional instruction. However, Vogel *et al*.'s meta-analysis has limitations. Despite providing an explicit definition of the concepts of *simulation* and *game*, Vogel *et al*. included several studies in which the experimental condition did not seem to consist of either a simulation or a game. In such studies (i.e. Andrews, Schwarz & Helme, 1992; Blank, Roy, Sahasrabudhe, Pottenger & Kessler, 2003; Brewster, 1996; Costabile, De Angeli, Roselli, Lanzilotti & Plantamura, 2003; Kekkonen-Moneta & Moneta, 2002), the experimental group was subjected to multimedia or hypermedia instruction consisting, for instance, of hypertext, images, videos, sounds, quizzes, or PowerPoint presentations, but not a simulation or a game. In addition, Vogel *et al.*'s meta-analysis computed general effect sizes for several knowledge domains considered as a whole, thus precluding a conclusion related specifically to science learning achievement.

In a 2011 meta-analysis, Sitzmann and Ely analyzed results from 65 studies from 1976 to 2009. This meta-analysis focused on pretest/post-test comparisons of computer-based *simulation games* versus more conventional instruction in the context of professional adult training. Studies included were in various knowledge domains, such as psychology, business, computer science, mathematics, science, etc. Sitzmann and Ely defined computer-based simulation games as a form of "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). Control conditions in the empirical studies analyzed by Sitzmann and Ely ranged from no training to trainees who received various forms of more conventional instruction as a substitute for the computer-based simulation game. Three types of cognitive outcomes were considered: (1) *declarative knowledge*, referring to trainees' memory of the facts and principles taught in training, (2) *procedural knowledge*, referring to information about how to perform a task or action, and (3) *retention*, referring to the delayed assessment of trainees' declarative knowledge weeks or months after leaving the training environment. Moderators were analyzed and consisted of five theoretical moderators, which were respectively (1) entertainment value of the simulation game (i.e. high vs. low, with the high value being described as the simulation game containing at least one characteristic commonly seen in board games or video games, such as rolling a virtual dice, moving pegs around a board, striving to make the list of top scorers, playing the role of a character in a fantasy world, or shooting foreign objects), (2) activity level of users during navigation (i.e. passive vs. active), (3) level of access to the simulation game (i.e.

unlimited vs. limited), (4) simulation game as sole instructional method vs. accompanied by other forms of instruction, and (5) activity level of the comparison group (i.e. passive instruction [i.e. lecture, reading] vs. active instruction [i.e. hands-on practice, discussion]). Four methodological moderators were also considered, namely (1) random assignment to conditions (vs. no random assignment), (2) rigor of study design (i.e. pretest/post-test comparison vs. post-test only), (3) publication status (i.e. published vs. unpublished in scientific journal), and (4) year of publication/presentation. For declarative knowledge (*d* = 0.28, 95% CI [0.20 – 0.38], *N*[5] = 2 758), procedural knowledge (*d* = 0.37, 95% CI [0.23 – 0.50], *N* = 936), and retention (*d* = 0.22, 95% CI [0.07 – 0.37], *N* = 824), Sitzmann and Ely found significantly higher positive effects for trainees receiving instruction via a simulation game in comparison with trainees receiving instruction via more conventional instruction. With respect to theoretical moderators, all proved significant, with the exception of entertainment value. For entertainment value, it was found that the benefits of high entertainment value (*d* = 0.26, 95% CI [0.11 – 0.41], *N* = 809) were not significantly higher, and were even possibly lower than those of low entertainment value (*d* = 0.38, 95% CI [0.31 – 0.45], *N* = 32163216). However, adults trained with a simulation game learned more, in comparison with a control group, when (1) they were active (*d* = 0.49, 95% CI [0.41 – 0.56], *N* = 32603260) rather than passive (*d* = -0.11, 95% CI [-0.29 – 0.07], *N* = 521) during navigation, (2) they had unlimited (*d* = 0.68, 95% CI [0.54 – 0.82], *N* = 925) rather than limited (*d* = 0.31, 95% CI [0.23 – 0.38], *N* = 2738) access to the simulation game, and (c) the simulation game was used as a supplement to other instructional methods (*d* = 0.51, 95% CI [0.43 – 0.58], *N* = 3109) rather than the standalone instruction (*d* = -0.12, 95% CI [-0.26 – 0.01], *N* = 946).

[5] In Sitzmann & Ely's meta-analysis, *N* represents the cumulative number of participants from primary studies on which overall effect sizes were computed.

Conversely, the control group learned significantly more than the simulation game group when the control group received instruction that actively (*d* = -0.19, 95% CI [-0.33 – -0.05], *N* = 832) rather than passively (*d* = 0.38, 95% CI [0.24 – 0.51], *N* = 970) engaged them in the learning experience. With respect to methodological moderators, random assignment, rigor of the study design and year of publication did not moderate learning from simulation games, relative to the comparison group. Random assignment to a group (*d* = 0.35, 95% CI [0.26 – 0.45], *N* =19971997) yielded a slightly lower, but not significantly different overall effect size than the lack of random assignment (*d* = 0.43, 95% CI [0.34 – 0.53], *N* = 19311931). The more rigorous pretest/post-test study design (*d* = 0.36, 95% CI [0.26 – 0.46], *N* = 18321832) yielded an identical overall effect size with the less rigorous post-test only study design (*d* = 0.36, 95% CI [0.27 – 0.45], *N* = 21932193). For year of publication, the inverse of the sampling error variance weight correlation between the year of publication or presentation and the effect size was not statistically significant (*r* = .16), thus suggesting that the effect of simulation games on learning, relative to the comparison group, has not changed over time. However, publication status produced a significant moderator effect, with effect sizes being significantly larger for published (*d* = 0.52, 95% CI [0.44 – 0.59], *N* = 30323032) than for unpublished (*d* = -0.10, 95% CI [-0.23 – 0.03], *N* = 993) studies, suggesting a probable publication bias effect. As in previous reviews, Sitzmann and Ely's meta-analysis has some limitations. Sitzmann and Ely calculated general effect sizes for several knowledge domains (e.g. psychology, business, computer science, mathematics, science) considered as a whole, and did not draw specific conclusions in connection with science learning achievement. Furthermore, Sitzmann and Ely's meta-analysis focused exclusively on adult workforce trainees, thus precluding generalization to younger learners specifically.

A meta-analysis conducted by Wouters, van Nimwegen, van Oostendorp and van der Spek followed in 2013, and analyzed 38 studies from 1990 to 2012. Wouters *et al*. included pretest/post-test as well as post-test-only comparisons of *serious games* versus more conventional instructional methods. Studies included covered several knowledge domains (e.g. language, mathematics, preparatory education, science) with 22 studies being in the domain of science (i.e. biology and engineering). Serious games were defined as “being interactive, based on a set of agreed rules and constraints, directed toward a clear goal often set by a challenge, and constantly providing feedback, either as a score or as changes in the game world, to enable players to monitor their progress toward the goal” (p. 250). Control conditions in the studies analyzed by Wouters *et al*. comprised a wide range of more conventional instructional methods, such as lectures, reading, drill and practice, or hypertext learning environments. Learners in the studies included ranged from elementary school children to adults. Three types of learning achievement outcomes were considered: (1) *knowledge* learning, as observed by immediate post-test; (2) *cognitive skills* learning, pertaining to more complex cognitive processes, such as when learners apply their knowledge to solve problems; and (3) *retention*, as observed by delayed post-test. Theoretical moderators consisted of (1) arrangement of the comparison group (i.e. active vs. passive instruction), (2) serious game being inclusive (i.e. combined with another instructional method) vs. exclusive (i.e. the only instructional method), (3) number of training sessions (single vs. multiple), (4) group size (i.e. single player vs. group play), (5) instructional domain, (6) age, (7) level of realism (schematic, cartoon-like, photorealistic), (8) presence of a narrative during gameplay (vs. absence). Methodological moderators consisted of (1) publication source (i.e. peer-reviewed journal, proceedings, dissertation), (2) presence of randomization (vs. absence), and (3) experimental design (post-test only vs. pretest/post-test design).

Wouters *et al*. found that serious games were significantly more effective than more conventional instruction concerning knowledge learning (d = 0.27, 95% CI [0.01 – 0.54], N[6] = 948), cognitive skills learning (d = 0.29, 95% CI [0.15 – 0.43], N = 45994599), and retention (d = 0.36, 95% CI [0.07 – 0.68], N = 499). With respect to theoretical moderators, no significant effect was found for arrangement of the comparison group and age. Serious games did not yield more learning, in comparison with more conventional instruction, when the control group engaged in passive instruction rather than active instruction ($z_{active\text{-}passive}$ = 1.38, $p > .05$). Likewise, comparisons of age groups did not yield any significant difference (ps > .1). However, significant effects were found for all other theoretical moderators. Serious games were significantly more effective ($z_{\text{inclusive vs. exclusive}}$ = 1.66, $p < .048$), compared with more conventional instruction, when they were supplemented with other instructional methods (d = 0.41, 95% CI [0.23 – 0.59], k[7] = 29) rather than presented alone (d = 0.20, 95% CI [0.03 – 0.37], k = 48). Games were significantly more effective ($z_{1\text{ session vs. multiple sessions}}$ = 3.94, $p < .003$) when played over multiple training sessions (d = 0.54, 95% CI [0.35 – 0.72], k = 30) rather than a single training session (d = 0.10, 95% CI [-0.07 – 0.26], k = 47). Games were significantly more effective ($z_{\text{individual vs. group}}$ = 2.34, $p < .01$) when learners played in a group (d = 0.66, 95% CI [0.32 – 1.00], k = 13) rather than individually (d = 0.22, 95% CI [0.09 – 0.36], k = 63). With regard to visual realism, schematic games (d = 0.46, 95% CI [0.27 – 0.65], k = 14) were significantly more effective ($z_{\text{schematic vs. cartoon-like}}$ = 1.89, p = .03) than cartoon-like games (d = 0.20, 95% CI [-0.01 – 0.40], k = 20) and significantly more effective ($z_{\text{schematic vs. realistic}}$ = 2.25, p = .01) than photorealistic games (d = 0.14, 95% CI [-0.08 – 0.35], k =

[6] In Wouters *et al*.'s meta-analysis, *N* represents the cumulative number of participants from primary studies on which overall effect sizes were computed.
[7] In Wouters *et al*.'s meta-analysis, *k* represents the cumulative number of effect sizes from primary studies on which overall effect sizes were computed.

32). Findings also suggested that the absence of a narrative (*d* = 0.46, 95% CI [0.18 – 0.73], *k* = 15), compared with the presence of a narrative (*d* = 0.25, 95% CI [0.11 – 0.39], *k* = 62), seemed to yield higher learning gains, although this result did not quite reach statistical significance ($z_{\text{narrative vs. no narrative}} = 1.34$, $p = .09$). For instructional domains, significant difference was observed for all domains except the ones related to natural sciences: biology (d = 0.11, 95% CI [0.11 – 0.33], k = 28) and engineering (d = -0.36, 95% CI [-0.80 – 0.09], k = 6) for which no significant effect was observed. With respect to methodological moderators, unpublished studies (*d* = -0.20, 95% CI [-0.83 – 0.43], *k* = 3) seemed to yield a lower effect size in favor of games than published studies (*d* = 0.36, 95% CI [0.24 – 0.48], *k* = 67), but this result did not reach significance (*p*s > .05). Random assignment (*d* = 0.08, 95% CI [-0.13 – 0.29], *k* = 35) significantly attenuated ($z_{\text{random vs. non-random}} = 2.75$, $p = .003$) the positive learning effect of serious games compared with lack of randomization (*d* = 0.44, 95% CI [0.29 – 0.60], *k* = 42). The experimental design had no effect on the magnitude of the effect size ($z_{\text{post-test only vs. pre-post-test}} = 0.55$, $p > .1$), with post-test only (*d* = 0.25, 95% CI [0.07 – 0.44], *k* = 27) and pre/post-test (*d* = 0.32, 95% CI [0.16 – 0.48], *k* = 50) studies yielding similar effect sizes. Wouters *et al*.'s main limitation with regard to science learning achievement is the rather small sample of 22 studies in the domain of natural sciences that were included in the analysis and the absence of significant effect in this domain.

The last relevant meta-analysis for this study was conducted by Clark, Tanner-Smith and Killingsworth (2015), and they analyzed 69 studies from 2000 to 2012. Clark *et al*. focused on pretest/post-test comparisons of *digital games* versus more conventional instructional methods. Studies included covered several knowledge domains (e.g. language, mathematics, social sciences, natural sciences) with 13 studies in the domain of natural sciences. Participants in primary studies ranged from ages 6 to 25. *Digital games* were defined either as a (1) "digital experience in which the participants (a) strive to achieve a set of fictive goals within the constraints of a set of rules that are enforced by the software, (b) receive feedback toward the completion of these goals (e.g. score, progress, advancement, win condition, narrative resolution), and (c) are intended to find some recreational value" (p. 9), or as a (2) "digital environment explicitly referred to by the authors of the study as a game in the title or abstract" (p. 9). Control conditions in the studies analyzed by Clark *et al*. included more conventional instructional methods such as lectures, reading, drill and practice, or hypertext learning environments. The learning achievement outcome consisted of a unified *cognitive learning outcome* comprising cognitive processes and strategies, knowledge and creativity. Most notable theoretical moderators consisted of (1) play duration (single vs. multiple sessions), (2) additional instruction (game with vs. without additional non-game instruction), (3) player grouping (e.g. single player vs. collaborative team), (4) game type (i.e. rudimentary game [adding points/badges] vs. more school-like game consisting of more than adding points/badges), (5) visual realism (schematic, cartoon, realistic) and (6) story relevance (none, irrelevant, relevant). Most notable methodological moderators consisted of (1) research design (quasi-experimental vs. experimental), (2) assessment type (pre-existing normed instruments vs. author-developed instruments) and (3) quality of game condition reporting, in terms of word count and number of screenshots. Analysis for the overall cognitive learning outcome showed that students in digital game conditions demonstrated a significantly better outcome relative to students in the non-game comparison conditions (g = 0.35, 95% CI [0.20, 0.51], N[8] = 173). For play duration, effect sizes were significantly smaller (b = −0.37, p = .03, 95% CI [−0.70, −0.04]) when games were

[8] In Clark *et al*.'s meta-analysis, *N* represents the cumulative number of effect sizes from primary studies which lead to computation of overall effect sizes.

played in one session (g = 0.08, 95% CI [-0.24, 0.39], N = 43) versus multiple sessions (g = 0.44, 95% CI [0.2, 0.59], N = 166). For additional instruction, the analysis found no evidence (b = 0.04, 95% CI [−0.23, 0.31]) that effect sizes were different depending on whether game conditions included additional non-game instruction (g = 0.36, 95% CI [0.19, 0.52], N = 72) or not (g = 0.32, 95% CI [0.11, 0.52], N = 137). For player grouping, average effects were significantly larger (b = 0.29, p = .03, 95% CI [0.02, 0.56]) for games with single players (g = 0.45, 95% CI [0.29, 0.61], N = 150) relative to those using collaborative teams (g = -0.22, 95% CI [0.32, 0.76], N = 12). For game type, no significant difference (b = 0.28, p = .07, 95% CI [−0.02, 0.57]) was found in the mean effect size across the two categories of rudimentary (g = 0.53, 95% CI [0.27, 0.79], N = 64) and school-like games (g = 0.25, 95% CI [0.08, 0.42], N = 145). For visual realism, effects were significantly larger (b = 0.45, p = .03, 95% CI [0.05, 0.84]) for schematic (g = 0.48, 95% CI [0.13, 0.82], N = 52) than for realistic games (g = -0.01, 95% CI [-0.34, 0.32], N = 36). For story relevance, effects were significantly larger (b = 0.46, p = .01, 95% CI [0.12, 0.81]) for game conditions using irrelevant story lines (g = 0.63, 95% CI [0.33, 0.94], N = 75) compared with those using relevant story lines (g = 0.17, 95% CI [-0.03, 0.37], N = 88). Research design was not significantly associated (b = -0.26, p = .051, 95% CI [-0.51, 0.001]) with effect size magnitude across studies, but was close to significance, with quasi-experimental studies yielding a notably larger average effect size (g = 0.43, 95% CI [0.22, 0.63], N = 96) than experimental studies (g = 0.17, 95% CI [0.004, 0.33], N = 113). For assessment type, results indicated no significant differences (ps > .05) in effect size magnitude across assessment types, although the mean effect sizes were slightly lower for author-developed instruments (g = 0.33, 95% CI [0.11, 0.56], N = 97) compared with pre-existing normed instruments (g = 0.40, 95% CI [0.22, 0.58], N = 89). Finally, for condition reporting, there were no significant differences in mean effect sizes filtering studies based on word counts of the game descriptions. Number of screenshots, however, was significantly correlated with effect sizes for the game versus non-game conditions in the media comparison analyses (b = 0.05, p = .02, 95% CI [0.01, 0.09]), suggesting that studies which included more screenshots of the game condition obtained larger effect sizes. Clark *et al*.'s main limitations with regard to the present work's objectives are the relatively small sample size (N = 13) of studies in the domain of science and the lack of differentiation of the learning outcome, which considered as a single whole knowledge, cognitive processes/strategies and creativity.

Thus, as can be noted from these previous reviews, the main limitation with regard to the present work's objectives is the lack of sufficient examination of the effect of serious games in the domain of natural sciences. Even if several of these meta-analyses (e.g. Clark *et al*., 2015; Sitzmann & Ely, 2011; Vogel *et al*., 2006) included some science-focused games, they did not differentiate the science learning outcome. Besides, for the only meta-analysis (e.g., Wouters et al., 2013) that specifically addressed this question, the science domain included a limited number of primary studies (N = 22) and was the only domain for wich no significant effect on learning was observed.

Research Hypotheses of the Present Meta-Analysis

Definition of Serious Games

Because the present meta-analysis aims to examine the impact of serious games on science learning achievement, it is important to specify what will be meant by a *serious game*. As could be noted in the preceding overview, there is some heterogeneity in the scientific literature with regard to the defining features of a serious game (Garris, Ahlers & Driskell, 2002; Hays, 2005; Manninen, 2002; O'Neil *et al*., 2005; Sitzmann & Ely, 2011; Vogel *et al*., 2006). Some scholars (e.g. Gredler, 1996, 2004; Prensky, 2001) have proposed that there are in fact two types of digital games, namely *serious games* and *simulations*. A *serious game* is generally defined as a digital software the primary purpose of which is learning, rather than entertainment (Bellotti, Kapralos, Lee, Moreno-Ger & Berta1, 2013; Giessen, 2015; Girard *et al*., 2013; Klopfer, Osterweil & Salen, 2009; Quinn, 2015; Wouters, van der Spek & Oostendorp, 2009). In addition, there is general agreement that a serious game has the following features: a goal to be reached (e.g. move to a higher difficulty level), rules and constraints (e.g. what a player can and cannot do, time limits), competition (e.g. against other players or against computer), a dose of fantasy (e.g. a specific context separate from real life in terms of time and space), and fun for users (Bright & Harvey, 1984; Dorn, 1989; Leemkuil, de Jong & Ootes, 2000; Lindley, 2003, 2004; Tobias & Fletcher, 2007; Vogel *et al*., 2006; Wouters *et al*., 2009). A *simulation* shares several of these same features, but, according to several scholars (Bell & Smetana, 2008; Gredler, 1996, 2004; Hays, 2005; Tobias & Fletcher, 2007), it differs from a serious game mainly because of its greater realism and more serious goals. For example, Gredler (1996, p. 522) noted that while a serious game's rules and constraints may be imaginative and exceed the limits of the real world, "a simulation is a realistic approximation of reality and its basis is a dynamic set of relationships among several variables that reflect authentic causal processes." The National Research Council (2011, p. 9) identifies four features in which computer games differ from simulations: 1) they are played for enjoyment in informal contexts, 2) they incorporate goals and rules, 3) they provide feedback on the player's progress; 4) they let the player influence the future state of the game. It has also been acknowledged (National Research Concil, 2010, p. 1) that "the technical and cultural boundaries between modeling, simulation, and games are increasingly blurring."

Despite possible conceptual distinctions between *serious games* and *simulations*, this distinction will not be made in the present work for three reasons. First, several authors have pointed out that these two concepts are frequently used interchangeably by researchers conducting studies in the field (Greenblat & Duke, 1981; Hays, 2005; Rieber, Smith & Noah, 1998; Thomas, Cahill & Santilli, 1997). Second, recent works (e.g. Sitzmann & Ely, 2011; Tennyson and Jorczak, 2008) proposed that there are no longer clear boundaries between these two concepts, because the scientific literature is rich with examples of digital software called *simulations,* which do not reproduce real-world contexts and examples of software called *serious games*, which have very serious goals. Third, the more recent meta-analyses discussed in the preceding overview (i.e. Clark *et al.*, 2015; Sitzmann & Ely, 2011; Wouters *et al*., 2013) did not distinguish between serious games and simulations. Thus, to help comparisons between these past reviews and the present work, a unifying concept of *serious games* with the four preceding features but not excluding simulations will be considered here.

Learning Outcomes

Although serious games could possibly advance many science learning goals (identified by National Research Council, 2011, p. 25) such as motivation, conceptual understanding, science process skills, scientific discourse and identity, research actually provides some evidence only for the first two (motivation and conceptual understanding) and the present work focuses on the second because it is a more essential goal for the educational system. The present meta-analysis aims to examine the overall effect of serious games and simulations versus more conventional instructional methods on science learning achievement outcomes. Of course, future work could focus on motivation as, for comparable learning outcomes, more engaged students is preferable.

Several taxonomies of learning outcomes have been proposed in the scientific literature (Kraiger, Ford & Salas, 1993; O'Neill, 2009; Wouters, van der Spek & van Oostendorp, 2009). To help build on previous reviews' findings, and because of available data, three cognitive learning outcomes will be considered here and will be the same as those frequently examined by previous reviews, namely: (1) *learning* of *declarative* knowledge, as observed by immediate post-test on facts, concepts, principles, theories or models (Chi & Ohlsson, 2005), (2) *retention* of declarative knowledge, as observed by delayed post-tests, and (3) *learning* of *procedural* knowledge, as observed by post-test on how to perform a task (Chi & Ohlsson, 2005). However, as future studies are conducted and provide more data, it will be interesting to focus on more ambitious goals for the rich learning contexts related to serious games, as proposed by Barab & Luehmann (2003) or Shaffer, Squire, Halverson & Gee (2004), to provide resources and tools supporting the complex learning-by-doing process of individuals participating in large communities.

For learning of declarative knowledge, because (i) of the previously discussed theorized advantages of serious games (i.e. more active) compared with more conventional instruction and (ii) the fact that previous meta-analyses (e.g. Sitzmann & Ely, 2011; Wouters *et al*., 2013) mostly concluded that serious games were more beneficial than more conventional instruction, the first hypothesis reads as follows.

> H_1: Instruction using serious games yields higher learning gains in terms of immediate declarative knowledge than instruction using more conventional methods under conditions of similar time investment or engagement.

Similarly, for declarative knowledge retention, because of the theorized advantages of serious games and because previous meta-analyses (e.g. Sitzmann & Ely, 2011; Wouters *et al*., 2013) mostly concluded that serious games were more beneficial than more conventional instruction, the second hypothesis reads as follows.

> H_2: Instruction using serious games yields higher learning gains in terms of declarative knowledge retention than instruction using more conventional methods under conditions of similar time investment or engagement.

For the learning of procedural knowledge, only one previous relevant meta-analysis (Sitzmann & Ely, 2011) examined this outcome and found a beneficial effect of simulation gaming compared with more conventional instruction. At a theoretical level, this finding seems to support the postulate of some scholars, which posited that serious gaming "fosters procedural thinking" (Johnson, Smith, Willis, Levine & Haywood, 2011, p. 7; McClarty *et al*., 2012, p. 16). Therefore, the third hypothesis reads as follows.

> H_3: Instruction using serious games yields higher learning gains in terms of procedural knowledge than instruction using more conventional methods under conditions of similar time investment or engagement.

With regard to the special relation between serious games and natural sciences, it could seem reasonable to propose that the first three hypotheses should be verified with stronger effect sizes when compared with other knowledge domains.

Moderators

As can be noted from the preceding overview, several moderators have been empirically found to affect the relationship between playing serious games and learning outcomes. The following list describes the expected significant moderators that will be tested in the present meta-analysis, and the hypothesis regarding each moderator's effect based on the overview and theoretical postulates from the scientific literature. A dichotomist categorization (i.e. theoretical and methodological moderators), as used by some previously discussed reviews (e.g. Clark *et al*., 2015; Sitzmann & Ely, 2011), will be used to classify moderators. The first four theoretical moderators categorize context (subject area, grade level, duration of intervention, activity level of the comparison group), three more theoretical moderators describe some qualities of the games (ludic content, level of realism, level of user control) and the last four moderators characterize methodology (randomization, experimental design, year of publication, publishing status).

Subject area. Examining the moderator effect of subject area consists of verifying if the impact of serious games, compared to conventional instruction, differs depending on subject area or, in the present case, on scientific discipline (i.e. physics, chemistry, biology, etc.). As pointed out in the present overview, one major limitation of previous reviews (e.g. Clark *et al*., 2015; Dekkers & Donatti, 1981; Lee, 1999; Sitzmann & Ely, 2011; Vogel *et al*., 2006) with regard to the present work's aims is the fact that overall effect sizes were computed by considering primary studies in various knowledge domains (e.g. science, mathematics, social sciences, computer science, etc.) as a single whole, thus precluding a precise conclusion with regard to the domain of science. Some recent meta-syntheses (Brinson *et al*., 2015; Ma & Nickerson, 2006; Rutten, van Joolingen & van der Veen, 2012; Young *et al*., 2012), which examined the impact of serious games on science learning achievement, suggest, however, that serious games could yield different learning outcomes depending on scientific discipline. For example, Rutten *et al*.'s review (2012) suggests that studies on serious games in biology and physics generally yield positive learning outcomes, but that studies in chemistry yield more negative outcomes, such as "no difference in long-term performance" and "not much contribution to conceptual understanding" (p. 142). Young *et al*.'s review (2012) suggests that serious games in physics seem to yield less positive outcomes compared with biology, and that one reason explaining this could be that they "may introduce misconceptions that could interfere with the learning of concepts like force, momentum, and inertia" (p. 73). Thus, because of these previous findings, the fourth hypothesis reads as follows.

> H_4: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is larger for some scientific disciplines than for others.

With regard to the special relationship between serious games and natural sciences, it could seem logical to propose that the observed differences with other knowledge domains should be larger than the differences between natural sciences.

Grade Level. Examining the moderator effect of grade level consists of verifying whether serious games are more beneficial, in comparison with more conventional instruction, depending on learners' educational level (i.e. elementary, secondary, college, etc.). Previous meta-syntheses (Brinson *et al*., 2015; Hew & Cheung, 2010; Oloruntegbe & Alam, 2010) note that a majority of primary studies which examined the impact of serious games on science learning achievement were carried out in higher education settings, such as polytechnics and universities. Brinson (2015, p. 230) suggests it is therefore important to gain a better understanding of serious games' "effectiveness relative to student grade level and cognitive/psychological development because of the proliferation of serious games and the increase in number of online elementary and secondary schools." As pointed out in the present overview, the moderator effect of grade level was examined by some previous meta-analyses. Vogel *et al*. (2006) found that younger learners (i.e. combination of preschool, elementary, middle, and high school children) seemed to benefit less ($z = 4.111$, $p < .0001$, $N = 6138$) from serious games, compared with older learners consisting of a combination of college students and adult populations ($z = 7.434$, $p < .0001$, $N = 2336$), although this comparison failed to quite reach statistical significance. Wouters *et al*. (2013) found similar results, with adults (d =

0.50, 95% CI [-0.10, 1.10]) seemingly benefiting more than children (*d* = 0.30, 95% CI [0.08, 0.52]) from playing serious games, but not to the point of statistical significance. Thus, because of previous meta-analytical findings, the fifth hypothesis reads as follows.

> H_5: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is greater for older learners compared with younger learners.

Duration of Intervention. Examining the moderator effect of duration of intervention consists of verifying whether the effect of serious games on learning achievement, in comparison with more conventional instruction, varies depending on the length of game play. In a 1977 meta-synthesis on serious games, Pierfy already cautioned that short duration studies, in which one group continues with the same instructional method they have been using previously while another group is subjected to a new technological method, being the serious game, are almost always influenced by a Hawthorne or novelty effect biasing the results. Novelty effect has been described as the greater learning *transitorily* achieved by participants in the experimental group because of their increased interest and motivation toward the new technological media, and as the attenuation of this beneficial effect as the duration of exposure to the technological media lengthens (Clark, 1983; Colorado, 1988; Lookatch, 1995; Poppenk, Moscovitch & Köhler, 2010; Timmerman & Kruepke, 2006). The presence of novelty effect in computer-assisted instruction (CAI) in general was confirmed by several meta-analyses (Bayraktar, 2001; Kulik, Bangert & Williams, 1983; Kulik & Kulik, 1986, 1991; Kulik, Kulik & Shwalb, 1986; Liao, 1998, 2007; Roh & Park, 2010; Schenker, 2007). These meta-analyses found that, when compared with more conventional instruction, longer duration of exposure to CAI (e.g. more than four weeks) was associated with smaller learning effect sizes than a median duration of exposure to CAI (e.g. between one and four weeks) which itself was associated with smaller learning effect sizes than a shorter duration of exposure to CAI (e.g. less than one week). For the specific type of CAI that are serious games, the present overview suggests more contradictory findings. Dekkers and Donatti (1981) found a negative correlation between duration of game play, and both learning (r = -.282; p < .05) and retention achieved (r = -.431; p < .05), a result suggesting the presence of a novelty effect and going along well with meta-analyses on CAI in general. On the contrary, Sitzmann and Ely (2011) found that learners benefited more from game play when they had unlimited (d = 0.68, 95% CI [0.54 – 0.82], N = 925) rather than limited (d = 0.31, 95% CI [0.23 – 0.38], N = 2738) access to the serious game. In the same vein as Sitzmann and Ely, Wouters *et al*. (2013) found that serious games were significantly more effective (z_1 = 3.94, p < .003) when played over multiple training sessions (d = 0.54, 95% CI [0.35 – 0.72], k = 30) rather than a single training session (d = 0.10, 95% CI [-0.07 – 0.26], k = 47), while Clark *et al*. (2015) found that effect sizes were significantly smaller (b = −0.37, p = .03, 95% CI [−0.70, −0.04]) when serious games were played in one game session (g = 0.08, 95% CI [-0.24, 0.39], N = 43) versus multiple sessions (g = 0.44, 95% CI [0.2, 0.59], N = 166). However, one limitation of the three last-mentioned meta-analyses on serious games is the fact that only two levels of game play duration were considered (e.g. one session vs. multiple sessions), thus precluding a finer analysis with regard to the effect of game play duration. In accordance with the previously discussed meta-analyses on CAI in general, it is thus possible that learning promoted by serious games needs more than one play session to reach its maximum effect, and that this effect wanes thereafter as game play duration lengthens. Therefore, the sixth hypothesis reads as follows.

> H_6: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, decreases with length of game play duration.

Activity Level of the Comparison Group. Examining the moderator effect of the activity level of the comparison group consists of verifying whether the group playing a serious game shows more learning outcomes when the comparison group receives passive rather than active conventional instruction. There is general agreement in the scientific literature that learning is enhanced when learners actively engage with the course material (Newell, Rosenbloom & Laird, 1989; Sitzmann, Kraiger, Stewart & Wisher, 2006; Webster & Hackley, 1997). Active learning has been found to be effective across several subject matter areas (Freeman *et al*., 2014) and should benefit learning regardless of whether learners are playing a serious game or learning from more conventional instruction. It has been posited (Bell & Kozlowski, 2008; Keith & Frese, 2005, 2008) that active learning is beneficial by helping learners develop a more refined mental model of course material (i.e. declarative knowledge) and the expertise required to apply their knowledge to different circumstances (i.e. procedural knowledge). According to previously discussed meta-analyses (e.g. Sitzmann & Ely, 2011; Wouters *et al*., 2013), learners in comparison groups are considered active when they are using a computerized tutorial, participating in a discussion, completing assignments or exercises, doing hands-on laboratory work, or a combination of these instructional methods. They are considered passive when they are listening to a lecture or PowerPoint presentation, reading textbook or other forms of expository texts, watching a video, or a combination of these instructional methods. The previously discussed Sitzmann & Ely meta-analysis (2011) found that adult trainees learning with serious games learned more than comparisons trainees learning with more conventional instructional methods, when the more conventional instruction was passive rather than active. Although this finding was not replicated by Wouters *et al*.'s (2013) meta-analysis, based on theoretical postulates and Sitzmann and Ely's (2011) finding, the seventh hypothesis reads as follows.

> H_7: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is larger when the comparison group receives passive rather than active instruction.

Ludic Content. Examining the moderator effect of ludic content consists of verifying whether the effect of serious games on learning achievement, in comparison with more conventional instruction, varies depending on how ludic a game is. The *ludic content* can be described as the "entertainment value" of the game (Sitzmann & Ely, 2011), the "enjoyment felt while playing" (Hays, 2005) or very simply, as Prensky (2001) wrote, how "fun" the game is. Baranowski, Buday, Thompson & Baranowski (2008) noted that "fun" is not a concept that is well understood and that typical measures of enjoyment (or fun) have used synonyms of fun (e.g. enjoy, like, interested, pleasurable, energizing). Several scholars (e.g. Goh, Ang & Tan, 2008; Johnson, 1991; Prensky, 2001) posited that more ludic serious games promote greater learning achievement. In a 2009 meta-synthesis examining the effects of serious games on health and physical education, Papastergiou (2009, p. 608) notably concluded that "enjoyment […] seems to account for the effectiveness of the games." In the present overview, it was pointed out that only Sitzmann and Ely's meta-analysis (2011) examined the effect of this moderator on the learning achieved playing serious games, and found learning attained with high entertainment value games (d = 0.26, 95% CI [0.11 – 0.41], N = 809) was not significantly higher than learning achieved with low entertainment value games (d = 0.38, 95% CI [0.31 – 0.45], N = 32163216). Thus, because of theoretical postulates and because of Papastergiou's meta-synthetic findings, the eighth hypothesis reads as follows.

> H_8: The beneficial effect of instruction using serious games on learning achievement outcomes is greater for games with higher ludic content than for games with lower ludic content.

One has to recognize here that the effects of ludic content could also have a complex relation to engagement and duration of gameplay. This testing would be beyond the scope of the present work.

Level of Realism. Examining the moderator effect of level of realism consists of verifying whether the effect of serious games on learning achievement, in comparison with more conventional instruction, differs depending on how realistic the game interface looks. Game realism can be defined as how accurately the game "replicates the real world environment" (Wilson *et al*., 2009, p. 232), or as the physical and psychological similarity between a game and the environment it represents (Crawford, 1984). Several scholars posited that a higher level of game realism enhances learning (e.g. Bell & Smetana, 2008; Dickey, 2007; Goh *et al*., 2008; McClarty *et al*., 2012; Prensky, 2001; Warren, Dondlinger & Barab, 2008). For example, according to Bell and Smetana (2008, p. 3), realism can "bring the subject matter to life" and favor the development of mental constructs about objects, phenomena and processes. Similarly, in a 2011 meta-synthesis on serious games, Mikropoulos and Natsis (p. 769) noted that "real world, authentic tasks […] enable context and content dependent knowledge construction." In the present overview, it was pointed out that three previous meta-analyses (Clark *et al*., 2015; Vogel *et al*., 2006; Wouters *et al*., 2013) examined the moderator effect of realism with rather concurring results. Contrary to previous postulates, Wouters *et al*. ($z_{\text{schematic vs. realistic}}$ = 2.25, p = .01) and Clark *et al*. (b = 0.45, p = .03) found that less realistic (schematic) games were significantly more beneficial than more realistic (photorealistic), while Vogel *et al*. found a slightly higher beneficial effect for schematic games (z = 5.447) than for photorealistic games (z = 4.105), although this comparison did not reach statistical significance. One possible interpretation for these findings is that visual complexity may "distract students from the intended learning content or provide alternative goals within the game that do not support improvement on the assessed outcome measures" (Clark *et al*., 2015, p. 34). Thus, because of the general agreement among these three previous meta-analyses, the ninth hypothesis reads as follows.

> H_9: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is greater for less realistic games than for more realistic games.

Level of User Control. Examining the moderator effect of level of user control consists of verifying whether the effect of serious games on learning achievement, in comparison with more conventional instruction, differs depending on whether a learner can decide what happens during his navigation through the game or not. This moderator can be described subjectively as the sense of being in control which is experienced by the user while using a medium (Csikszentmihalyi, 1990; Gredler, 1996) or objectively as the "users' ability to influence elements of their learning environment, such as […] how they navigate through content, and their pace through the game" (Wilson *et al*., 2009, p. 234). It was posited that learner control over navigation through tasks and activities is a "surprisingly important feature of effective learning games" (Mayo, 2009, p. 80). Similarly, Gifford (1991) emphasized the importance of control in a good serious game. In the present overview, it was pointed out that only one previous meta-analysis examined this moderator. Vogel *et al*. (2006) found that significantly higher learning gains were observed for studies in which learners had control over content, sequence or pace of navigation (z = 7.038, p < .0001, N = 3656), and lower, albeit not significantly different, learning gains for studies in which learners did not have control over any of these elements (z = -2.099, p = .018, N = 94). Another meta-analysis (Sitzmann *et al*., 2006), not discussed in the overview, examined the effect of this moderator in studies comparing the impact of Web-based instruction (WBI) in relation to more conventional classroom instruction and found concurring results. The cumulative findings showed that the extent to which Web-based trainees learned more than classroom trainees was greater when they were afforded a high (*d* = 0.30) rather than a low level of control (*d* = 0.07) during WBI. Thus, the tenth hypothesis reads as follows.

> H_{10}: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is greater when learners control their navigation through the game than when they do not.

Randomization. Examining the moderator effect of randomization consists of verifying whether the effect of serious games on learning achievement, in comparison with more conventional instruction, differs depending on whether a pure experimental (participants randomly assigned to group or randomized control trial) or quasi-experimental (participants not randomly assigned to group) study design is used. Randomized controlled trials are the most rigorous way of determining whether a cause-effect relation exists between treatment and outcome (Girard *et al*., 2013; Sibbald & Roland, 1998), but only a minority of studies which examined the impact of serious games used this design (Connolly, Boyle, MacArthur, Hainey & Boyle, 2012; Oloruntegbe & Alam, 2010). Because an experimental design is more rigorous, it is often posited that effect sizes related to the effect of serious games on learning achievement should be lesser in such design (Bisoglio, Michaels, Mervis & Ashinoff, 2014). In the present overview, it was pointed out that three meta-analyses examined the moderator effect of randomization and all concur that randomization is associated with lower overall effect sizes. Sitzmann and Ely (2011) found that random assignment to a group (*d* = 0.35) yielded a slightly lower, but not significantly different, overall effect size than the lack of random assignment (*d* = 0.43). Wouters *et al*. (2013) found that random assignment to a group (*d* = 0.08) significantly attenuated ($z_{\text{random vs. non-random}} = 2.75$, $p = .003$) the positive learning effect of serious games compared with lack of randomization (*d* = 0.44). Clark *et al*. (2015) found that randomization was not significantly associated (b = -0.26, p = .051) with effect size magnitude across studies, but that it was close to significance, with quasi-experimental studies yielding a notably larger average effect size (g = 0.43) than experimental studies (g = 0.17). Thus, the eleventh hypothesis reads as follows.

> H_{11}: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is greater for studies using a quasi-experimental design than for studies using an experimental design**.**

Experimental Design. Examining the moderator effect of experimental design consists of verifying whether the effect of serious games on learning achievement, in comparison with more conventional instruction, differs depending on whether a study used a pretest/post-test design or a post-test only design. Some scholars (e.g. Pierfy, 1977; Salthouse & Tucker-Drob, 2008) posited that pretest/post-test design poses a risk of a test-retest effect, especially so when the same test is used at both times, thus enhancing the observed effect size due to game playing. Other scholars (e.g. Hays, 2005; Liao, 1998) posited that, on the contrary, repeated measure design is methodologically more rigorous and takes into account learners' various entry levels of knowledge before playing the game. In the present overview, it was pointed out that two meta-analyses examined this moderator and concurred that there is no significant effect with regard to this moderator. Sitzmann and Ely (2011) found that the pretest/post-test study design (d = 0.36) yielded an identical overall effect size with the post-test only study design (d = 0.36). Wouters *et al*. (2013) found that the experimental design had no effect on the magnitude of the effect size ($z_{\text{post-test only vs. pre-post-test}}$ = 0.55, p > .1), with post-test only (d = 0.25) and pretest/post-test (d = 0.32, 95% CI [0.16 – 0.48], k = 50) designs yielding rather similar effect sizes. Thus, the twelfth and last hypothesis reads as follows.

> H_{12}: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is greater for studies using a quasi-experimental design than for studies using an experimental design.

Year of Publication. Examining the moderator effect of year of publication consists of verifying whether the group playing a serious game learns more, compared with the more conventional instruction group, in more recent studies rather than in older studies. Several scholars posited that because technology "advances very rapidly" (Roh & Park, 2010, p. 150) and "is becoming more manipulative, interactive, and 'real' by the day" (Brinson, 2015, p. 230), more recent studies that compared any form of computer-assisted instruction (CAI) to more conventional instruction should probably yield better results in favor of CAI than older studies. According to Liao (1998, p. 353), year of publication is thus an important moderator because it "allows an assessment of the effects of media over time." As can be noted from the present overview, rather conflicting results were found for this moderator. Dekkers and Donatti's (1981) meta-analysis found a positive correlation between date of publication and retention (r = .539; p < .05). Sitzmann & Ely found that the correlation between year of publication and overall effect size for learning was not statistically significant (r = .16, p > .05). Thus, because of the aforementioned theoretical postulates and meta-analytical finding of Dekkers and Donatti (1981), the thirteenth hypothesis reads as follows.

> H_{13}: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is larger in more recent studies compared with older studies.

Publication Status. Examining the moderator effect of publication status consists of verifying whether the effect of serious games on learning achievement, in comparison with more conventional instruction, differs depending on whether a study is published in a peer-reviewed journal or unpublished in this context. We will hereafter designate as “unpublished” the studies (e.g. dissertation theses, acts of proceedings, research reports) that have become available on platforms other than peer-reviewed journals. This moderator is related to the “file drawer problem” (Glass, McGaw & Smith, 1981) that is a potential bias affecting meta-analysis. It concerns the fact that studies included in a meta-analysis may not be a correct representation of all studies that were actually conducted on a subject (Ellis, 2010; Rosenberg, 2005; Rosenthal, 1995). Because statistical analyses of studies published in peer-reviewed journals are more likely to have reached statistical significance and larger effect sizes, compared with studies unpublished in such journals (e.g. acts of proceedings, dissertation theses), to gain a balanced view of the literature on a subject, both published and unpublished studies should be examined (Rothstein, Sutton & Borenstein, 2005). In the present overview, it was pointed out that three previous meta-analyses examined this moderator and they all concur that published studies yield larger effect sizes. Dekkers and Donatti (1981) found a positive correlation between publication status and both learning ($r = .326$; $p < .05$) and retention ($r = .477$; $p < .05$), suggesting that published studies reported larger effect sizes than unpublished studies. Sitzmann and Ely (2011) found that effect sizes were significantly larger across published ($d = 0.52$) compared with unpublished ($d = -0.10$) studies. Wouters *et al*. (2013) found that published studies ($d = 0.36$) seemed to yield larger effect sizes than unpublished studies ($d = -0.20$), but this result did not reach significance ($p > .05$). Thus, the fourteenth hypothesis reads as follows.

> H_{14}: The beneficial effect of instruction using serious games on learning achievement outcomes, in comparison with more conventional instruction, is greater for published than unpublished studies.

Method

The present work used the same classical meta-analytical approach (Glass, McGaw & Smith, 1981; Hunter & Schmidt, 1990; Hunter, Schmidt & Jackson, 1982; Kulik, Kulik & Bangert-Drowns, 1985; Liao, 1998) which was used by recent meta-analyses discussed in the overview. This method can be declined in the following four steps: (1) locate studies through objective and replicable literature search, (2) describe outcomes using a common scale, (3) code these studies for salient characteristics (i.e. moderator variables), (4) use statistical methods to relate the study characteristics to the outcomes.

Literature Search

Computer-based literature searches of three databases (ERIC, Google Scholar, PsycNet) were conducted to locate relevant studies. The algorithm which was used for the search consisted of a set of keywords related to media (e.g. “serious game,” “simulation game,” “virtual simulation”), knowledge domain (e.g. “science,” “physics,” “chemistry,” “biology”), outcome (e.g. “learning,” “learning achievement,” “knowledge gain”) and presence of a control group subjected to more conventional instruction (e.g. “experimental,” “comparison group,” “traditional,” “conventional”). To be included in the initial review, studies had to contain terms relevant to each of these four sets of keywords. This initial search yielded > 1,000 possible studies. A review of abstracts, combined to skim reading of studies when necessary, limited the initial list to 62 potentially relevant studies. Using a snowball technique (Greenhalgh & Peacock, 2005), reference lists of these studies were manually searched for more relevant studies, resulting in a new total of 195 potentially relevant studies. In addition, a manual search was performed through the reference lists of several meta-syntheses (e.g. Brinson, 2015; Girard *et al*., 2013; Lee, 1999; Randel *et al*., 1992) and meta-analyses (e.g. Clark *et al*., 2015; Sitzmann & Evy, 2011; Vogel *et al*., 2006; Wouters *et al*., 2013) on the impact of serious games, resulting in a new total of 238 potentially relevant studies. Finally, researchers with some expertise in the field of serious games were asked to provide leads on possible supplementary studies, resulting in a final total of 242 potentially relevant studies, of which 79 survived application of inclusion criteria.

Inclusion Criteria

To be included in the present meta-analysis, studies had to meet five inclusion criteria: (1) the article had to describe an experimental or quasi-experimental study comparing a group subjected to playing a serious game with a group subjected to non-game instruction; (2) the subject matter had to be in a scientific discipline within the natural and physical sciences (e.g. physics, chemistry, biology, etc.); (3) the article had to present quantitative learning outcome measurement of either declarative knowledge, procedural knowledge or retention; (4) compared groups had to consist of non-disabled learners; (5) the study had to provide sufficient data to allow for calculation of standardized effect sizes (d), such as means (M) with standard deviations (SD), sample sizes (N), t-values or univariate F-values.

Coding of Moderators

With respect to moderators, all the coding was performed on a SPSS 23 data sheet by two of the study's authors. The original inter-rater agreement was 90 percent. Most disagreements concerned randomization, level of user control, level of realism and activity level of the comparison group for which agreement went down to 79 percent. For these moderators, we found that papers were not always explicit and some implicit deduction was necessary. All disagreements on the coding were afterward resolved by discussion and consensus after careful re-examination of corresponding papers. For every moderator, codes used by previously discussed meta-analyses on serious games were used to help comparisons, with one exception: three categories were used to code *duration of intervention* to better account for novelty effect, instead of two categories used by Sitzamnn and Ely (2011), Wouters *et al*. (2013), and Clark *et al*. (2015). Table 2, at the end of the present text, summarizes coding used for each moderator and Table 3 summarizes coding for each moderator and each study.

Calculating Standardized Effect sizes (d) for Learning Outcomes

With respect to the three learning outcomes, all relevant quantitative data (i.e. *M*, *SD*, *N*) and results of statistical inference tests (i.e. *t*-values, *F*-values) were entered onto the same SPSS 23 data sheet. The approach developed by Hedges (1981) and Hedges and Olkin (1985) was used to analyze the data. The standardized effect size computed for every learning outcome in primary studies was *d*, which quantifies the difference between learning achieved by the group subjected to a serious game and the group subjected to more conventional instruction. To this end, Glass, McGaw and Smith's (1981) formulas (i, ii, iii, iv, v) revised by Hunter, Schmidt & Jackson (1982) were used. When means and standard deviations were available, formulas i and ii were used. When a study conducted only a post-test, formula i was used. The mean for the control group (*M*ctrl) was subtracted from the mean for the experimental group (*M*exp), and this difference was divided by the pooled standard deviation for the two groups (*SD*pooled) obtained with formula v. When a study conducted a pretest and a post-test, formula ii was used. The post-test mean for the control group was subtracted from the post-test mean for the experimental group, and this difference was divided by the pooled standard deviation for the two groups at post-test to obtain a post-test quotient. The same computation was done for pretest to obtain a pretest quotient. Then, the pretest quotient was subtracted from the post-test quotient to obtain *d*. When means and standard deviations were not available, either formula iii (*t*-value) or iv (*F*-value) was used to compute *d*.

$$d = (M\text{exp} - M\text{ctrl}) / SD_{\text{pooled}} \quad \text{(i)}$$

$$d = [(M\text{exp} - M\text{ctrl}) / SD_{\text{pooled}}]_{\text{post-test}} - [(M\text{exp} - M\text{ctrl}) / SD_{\text{pooled}}]_{\text{pre-test}} \quad \text{(ii)}$$

$$d = \text{t}\sqrt{}(\frac{1}{N\text{exp}} + \frac{1}{N\text{ctrl}}) \quad \text{(iii)}$$

$$d = \sqrt{}\text{F}\sqrt{}(\frac{1}{N\text{exp}} + \frac{1}{N\text{ctrl}}) \quad \text{(iv)}$$

$$SD_{\text{pooled}} = \sqrt{\frac{(N\text{exp}-1)SD\text{exp}\verb|^|2 + (N\text{ctrl}-1)SD\text{ctrl}\verb|^|2}{(N\text{exp} + N\text{ctrl}-2)}} \quad \text{(v)}$$

$$d_{\text{corr}} = [1 - (\frac{3}{4N-9})] * d \quad \text{(vi)}$$

$$95\%\ \text{CI} = d_{\text{corr}} \pm z_{95\%} * (\frac{SDd}{\sqrt{N}})] \quad \text{(vii)}$$

In addition, because of a possible small sample bias, all effect sizes were then adjusted with the small-sample correction formula (vi) to provide unbiased estimates of effect sizes for learning outcomes (Hedges, 1981; Hedges & Olkin, 1985). In this formula, d_{corr} corresponds to the corrected effect size, *N* corresponds to the cumulative post-test sample size for the experimental and control groups, and *d* corresponds to the original standardized difference effect size computed with formulas i to iv. Then, overall mean *d*s were computed for every learning outcome by weighting for the total sample size of every primary study. A total number of 138 effect sizes were analysed.

Some of the primary studies reported data, for the same learning outcome, from two experimental groups subjected to serious games and/or two control groups subjected to more conventional instruction. In such cases, a weighted mean and pooled standard deviation based on the sample size of each group was first computed for all experimental groups and/or for all control groups, before applying the aforementioned formulas. Likewise, some of the primary studies reported multiple quantitative data (*M*, *SD*) based on the same experimental and/or control group for a single learning outcome. In such cases, one mean and one standard deviation per group was computed per group for that learning outcome, before applying the aforementioned formulas. Finally, 95 percent confidence intervals (CI 95%) were computed around the weighted overall *d*s by applying formula vii. In this formula, $z_{95\%}$ corresponds to the *z*-value (i.e. 1.96) associated with a 5 percent risk of type I error, *SD*d corresponds to the standard deviation on d_{corr}, and *N* corresponds to the total number of participants that contributed to computation of d_{corr}. Confidence intervals present the benefit of assessing the accuracy of the estimate of the mean overall *d*s and provide a reliable estimate of the extent to which the overall mean *d*s are different from zero (Whitener, 1990).

Moderator Analyses

A test for heterogeneity (Hedges & Olkin, 1985; Higgins, Deeks & Altman, 2008; Higgins & Thompson, 2002; Higgins, Thompson, Deeks & Altman, 2003; Kulinskaya & Dollinger, 2015) was first conducted to verify whether effect sizes were consistent across studies. A test for heterogeneity examines the null hypothesis that all studies are evaluating the same effect. For the overall main effects of the three learning outcomes, the set of effect sizes was tested for heterogeneity by performing a Cochran's *Q* analysis (*Q*T statistic). *Q*T is computed by summing the squared deviations of each study's estimate from the overall meta-analytic estimate, weighting each study's contribution in the same manner as in the meta-analysis. QT has an approximate χ^2 distribution with k - 1 degrees of freedom, where k is the number of effect sizes. P-values are obtained by comparing the statistic with this expected χ^2 distribution. If QT is beyond the statistical threshold ($p < .05$), then the null hypothesis of homogeneity is rejected, and the alternative hypothesis of heterogeneity is accepted. Acceptance of heterogeneity indicates that there is more variability in the overall mean effect size than would be expected by chance alone, thus suggesting that it is appropriate to conduct moderator analyses, which was the case here.

The objective of moderator analyses was to determine whether the impact of serious games, relative to more conventional instruction, differed based on each theoretical and methodological study's features described above. Moderating effects were tested by classifying studies according to moderator categories and testing for heterogeneity between categories (Liao, 1998). Because all moderators comprised three or more categories, a univariate analysis of variance was computed to test whether mean effect sizes across categories for each moderator differed by more than chance alone. When that was the case, it was concluded that the moderator had a significant effect on the relation between playing a serious game and learning achieved, and a pairwise comparison of mean effect sizes was performed using a Bonferroni post hoc test to determine which means were significantly different.

Results

First, at a descriptive level, a majority of studies included in the present meta-analysis were conducted in North America (N = 41). A considerable number of studies were conducted in Europe (N = 25). A lesser number of studies were conducted in Asia

(N = 10), while only two studies were conducted in Africa and one study was conducted in Australia. A majority of studies (N = 69) were published in a peer-reviewed journal, three studies were dissertation theses and seven studies were acts of proceedings. A majority of the included studies (N = 46) found that serious games produced significantly greater science learning achievement than more conventional instruction. A considerable number of studies (N = 26) found that there was no significant difference between science learning achieved playing serious games compared with traditional instruction. A lesser number of studies (N = 7) found that more conventional instruction was significantly more effective on science learning achievement than serious games. It is important to note that, given the selection criteria for studies, the selected games could not be considered representative of commercially successful games played by many but less related to school and consequently less studied in this context.

Results for the main effects analyses are presented in Table 4, at the end of this text. Hypothesis $\mathbf{H_1}$ predicted that instruction received with serious games would yield higher learning gain in terms of declarative knowledge compared with instruction received with more conventional methods. Across 65 studies that measured this outcome and provided enough statistics to allow calculation of effect sizes, it was found that declarative knowledge was higher for learners subjected to serious games instruction than for learners subjected to more conventional instruction (d = 0.34, 95% IC [0.25, 0.43], k = 65, N = 7354). The confidence interval for this outcome excluded zero, thus leading to acceptance of hypothesis H_1. This finding suggests that, with respect to declarative knowledge gained, approximately 63 percent of learners in the experimental group would be above average in the control group (Sullivan & Feinn, 2012). The QT statistic (QT = 296.17, df = 64, p < .001) was very significant, meaning that the individual effect sizes which contributed to this overall effect size were heterogeneous and it was thus appropriate to test for moderators.

Hypothesis **H_2** predicted that instruction received with serious games would yield higher learning gain in terms of knowledge retention compared with instruction received through more conventional methods. Across eight studies that measured this outcome and provided enough statistics to allow calculation of effect sizes, it was found that retention of knowledge was higher for learners subjected to serious games instruction than for learners subjected to more conventional instruction (d = 0.31; k = 8; N = 10251025; 95% IC [0.10, 0.52]). The confidence interval for this outcome excluded zero, thus leading to acceptance of hypothesis H_2. This finding suggests that, with respect to knowledge retention, approximately 61 percent of learners in the experimental group would be above average in the control group. The QT statistic (QT = 5.28, df = 7, p > .05) was not significant, meaning that there is not enough heterogeneity between the individual effect sizes which contributed to this overall effect size to justify moderator analyses.

Hypothesis **H_3** predicted that instruction received with serious games would yield higher learning gain in terms of procedural knowledge compared with instruction received through more conventional methods. Across seven studies that measured this outcome and provided enough statistics to allow calculation of effect sizes, it was found that procedural knowledge gain was higher for learners subjected to serious games instruction than for learners subjected to more conventional instruction (d = 0.41; k = 7; N = 556; 95% IC [0.11, 0.71]). In addition, the confidence interval for this outcome excluded zero, thus leading to acceptance of hypothesis H_3. This finding suggests that, with respect to procedural knowledge gain, approximately 66 percent of learners in the experimental group would be above average in the control group. The QT statistic (QT = 7.12, df = 6, p

> .05) was not significant, meaning that there is not enough heterogeneity between the individual effect sizes that contributed to this overall effect size to justify moderator analyses.

As shown in Table 1, the *Q*T value was found to be significant for all three main overall effect sizes, indicating they are heterogeneous and that it is appropriate to conduct moderator analyses. Furthermore, because the three overall mean effect sizes for learning outcomes were relatively similar and their confidence intervals overlapped, a *Q*B analysis was conducted to test for heterogeneity between them. The analysis yielded a result that was not significant ($\chi^2(2) = 1.96$, $p > .05$), suggesting that the overall mean effect sizes for declarative knowledge, knowledge retention, and procedural knowledge were not different by more than simple sampling error. The three learning outcomes were thus combined for the moderator analyses.

Table 5, at the end of this text, summarizes results of the univariate analyses of variance (*F*) that were conducted for every moderator to verify whether its categories influenced significantly the overall effect size. An *F* statistic beyond the threshold for a moderator indicates that the mean effect sizes across the moderator's categories differ by more than chance alone. Such finding suggests that the moderator does have a significant effect affecting the relation between playing a serious games and learning achieved (Lipsey & Wilson, 2001). However, effect size, more than p-value, indicates the possible amplitude or importance of an effect and consistent variation of effect sizes could suggest that, with more studies and larger samples, significant results would be observed. To identify the possible needs for future research, examination of effect sizes was also done systematically for each moderator and consistent variations reaching the limit of 0.10 were identified.

For *subject area*, a nonsignificant *F* statistic was found ($F[3, 61] = 0.179$; $p = .911$), suggesting that scientific discipline did not seem to have a moderating effect on the link between playing a serious game and science learning achieved. Because of this nonsignificant result, a post hoc Bonferroni test was not performed. Examination of the overall mean effect sizes associated with each scientific discipline showed that physics appeared to be associated with the highest overall mean effect size ($d = 0.38$). Physics appeared to be followed in decreasing order by life science ($d = 0.33$), chemistry ($d = 0.24$) and astronomy or Earth science ($d = 0.21$).

For *grade level*, a significant *F* statistic was found ($F[2, 60] = 6.064$; $p = .004$). The pairwise post hoc comparison using the Bonferroni correction showed (1) a significant difference ($\Delta d = 0.53$; $p = .005$) between the overall mean effect sizes for high school students and college/adult students in favor of high school students; (2) a marked difference ($\Delta d = 0.40$; $p = .103$) between the overall means for elementary school students and high school students in favor of high school students, although this difference did not quite reach statistical significance; (3) a small and nonsignificant difference ($\Delta d = 0.13$; $p = .803$) between the overall means for elementary school students and college/adult students in favor of elementary school students.

For *duration of intervention*, a significant *F* statistic was found ($F[3, 61] = 4.227$; $p = .022$). The pairwise post hoc comparison using the Bonferroni correction showed a significant difference ($\Delta d = 0.12$; $p = .026$) between the overall mean effect sizes obtained for (1) studies with an intervention length of less than one week in comparison with (2) studies with an intervention length between one and four weeks, with the difference being in favor of (1). Likewise, the Bonferroni correction showed a significant difference ($\Delta d = 0.13$; $p = .020$) between the overall means obtained for (2) studies with an intervention length between one and four weeks and (3) studies with an intervention length longer than four weeks, with the difference being in favor of (2).

For *activity level of the comparison group*, a marginal, but nonsignificant F statistic was found (F[2, 62] = 2.883; p = .064), suggesting that the activity level of the comparison group did not seem to have a moderating effect on the relation between playing a serious game and science learning achieved. Because of this nonsignificant result, a post hoc Bonferroni test was not performed. However, examination of the overall mean effect sizes associated with the two categories defining the level of activity of the control group (i.e. 1 = more passive vs. 2 = more active) showed that a more passive comparison group appeared to be associated with a higher overall mean effect size (Δd = 0.24) in favor of serious games relative to a more active comparison group.

For *ludic content* of the serious game, a nonsignificant F statistic was found (F[2, 44] = 0.249; p = .781), suggesting that the presence of a higher ludic content in the serious game did not benefit learning more in comparison with a lower ludic content. Moreover, the mean overall effect sizes associated with a lower or a higher ludic content were almost identical. (Δd = 0.03).

For *level of realism* of the serious game, a nonsignificant F statistic was found (F[1, 63] = 1.854; p = .179), suggesting that serious games have no differential effect on science learning achieved depending on their level of realism. Despite this nonsignificant result, the mean overall effect size associated with schematic/unrealistic games (Δd = 0.13) appeared to be larger than the mean overall effect size associated with photorealistic games.

For *level of user control* during the game, a significant *F* statistic was found ($F[2, 55] = 3.582$; $p = .012$). The pairwise post hoc comparison using the Bonferroni correction showed that the mean overall effect size when learners had control over content, sequence or pace of the serious game was larger ($\Delta d = 0.24$) than the mean overall effect size when learners did not have control over any of these three elements.

For *randomization* to group, a nonsignificant *F* statistic was found ($F[2, 57] = 1.741$; $p = .194$). Nevertheless, the overall mean effect size of studies not using randomization to assign learners to groups appeared to be similar ($\Delta d = 0.08$) than the overall mean effect size of studies using randomization.

For *experimental design*, a nonsignificant *F* statistic was found ($F[2, 51] = 0.808$; $p = .608$). The overall mean effect size of studies using a post-test only design appeared to be very similar ($\Delta d = 0.03$) to the overall mean effect size of studies using a pretest-post-test design.

For *year of publication,* a significant *F* statistic was found ($F[2, 62] = 6.993$; $p = .002$). The pairwise post-hoc comparison using the Bonferroni correction showed a significant difference ($\Delta d = 0.26$; $p = .002$) between the overall mean effect sizes for (1) studies published from 2010 to the present day and (3) studies published before 2000, in favor of (1). Bonferroni correction also showed a significant difference ($\Delta d = 0.26$; $p = .018$) between the overall mean effect sizes for (1) studies published from 2010 to the present day and (2) studies published between 2000 and 2009, again in favor of (1).

For *publication status*, a significant *F* statistic was found ($F[2, 62] = 3.198$; $p = .038$). The overall mean effect size for studies published in a peer-reviewed journal was significantly higher (Δd = 0.25) than the overall mean effect size of studies unpublished in a peer-reviewed journal.

Discussion

Learning Outcomes

With regard to hypotheses $\mathbf{H_1}$ to $\mathbf{H_3}$, findings suggest that serious games are more beneficial, in the context of natural sciences and with equivalent instructional time, than more conventional instructional methods on declarative knowledge gain, knowledge retention and procedural knowledge gain. Thus, hypotheses H_1 to H_3 are accepted. These findings are in accordance with findings of meta-analyses discussed in the preceding overview (and not specifically related to natural sciences), which have mostly come to the same conclusions. Moreover, it is interesting to note the similarity between mean overall effect sizes computed in the present work and mean overall effect sizes of past meta-analyses. For example, for *declarative knowledge* gain, an overall effect size of 0.34 was found in the present work, while previous works have found quite similar overall effect sizes of 0.35 for *cognitive learning* outcomes (Clark *et al*., 2015), of 0.27 for *knowledge learning* (Wouters *et al*., 2013) and of 0.28 for *declarative knowledge* gain (Sitzmann & Ely, 2011). For *knowledge retention*, an overall effect size of 0.31 was found in the present work, while previous works have found quite similar overall effect sizes of 0.36 (VanSickle, 1986 Wouters *et al*., 2013). Sitzmann and Ely (2011) found a slightly lower effect size of 0.22 for this same outcome, a result which could possibly be explained by the fact that Sitzmann and Ely's study focused exclusively on adult trainees and, thus, on a very circumscribed educational context. Although points of comparison are more difficult to establish for *procedural knowledge*, the overall effect size of 0.41

found in the present work was relatively similar to the overall effect size of 0.37 found by Sitzmann and Ely's (2011) meta-analysis. Thus, findings of the present meta-analysis seem to confirm that instruction with serious games, compared with more conventional instruction, is associated with a small to moderate positive overall effect size on science learning achievement. Findings also seem to confirm that the significant effects do not differ when considering declarative knowledge gain, knowledge retention or procedural knowledge. In addition, the impact of serious games on science learning achievement does not appear to be different from their overall impact on learning achievement in other domains of knowledge. This general result, before considering the effects of the moderators, does not support the proposition that natural sciences have a special relation to serious games.

Moderators

With regard to hypothesis $\mathbf{H_4}$, findings suggest that serious games are not differentially beneficial on science learning achievement depending on scientific subject area. Thus, hypothesis H_4 is rejected. Despite this finding, the discipline of physics appeared to be associated with the highest overall mean effect size, which contradicts some theoretical postulates. Indeed, because physics is a discipline in which misconceptions, or erroneous beliefs about natural phenomena, are particularly well entrenched among learners, it was earlier pointed out that some scholars (e.g. Young *et al*., 2012) posited that serious games could be detrimental for learning physics because they might induce or consolidate misconceptions. Thus, findings of the present meta-analysis do not support this claim. In addition, the nonsignificant present finding for this moderator could be considered slightly conflictual with previous meta-analyses and meta-syntheses on serious games that examined the moderator effect of subject area. As pointed out in the overview, subject areas compared in these past reviews consisted of various knowledge domains (e.g. mathematics, science, language, etc.) and thus, an inter-domain comparison was conducted. Usually, the conclusion reached was to the effect that learning achieved differed depending on subject area (e.g. Randel *et al*., 1992; Wouters *et al*., 2013). This observed nonsignificant difference (while previous meta-analyses observed significant differences for other knowledge domains) could support the claim that all natural science disciplines (physics, chemistry, biology, etc.), because they are based on the same type of quantitative predictive models, have some special relation to serious games that leads them to be homogeneously beneficial.

With regard to hypothesis $\mathbf{H_5}$, findings suggest that high school students appear to benefit the most from playing serious games. Indeed, high school students appear to achieve higher, although not significantly different, learning gains than the younger elementary school students. Thus, hypothesis H_5 is partly accepted. This finding concurs with findings of previous meta-analyses on serious games (e.g. Vogel *et al*., 2006; Wouters *et al*., 2013), which reported higher learning achieved by older learners. However, these findings also partly conflict with H_5 and previously discussed meta-analyses (e.g. Vogel *et al*., 2006; Wouters *et al*., 2013) and in other meta-analyses on computer-based instruction (CAI) in general. For example, Liao’s (1998) meta-analysis on CAI found that the lowest overall means were achieved by high school students. In the present meta-analysis, high school students achieved significantly higher learning gains than the older college/adult population, while the overall mean gain achieved by elementary school students was slightly higher, although not significantly different, than the college/adult population. This finding suggests a non-monotonic relationship between age, learning sciences and serious games.

With regard to hypothesis **H_6**, findings suggest that a shorter duration of instruction with serious games yields higher learning achievement outcomes than a median duration, which itself yields higher learning achievement outcomes than a longer duration. Thus, hypothesis H_6 is accepted, leading to the conclusion that the probable dissipation of a novelty effect explains the progressive decrease of the main effects. One cannot exclude that it could also mean that games designed to act on short engagement time have some other beneficial effect on learning. This finding concurs with the previously discussed Dekkers and Donatti's (1981) meta-analysis on serious games, which found a negative correlation between game play duration and learning achieved. It also concurs with the previously discussed considerable corpus of meta-analyses which found that shorter computer-based instruction (CBI) was associated with better learning outcomes than longer CBI, and generally concluded the presence of a novelty effect (i.e. Bayraktar, 2001; Kulik, Bangert & Williams, 1983; Kulik & Kulik, 1986, 1991; Kulik, Kulik & Shwalb, 1986; Liao, 1998, 2007; Roh & Park, 2010; Schenker, 2007). However, this finding conflicts slightly with findings of recent meta-analyses on serious games discussed earlier (e.g. Clark *et al*., 2015; Wouters *et al*., 2013), which found that multiple-session gaming was associated with higher learning achievement than single-session gaming. As already pointed out, the most likely explanation for this is the fact that the coding used by Wouters *et al*. (2013) and Clark *et al*. (2015) for length of game play consisted of only two categories, thus not allowing a finer analysis of this moderator. In addition, the category representing the shorter length of game play in the present work (i.e. less than one week) included several studies in which game play lasted for two or three sessions (e.g. Akcay *et al*., 2006; Pyatt & Simms, 2012) or for a single long session (e.g. Barab *et al*., 2009; Tarekegn, 2009), thus allowing more learning to take place.

With regard to hypothesis **H_7**, findings suggest that serious games are not differentially beneficial on science learning achievement depending on the activity level of the comparison group. Thus, hypothesis H_7 is rejected. Despite this finding, it was observed that instruction with serious games appears to be associated with a higher overall mean effect size when instruction received by the comparison group is passive ($d = 0.43$) rather than active ($d = 0.19$). This finding, although not statistically significant, appears to concur with the previously discussed general agreement in the scientific literature that active instruction benefits learning more than passive instruction (Freeman *et al*., 2014; Newell *et al*., 1989; Sitzmann *et al*., 2006; Webster & Hackley, 1997). This finding also partially concurs with the previously discussed Sitzmann & Ely's meta-analysis (2011) on serious games, which found that adult trainees instructed with serious games learned more than trainees instructed with more conventional instructional methods, when the conventional methods were passive ($d = 0.38$) rather than active ($d = -0.19$). It is interesting to note, however, that Sitzmann and Ely's meta-analysis found that an active comparison group appeared to learn more ($d = -0.19$) than a serious game group, a result not replicated here. One possible explanation with regard to these somehow divergent results is the fact that Sitzmann and Ely focused exclusively on adult trainees and, therefore, on a different learning context. Another possible interpretation, which could be put forward, is that it is the activity level of learners, rather than the method of instruction, that influences learning most. It would be consistent with the results from Freeman (2014) that active learning increases students' performance in natural sciences.

With regard to hypothesis **H_8**, findings suggest that serious games with a higher ludic content do not appear to be more beneficial than serious games with a lower ludic content. Thus, hypothesis H_8 is rejected. At a theoretical level, this finding contradicts the previously discussed postulate of several scholars (e.g. Goh, Ang & Tan, 2008; Johnson, 1991; Prensky, 2001), according to which more ludic serious games promote greater learning achievement. This finding also contradicts Papastergiou's (2009, p. 608) meta-synthesis on the effects of serious games in the domain of health and physical education, which observed that "enjoyment […] seems to account for the effectiveness of the games." However, this finding concurs with Sitzmann and Ely's meta-analytical finding (2011), to the effect that learning attained with more ludic games (d = 0.26) was not higher, and even appeared to be lower than learning attained with less ludic games (d = 0.38). One possible explanation for these results is that, when subjected to a serious game with high ludic content, many students might consider the science learning tasks as "fun" (Falvo, 2008) or "enjoyable" (Stobart and Chau, 2002) the same way consider take movies, computer games and other recreation modes as fun and enjoyable, thus deterring them from achieving learning.

With regard to hypothesis $\mathbf{H_9}$, findings suggest that serious games do not appear to be differentially beneficial depending on the level of realism of their interface. Thus, hypothesis H_9 is rejected. Nevertheless, as was pointed out, the mean overall effect size (d = 0.39) associated with schematic, unrealistic games appears to be larger than the mean overall effect size (d = 0.26) associated with photorealistic games. Thus, this finding contradicts the previously discussed postulates of several scholars, which posited that a higher level of game realism enhances learning (e.g. Bell & Smetana, 2008; Dickey, 2007; Goh *et al*., 2008; McClarty *et al*., 2012; Prensky, 2001; Warren, Dondlinger & Barab, 2008). This finding also contradicts the 2011 finding of the meta-synthesis of Mikropoulos and Natsis, who concluded that realistic serious games favor knowledge construction. This apparent contradiction with previous studies could support the proposition that natural sciences have a special relation to serious games favoring schematic interfaces. However, the finding of the present meta-analysis concurs with findings by three recent meta-analyses discussed in the overview that concluded that less realistic (or schematic) serious games significantly benefited (Clark *et al*., 2015; Wouters *et al*., 2013) or appeared to benefit (Vogel *et al*., 2006) learning more than more realistic (or photorealistic) serious games. This finding, as already mentioned, could be explained by the enhanced visual complexity of photorealistic games and the state of cognitive overload it induces. For example, Dansereau (2005, p. 77) note that too great a visual complexity could make the "topic material appear overwhelming to the user, which may cause procrastination and false starts and […] interfere with the imagery required to create mental models of the information being presented." One cannot rule out, however, that this finding could also mean that realism benefits different types of games in a different way.

With regard to hypothesis $\mathbf{H_{10}}$, findings suggest serious games appear to be more beneficial when learners have control over content, sequence or pace of the game compared with when they do not have control over any of these elements. Thus, hypothesis H_{10} is accepted. This finding agrees with previously discussed postulates, by which the sense of being in control during media navigation (Csikszentmihalyi, 1990; Gredler, 1996) and the learner's ability to influence elements of the media environment during navigation (Wilson *et al*., 2009, p. 234) should enhance learning. Moreover, this finding concurs with Vogel *et al*.'s (2006) meta-analysis discussed in the overview, as they observed seemingly higher learning gains when learners had control over any of the same three elements of a serious game, compared with when learners did not have control over any of these three elements. This finding could be explained by the *linearity* of the way the information is presented to the learner during navigation of a medium (Gredler, 1996; Lawless & Brown, 1997; O'Neil *et al*., 2005). For example, Lawless and Brown (1997, p. 118) note that a higher level of learner control over navigation enhances learning, because the learner then has the "opportunity to select what information to access as well as how to sequence the information in a manner that is meaningful to him or her." This finding is also consistent with the constructivist view of learning and suggests that user-centered serious games are more beneficial.

With regard to hypothesis $\mathbf{H_{11}}$, findings suggest that serious games do not yield different learning outcomes when they serve as the experimental treatment in the more rigorous experimental studies compared to the less rigorous quasi-experimental studies. Thus, hypothesis H_{11} is rejected. This result does not seem to concur with findings of meta-analyses previously discussed in the overview. Indeed, both Sitzmann and Ely (2011) and Clark et al. (2015) found that random assignment to group was associated with a lower, although not significantly different overall effect size than the lack of random assignment to group. Wouters *et al*. (2013) found the same, although statistically significant result. Thus results, of the present meta-analysis don't appear to confirm the fact that experimental, more rigorous studies on serious games yield lower effect sizes than quasi-experimental, less rigorous studies.

With regard to hypothesis $\mathbf{H_{12}}$, findings suggest that serious games do not yield different learning outcomes when they are examined with a pretest/post-test design compared with a post-test only design. Thus, hypothesis H_{12} is rejected. This finding thus in agrees with findings of meta-analyses discussed in the overview (Sitzmann & Ely, 2011; Wouters *et al*., 2013), which also concluded that these two types of study designs are associated with similar mean overall effect sizes. Thus, results of the present meta-analysis appear to confirm the fact that the presence of a pretest does not appear to be a significant moderator in the relation between serious game play and learning achieved.

With regard to hypothesis **H_{13}**, findings suggest that more recent studies yield higher learning achievement outcomes than older studies. Thus, hypothesis H_{13} is accepted. This is consistent with the postulate (e.g. Brinson, 2015; Roh & Park, 2010) by which advancements in serious games' technology over time seem to be associated with more beneficial science learning achievement outcomes. This finding also concurs with Dekkers and Donatti's meta-analytical finding discussed earlier, which reported, as early as 1981, a positive correlation between year of publication and learning outcomes. However, this finding does not concur with Sitzmann and Ely's (2011) meta-analytical finding, which reported a nonsignificant correlation between year of publication and learning outcomes. One possible explanation in connection with this is that Sitzmann and Ely's meta-analysis, despite covering a period of more than 30 years (i.e. 1976 to 2009) focused exclusively on adult trainees, and thus was restricted to a limited part of the field of literature research on serious games. Another possible explanation is that recent years, following Sitzmann and Ely's meta-analysis, have seen a particularly rapid rate of technology development in the field of serious games (Ma, Oliveira & Baalsrud Hauge, 2014), thus explaining the higher overall mean effect size associated with studies published from 2010 onwards.

Finally, with regard to hypothesis **H_{14}**, findings suggest that studies published in peer-reviewed journals yield a significantly larger overall mean effect size than studies unpublished in such journals (e.g. acts of proceedings, dissertation theses). Thus, hypothesis H_{14} is accepted. This finding is consistent with findings of previous meta-analyses discussed in the overview (Dekkers & Donatti, 1981; Sitzmann & Ely, 2011; Wouters *et al*., 2013) which came to the same conclusion. Thus, results of the present meta-analysis appear to confirm the reported effect, that studies statistically significant and with larger effect sizes are more likely published in a peer-reviewed journal.

To summarize, among theoretical moderators that categorize context, findings of the present work suggest that students appear to learn more with a serious game, compared with comparison group receiving conventional instruction, when (1) serious games are implemented in secondary schools and (2) serious games are implemented for a shorter duration. Consistent variation of effect sizes also suggests that, with more studies and larger samples, (3) serious games could be shown to have more impact in some disciplines (physics, life science) than others (chemistry, astronomy or Earth science). The finding that high school students appear to benefit more from serious games than younger or older students differs from previous studies not specifically related to natural sciences and cannot be explained with a simple linear (or even monotone) relation between benefit and age. It would require more complex propositions about the relationships between age, learning and serious games that are beyond the scope of the present study. One cannot rule out that these new propositions might contribute to characterizing a special relationship between natural sciences and serious games. This would require more research.

Among theoretical moderators that categorize qualities of the games, findings of the present work suggest that students appear to learn more when (1) they can control their navigation through the game. Consistent variation of effect sizes also suggests that, with more studies and larger samples, (2) schematic/unrealistic games could be shown to have more impact than photorealistic ones. While the presence of ludic content somewhat surprisingly doesn’t seem to produce any effect at all, it would be interesting in future studies to analyze other qualities more precisely (for example, game mechanics, complexity of storyline, multiplayer mode, etc.) to identify the most significant.

Among methodological moderators, findings suggest that (1) year of publication and (2) publication status have a significant effect on the link between instructing students with serious games and science learning achieved. This could suggest that more recent efforts in the serious games space seem to be more efficacious and that there may be a publication bias in the literature on this topic.

Conclusion

The present work aimed to determine whether serious games were more effective, compared with more conventional instruction, on science learning achievement. For all three learning outcomes examined (i.e. declarative knowledge, knowledge retention, procedural knowledge), serious games were found to be more beneficial than conventional instructional methods. The effect size of this benefit was found to be small to moderate, which is consistent with previous meta-analytical findings on the effects of serious games in other domains. The present work thus concludes, about the special relationship of serious games to natural sciences, that the overall effect is as significant and with an amplitude comparable with other domains of knowledge.

Moreover, several theoretical and methodological moderators were found to affect the link between instruction with serious games and science learning achieved. Findings of the present work suggest that five moderators' effects were significant (grade level, duration of intervention, level of user control, year of publication and publication status). Among those that were not significant, three moderators showed small consistent variations of mean effect size (subject area, activity level of comparison group, level of realism) that could lead to significance with more studies and larger samples. Furthermore, some findings about moderators are intriguing and require more research and new proposals that could contribute to characterizing the special relationship between natural sciences and serious games.

Similar to previous meta-analyses on serious games, the present meta-analysis has limitations. For example, it did not examine some moderators frequently found in the literature to have a significant effect between instruction with serious games and learning achieved, such as participants' gender, grouping during game play, or overall quality of the game. It also did not analyze the effect of serious games on other variables, such as motivation.

References

**References marked with an asterisk indicate studies included in the meta-analysis.*

*Akçay, H., Durmaz, A., Tüysüz, C., & Feyzioglu, B. (2006). Effects of computer-based learning on students' attitudes and achievements towards analytical chemistry. *The Turkish Online Journal of Educational Technology*, *5*(1), 44-48.

*Akpan, J. P. (2002). Which Comes First: Computer Simulation of Dissection or a Traditional Laboratory Practical Method of Dissection? *Electronic Journal of Science Education*, *6*(4). Retrieved from http://ejse.southwestern.edu/article/view/7686

*Anderson, J. L. & Barnett, M. (2013). Learning Physics with Digital Game Simulations in Middle School Science. *J Sci Educ Technol*, *22*(6), 914-926.

Andrews, P. V., Schwartz, J., & Helme, R. D. (1992). Students can learn medicine with computers: Evaluation of an interactive computer learning package in geriatric medicine. *Medical Journal of Australia*, *157*, 693-695.

*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 & Education*, *53*, 74–85.

*Arici, A. D. (2008). *Meeting kids at their own game: A comparison of learning and engagement in traditional and 3D MUVE educational-gaming contexts* (Doctoral dissertation, Indiana University, Indiana, USA). Retrieved from https://search.proquest.com/docview/287987558?pq-origsite=gscholar

Avvisati, F. (2011). *Effective Teaching for Improving Students' Motivation, Curiosity, and Self-Confidence in Science: Evidence from PISA 2006* [PowerPoint slides]. Retrieved from https://www.oecd.org/edu/ceri/48077389.ppt

Aziz, Z., Nor, S. H. M., & Rahmat, R. (2011). Teaching Strategies to Increase Science Subject Achievement: Using Videos for Year Five Pupils in Primary School. *World Applied Sciences Journal*, *14*, 8-14.

Barab, S. A., & Luehmann, A. L. (2003). Building sustainable science curriculum: Acknowledging and accommodating local adaptation. *Science Education, 87*(4), 454-467.

*Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., & Warren, S. (2009). Transformational Play as a Curricular Scaffold: Using Videogames to Support Science Education. J*ournal of Science Education and Technology, 18*, 305-320.

Baranowski, T., Buday, R., Thompson, D. I., & Baranowski, J. (2008). Playing for Real: Video Games and Stories for Health-Related Behavior Change. *Am. J. Prev. Med*., *34*(1), 74–82.

*Barnea, N. & Dori, Y. J. (1999). High-School Chemistry Students' Performance and Gender Differences in a Computerized Molecular Modeling Learning Environment. *Journal of Science Education and Technology, 8*(4), 257-271.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Bartolucci, A. A. & Hillegass, W. B. (2010). Overview, Strengths, and Limitations of Systematic Reviews and Meta-Analyses. In F. Chiappelli et al. (eds.), *Evidence-Based Practice: Toward Optimizing Clinical Outcomes* (pp. 17-33). Berlin: Springer.

*Baser, M. & DurmusDurmus, S. (2010). The Effectiveness of Computer Supported Versus Real Laboratory Inquiry Learning Environments on the Understanding of Direct Current Electricity among Pre-Service Elementary School Teachers. *Eurasia Journal of Mathematics, Science & Technology Education*, *6*(1), 47-61.

*Baxter, J. H. & Preece, P.F.W. (1999). Interactive multimedia and concrete three-dimensional modelling. *Journal of Computer Assisted Learning*, *15*, 323–331.

*Bayrak, C. (2008). Effects of computer simulations programs on university students' achievements in physics. *Turkish Online Journal of Distance Education, 9*(4), 53-62.

Bayraktar, S. (2001). A Meta-analysis of the Effectiveness of Computer-Assisted Instruction in Science Education. *Journal of Research on Technology in Education*, *34*(2), 173-188.

Bell, B. S. & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93*, 296-316.

Bell, R. L. & Smetana, L. K. (2008). Using Computer Simulations to Enhance Science Teaching and Learning, In R. L. Bell, J.-G. Newsome, & J. Luft (eds.), *Technology in the Secondary Science Classroom* (pp. 23-32). Arlington, VA: NSTA Press.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P. & Berta1, R. (2013). Assessment in and of Serious Games: An Overview. *Advances in Human-Computer Interaction, Volume 2013,* Article ID 136864, 11 pages.

Bisoglio, J., *Michaels, T., Mervis, J. E., & Ashinoff, B. K.* (2014). Cognitive enhancement through action video game training: Great expectations require greater evidence. *Front. Psychol*., *5*, Article 136.

Blank, G. D., Roy, S., Sahasrabudhe, S., Pottenger, W., & Kessler, G. D. (2003). Adapting multimedia for diverse student learning styles. *JCSC*, *18*(3), 45-58.

Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports.* ERIC Clearinghouse on Higher Education, The George Washington University.

*Boothby, C. & Mountjoy, S. (2009, October). The Dissection Dilemma: Real Dissection versus Virtual Dissection in a Middle School Classroom. In *27th Annual NRMERA Conference, Bridging Theory & Practice: Educational Research in the 21st Century*, Jackson, Wyoming.

*Bozkurta, E. & Ilika, A. (2010). The effect of computer simulations over students' beliefs on physics and physics success. *Procedia Social and Behavioral Sciences, 2*, 4587–4591.

Brewster, J. (1996). Teaching abnormal psychology in a multimedia classroom. *Teaching of Psychology*, *23*(4), 249-252.

Bright, G. W., & Harvey, J. G. (1984). Computer games as instructional tools. *Computers in the Schools, 13,* 73-79.

Brinson, J. R. (2015). Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research. *Computers & Education*, *87*, 218-237.

*Brom C., Preuss M. & Klement D. (2011) Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study.. *Computers & Education 57*, 1971–1988.

Camerini, A.-L., Schulz, P. J., & Jeannet, A.-M. (2018). The social inequalities of Internet access, its use, and the impact on children's academic performance: Evidence from a longitudinal study in Switzerland. *New Media & Society*, *20*(7), 2489–2508

Cameron, B. (2003). The effectiveness of simulation in a hybrid and on-line networking course. *TechTrends*, *47*(5), 18-21.

Chi, M. T. H. & Ohlsson, S. (2005). Complex declarative learning. In K. J. Holyoak & R. G. Morrison (eds.), *Cambridge Handbook of Thinking and Reasoning* (pp. 1-59). New York: Cambridge University Press.

*Chini, J. J., Carmichael, A., Rebello, N. S., & Gire, E. (2010). Comparing students' performance with physical and virtual manipulatives in a simple machines curriculum. Paper presented at the 2010 *Annual Meeting of the American Educational Research Association (AERA): Understanding Complex Ecologies in a Changing World.* Denver, Colorado, USA.

*Civelek, T., Ucar, E., Ustunel, H., & Aydın, M. K. (2014). Effects of a Haptic Augmented Simulation on K-12 Students' Achievement and their Attitudes towards Physics. *Eurasia Journal of Mathematics, Science & Technology Education*, *10*(6), 565-574.

Clark, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Educational Research*, *53*(4), 445-459.

Clark, D.B., Nelson, B., Sengupta, P., and D'Angelo, C. (2009). *Rethinking science learning through digital games and simulations: Genres, examples, and evidence.* Paper commissioned for the National Research Council Workshop on Gaming and Simulations, October 6-7, Washington, DC.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2015). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *XX*(X), 1– 44.

*Climent-Bellido, M. S., Martínez-Jimenez, P., Pones-Pedrajas, A., & Polo, J. (2003). Learning in chemistry with virtual laboratories. *Journal of Chemical Education*, *80*(3), 346-352.

Colorado, R. J. (1988). Computer-assisted instruction research: A critical assessment. *Journal of Research on Computing in Education*, *20*, 226-233.

Coburn, C.E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, *77*(3), 211–244.

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

Corter, J. E., Esche, S. K., Chassapis, C., Ma, J., & Nickerson, J. V. (2011). Process and learning outcomes from remotely-operated, simulated, and hands-on student laboratories. *Computers and Education*, *57*(3), 2054-2067.

Costabile, M., De Angeli, A., Roselli, T., Lanzilotti, R., & Plantamura, P. (2003). Evaluating the educational impact of a tutoring hypermedia for children. *Information Technology in Childhood Education Annual*, *15*(1), 289-308.

Crawford, C. (1984). *The art of computer design.* Berkeley, CA: Osborne/McGraw-Hill.

*Cross, T. R. & Cross, V. E. (2004). Scalpel or mouse: Statistical Comparison of Real & Virtual Frog Dissections. *The American Biology* Teacher, 66(6), 408-411.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper & Row.

Dansereau, D. F. (2005). Node-link mapping principles for visualizing knowledge and information. In S. O. Tergan & T. Keller (Eds.), *Knowledge and information visualization: Searching for synergies* (pp. 61-81). Helderberg, Germany: Springer-Verlag.

*Darrah, M., Humbert, R., Finstein, J., Simon, M., & Hopkins, J. (2014). Are virtual labs as effective as hands-on labs for undergraduate physics? A comparative study at two major universities. *Journal of Science Education and Technology, 23*(6), 803-814.

Dekkers, J. & Donatti, S. (1981). The Integration of Research Studies on the Use of Simulation as an Instructional Strategy. *The Journal of Educational Research, 74*(6), 424-427.

Dickey, M. D. (2007). Game design and learning: A conjectural analysis of how Massively Multiple Online Role Playing Games (MMORPGs) foster intrinsic motivation. *Educational Technology, Research, and Development, 55,* 253–273.

*Dobson, J. (2009). Evaluation of the virtual physiology of exercise laboratory program. *Advances in Physiology Education*, *33*(4), 335-342.

Dorn, D. S. (1989). Simulation games: One more tool on the pedagogical shelf. *Teaching Sociology, 17*(1), 1-18.

*Elangovan, T. & Ismail, Z. (2013). The Effects of Realistic Simulation and Non-Realistic Simulation on Biology Students' Achievement. *Education*, *3*(4), 231-241.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, England: Cambridge University Press.

*Engum, S. A., Jeffries, P., & Fisher, L. (2003). Intravenous catheter training system: Computer-based education versus traditional learning methods. *The American Journal of Surgery*, *186*, 67–74.

Fagard, R. H., Staessen, J. A., & Thijs, L. (1996). Advantages and disadvantages of the meta-analysis approach. *J. Hypertens. Suppl*., *14*(2), 9-12.

Falvo, D. (2008). Animations and simulations for teaching and learning molecular chemistry. *Int. J. Technol. Teach. Learn*. *4*(1), 68-77.

*Farrokhnia, M. R. & Esmailpour, A. (2010). A study on the impact of real, virtual and comprehensive experimenting on students' conceptual understanding of DC electric circuits and their skills in undergraduate electricity laboratory. *Procedia - Social and Behavioral Sciences, 2*(2), 5474-5482.

*Finkelstein, N. D., Perkins, K. K., Adams, W., Kohl, P., & Podolefsky, N. (2005). Can Computer Simulations Replace Real Equipment in Undergraduate Laboratories? *Phys. Rev. ST Phys. Educ. Res*., *1*(010103), 1-8.

*Frederick, M. J. M. (2013). Comparison of student outcomes between computer-based simulated and hands-on lab environments. *International Journal of University Teaching and Faculty Development, 4*(1), 1-8.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordta, H., & Wenderotha, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*, 441-467.

*Gelbart, H., Brill, G., & Yarden, A. (2009). The Impact of a Web-Based Research Simulation in Bioinformatics on Students' Understanding of Genetics. *Res Sci Educ, 39*, 725–751.

Giessen, H. W. (2015). Serious Games Effects: An Overview. *Procedia - Social and Behavioral Sciences*, *174*, 2240 – 2244.

Gifford, B. R. (1991). The learning society: Serious play. *Chronicle of Higher Education*, *7*(August), 7.

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, *29*, 207–219.

Glass, G. V, McGaw, B., & Smith, T. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications.

Goh, D. H., Ang, R. P., & Tan, H. C. (2008). Strategies for designing effective psychotherapeutic gaming interventions for children and adolescents. *Computers in Human Behavior*, *24*, 2217–2235.

Gredler, M. E. (1996). Educational games and simulations: A technology in search of a (research) paradigm. In D. H. Jonassen (ed.), *Handbook of research for educational communications and technology: A project of the association for educational communications and technology* (pp. 521-540). New York: Simon & Schuster Macmillan.

Gredler, M. E. (2004). Games and simulations and their relationship to learning. In D. H. Jonassen (ed.), *Handbook of Research on Educational Communications and Technology* (2nd edition, pp. 571-581). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Greenblat, C. S., & Duke, R. D. (1981). *Principles and practices of gaming simulation.* Beverly Hills, CA: Sage Publications.

Greenhalgh, T. & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *British Medical Journal*, *331*, 1064-1065.

Griffiths, M. (2002). The educational benefits of videogames. *Education and Health*, *20* (3), 47-51.

*Harris, D. (2008). *A comparative study of the effect of collaborative problem solving in a massively multiplayer online game (MMOG) on individual achievement* (Doctoral dissertation, University of San Francisco, SF, CA, USA). Retrieved from http://repository.usfca.edu/cgi/viewcontent.cgi?article=1259&context=diss

*Hawkins, I. & Phelps, A. J. (2013). Virtual laboratory vs. traditional laboratory: which is more effective for teaching electrochemistry? *Chemistry Education Research and Practice, 14*(4), 516-523.

Hays, R. T. (2005). *The effectiveness of Instructional Games: A Literature Review and Discussion* [Technical Report 2005-004]. Naval Air Warfare Center Training Systems Division, Orlando, FL.

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. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Helsper, Ellen and Eynon, Rebecca (2010) Digital natives: where is the evidence? *British Educational Research Journal, 36* (3), 503-520.

*Herga, N. R., Grmek, M. I., & Dinevski, D. (2014). Virtual laboratory as an element of visualization when teaching chemical contents in science class. *The Turkish Online Journal of Educational Technology*, *13*(4), 157-165.

Hew, K. F. & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology, 41*(1), 33–55.

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

Higgins, J. P. T. & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. British Medical Journal, 327, 557-560.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage Publications.

Hunter, J. E., Schmidt, F. L., & Jackson, G, B. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.

*Huppert, J., Lomask, S. M., & Lazarowitz, R. (2002). Computer simulations in the high school: Students' cognitive stages, science process skills and academic achievement in microbiology. *International Journal of Science Education*, *24*(8), 803-821.

*Husmann, P. R., O'Loughlin, V. D., & Braun, M. W. (2009). Quantitative and qualitative changes in teaching histology by means of virtual microscopy in an introductory course in human anatomy. *Anatomical Sciences Education*, *2*(5), 218-226.

Ioannidis, J. P. & Lau, J. (1999). Pooling research results: benefits and limitations of meta-analysis. *Jt Comm. J. Qual. Improv.*, *25*(9), 462-9.

*Jaakkola, T. Nurmi, S. (2008). Fostering elementary school students’ understanding of simple electricity by combining simulation and laboratory activities. *Journal of Computer Assisted Learning*, *24*(4), 271–283.

*Jaakkola, T., Nurmi, S., & Veermans, K. (2011). A comparison of students' conceptual understanding of electric circuits in simulation only and simulation-laboratory contexts. *Journal of Research in Science Teaching*, *48*(1), 71-93.

Johnson, C. (1991). Taking fun seriously: Using cognitive models to reason about interaction with computer games. *Personal and Ubiquitous Computing*, *3*(3), 105–116.

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report.* Austin, Texas: The New Media Consortium. Retrieved from http://net.edu cause.edu/ir/library/pdf/HR2011.pdf

Keith, N. & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology, 90*, 677-691.

Keith, N. & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93*, 59-69.

Kekkonen-Moneta, S., & Moneta, G. B. (2002). E-Learning in Hong Kong: Comparing learning outcomes in online multimedia and lecture versions of an introductory computing course. *Journal of Computers in Mathematics & Science Teaching*, *21*(4), 361-379.

*Kerr, M. S., Rynearson, K., & Kerr, M. C. (2004). Innovative Educational Practice: Using Virtual Labs in the Secondary Classroom. *The Journal of Educators Online*, *1*(1), 1-9.

*Keyser, D. (2010). A Comparative Analysis of Student Learning with a Collaborative Computer Simulation of the Cardiopulmonary System. Paper presented at the *Annual Meeting of the American Educational Research Association*, Denver, CO.

*Kiboss, J. K., Ndirangu, M., & Wekesa, E. W. (2004). Effectiveness of a Computer-Mediated Simulations Program in School Biology on Pupils' Learning Outcomes in Cell Theory. *Journal of Science Education and Technology*, *13*(2), 207-213.

*Kim, J. H., Kim, W. O., Min, K. T., Yang, J. Y., & Nam, Y. T. (2002). Learning by Computer Simulation Does Not Lead to Better Test Performance Than Textbook Study in the Diagnosis and Treatment of Dysrhythmias. *Journal of Clinical Anesthesia*, *14*, 395–400.

*Kim, P. (2006). Effects of 3D Virtual Reality of Plate Tectonics on Fifth Grade Students'Achievement and Attitude Toward Science. *Interactive Learning Environments*, *14*(1), pp. 25 – 34.

*Kinzie, M. B., Strauss, R., & Foss, J. (1993). The Effects of an Interactive Dissection Simulation on the Performance and Achievement of High School Biology Students. *Journal of Research in Science Teaching*, *30*(8), 989–1000.

*Klahr, D. (2003) Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction*, *21*(2), 149-173.

*Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*, *44*(1), 183-203.

Klopfer, E., Osterweil, S., & Salen, K. (2009). *Moving learning games forward.* Cambridge, MA: Education Arcade.

*Kollöffel, B. & de Jong, T. (2013). Conceptual understanding of electrical circuits in secondary vocational engineering education: Combining traditional instruction with inquiry learning in a virtual lab. *Journal of Engineering Education*, *102*(3), 375-393.

*Kozhevnikov, M., Gurlitt, J., & Kozhevnikov, M. (2013). Learning Relative Motion Concepts in Immersive and Non-immersive Virtual Environments. *J Sci Educ Technol*, *22*, 952–962....

Kraiger, K., Ford, J. K. & Salas, E. (1993). Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. *Journal of Applied Psychology*, *78*(2), 311-328.

*Krippendorf, B. B. & Lough, J. (2005). Complete and Rapid Switch From Light Microscopy to Virtual Microscopy for Teaching Medical Histology. *The Anatomical Record (Part B: New Anat.), 285*(B), 19–25.

Kulik, J. A., Bangert, R. L., & Williams, G. W. (1983). Effects of Computer-Based Teaching On Secondary School Students. *Journal of Educational Psychology, 75*(1), 19-26.

Kulik, C.-L. C. & Kulik, J. A. (1986). *Effectiveness of Computer -Based Education in Colleges*. Paper presented at the 69th Annual Meeting of the American Educational Research Association, Chicago, IL, USA.

Kulik, C.-L. C. & Kulik, J. A. (1991). Effectiveness of Computer-Based Instruction: An Updated Analysis. *Computers in Human Behavior, 7*, 75-94.

Kulik, J. A., Kulik, C.-L. C. & Bangert-Drowns, R.-L. (1985). Effectiveness of computer-based education in elementary schools. *Computers in human behavior, 1*, 59-74.

Kulik, C.-L. C., Kulik, J. A., & Shwalb, B. J. (1986). The Effectiveness of Computer-Based Adult Education: A Meta-Analysis. *Journal of Educational Computing Research*, *2*(2), 235-252.

Kulinskaya, E. & Dollinger, M. B. (2015). An accurate test for homogeneity of odds ratios based on Cochran's *Q*-statistic. *BMC Medical Research* Methodology, *15* (49).

*Lalley, J. P., Piotrowski, P. S., Battaglia, B., Brophy, K., & Chugh, K. (2010). A comparison of V-Frog© to physical frog dissection. *International Journal of Environmental & Science Education*, *5*(2), 189-200.

Lang, J. (2012). Comparative study of hands-on and remote physics labs for first year university level physics students. *Transformative Dialogues: Teaching and Learning Journal*, *6*(1), 1-25.

Lawless, K. A., & Brown, S. W. (1997). Multimedia learning environments: Issues of learner control and navigation. *Instructional science, 25*(2), 117-131.

Lee, J. (1999). Effectiveness of computer-based instructional simulation: A meta-analysis. *International Journal of Instructional Media*, *26*(1), 71-85.

Leemkuil, H., de Jong, T., & Ootes, S. (2000). Review of educational use of games and simulations (Knowledge management Interactive Training System Project No. IST-1999-13078). The Netherlands: University of Twente. Retrieved on 8/1/2005 from http://kits.edte.utwente.nl/documents/D l.pdf

Liao, Y-K. C. (1998). Effects of Hypermedia Versus Traditional Instruction on Students' Achievement. *Journal of Research on Computing in Education*, *30*(4), 341-359.

Liao, Y-K. C. (2007). Effects of computer-assisted instruction on students achievement in Taiwan: A meta-analysis. *Computers & Education*, *48*, 216–233.

*Limniou, M., Papadopoulos, N., Giannakoudakis, A., Roberts, D., & Otto, O. (2007). The integration of a viscosity simulator in a chemistry laboratory. *Chemistry Education Research and Practice*, *8*(2), 220–231.

Lindley, C. A. (2003). *Game Taxonomies: A High Level Framework for Game Analysis and Design*. [PDF document]. Retrieved from http://www.gamasutra.com /features/20031003/ lindley_01.shtml

Lindley, C. A. (2004). Narrative, Game Play, and Alternative Time Structures for Virtual Environments. In S. Göbel *et al.* (Eds.), 2*nd International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, LNCS 3105, pp. 183–194, Darmstadt, Germany.

Lipsey, M. W. (2003). Those Confounded Moderators in Meta-Analysis: Good, Bad, and Ugly. *Annals of the American Academy of Political and Social Sciences*, *587*, 69-81.

Lipsey, M. W. & Wilson, D. B. (2001). *Applied social research methods series Vol. 49: Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

*Liu, H.-C., & Su, I.-H. (2011). Learning residential electrical wiring through computer simulation: The impact of computer-based learning environments on student achievement and cognitive load. *British Journal of Educational Technology*, *42*(4), 598–607.

Lookatch, R. P. (1995). Technology for teaching? The research is flawed. *Education Digest*, *61*(3), 4-8.

Ma, J. & Nickerson, J. V. (2006). Hands-On, Simulated, and Remote Laboratories: A Comparative Literature Review. *ACM Computing Surveys*, *38*(3), Article 7.

Ma, M., Oliveira, M., & Baalsrud Hauge, J. (2014) Serious Games Development and Applications: Proceedings of the Fifth International Conference on Serious Games Development and Applications (SGDA 2014). *Lecture Notes in Computer Science*, 8778. Berlin, Germany: Springer International Publishing.

*Mamo, M., Namuth-Covert, D., Guru, A., Nugent, G., Phillips, L., Sandall, L., Kettler,T., & McCallister, D. (2011) Avatars Go to Class: A Virtual Environment

Soil Science Activity. *Journal of Natural Resources & Life Sciences Education*, 40, 114-121.

Manninen, T. (2002). Contextual virtual interaction as part of ubiquitous game design and development. *Personal and Ubiquitous Computing*, *6*(5), 390-406.

*Marszalek, C. S. & Lockard, J. (1999, February). Which Way To Jump: Conventional Frog Dissection CD-Tutorial, or Microworld? In *21st Proceedings of Selected Research and Development Papers Presented at the National Convention of the Association for Educational Communications and Technology*, Houston, TX.

Martin, M. O., Mullis, I., Foy, P., & Stanco, G. M. (2011). *TIMMS 2011: International results in science*. Retrieved from http://timssandpirls.bc.edu/timss 2011/downloads/T11_IR_S_Chapter1.pdf

*Martinez, G., Naranjo, F. L., Perez, A. L., Suero, M. I., & Pardo, P. J. (2011). Comparative study of the effectiveness of three learning environments: Hyper-realistic virtual simulations, traditional schematic simulations and traditional laboratory. *Physical Review Special Topics – Physics Education Research*, *7*, 020111 (2011).

Mayo, M. J. (2009). Video Games: A Route to Large-Scale STEM Education? *Science*, *32*(2), 79-82.

McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). *A Literature Review of Gaming in Education: Research Report*. Retrieved from http://www.academia.edu/30930659/A_Literature_Review_of_Gaming_in_Education_Research_Report

McLaren, H. J. & Kenny, P. L. (2015). Motivating change from lecture-tutorial modes to less traditional forms of teaching. *Australian Universities Review*, *57*(1), 26-33.

*Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Cifuentes, L., Kwok, O., & Davis, T. J. (2013). Exploring 3-D virtual reality technology for spatial ability and chemistry achievement. *Journal of Computer Assisted Learning*, *29*, 579-590.

Millar, R. (2011). *Effective science education for innovation*. Retrieved from https://www.oecd.org/edu/ceri/48076916.ppt

*Montgomery, L. (2008). *A Comparison of the Effectiveness of Virtual and Traditional Dissection on Learning Frog Anatomy in High School* (Doctoral dissertation, Wilmington University, New Castle, DE, USA).

Munienge, M. & Muhandji, K. (2012). The use of ICT in Education: a comparison of traditional pedagogy and emerging pedagogy enabled by ICT's. In *Proceedings of* *the 11th International Conference on Frontiers in Education: Computer Science & Computer Engineering* (FECS'12), Las Vegas, Nevada, USA.

*Myneni, L. S., Narayanan, N. H., Rebello, S., Rouinfar, A., & Pumtambekar, S. (2013). An interactive and intelligent learning system for physics education. *IEEE Transactions on Learning Technologies*, *6*(3), 228-239.

National Research Council. (2010). *The rise of games and high-performance computing for modeling and simulation.* Committee on Modeling, Simulation, and Games. Division on Engineering and Physical Sciences. Washington, DC: The National Academies Press.

National Research Council. (2011). Learning Science Through Computer Games and Simulations. In M. A. Honey & M. L. Hilton (Eds.), *Committee on Science Learning: Computer Games, Simulations, and Education*. Washington, DC: The National Academies Press.

National Science Teachers Association. (2013). *NSTA position statement: The integral role of laboratory investigations in science instruction*. Retrieved from http://www.nsta.org/about/positions/laboratory.aspx.

Newell, A., Rosenbloom, P. S., & Laird, J. E. (1989). Symbolic architectures for cognition. In M. I. Posner (eds.), *Foundations of cognitive science* (pp. 93-131). Cambridge, MA: Bradford Books/MIT Press.

Oloruntegbe, K. O. & Alam, G. M. (2010). Evaluation of 3d environments and virtual realities in science teaching and learning: The need to go beyond perception referents. *Scientific Research and Essays*, *5*(9), 948-954.

*Olympiou, G. & Zacharias, C. (2012). Blending Physical and Virtual Manipulatives: An Effort to Improve Students' Conceptual Understanding Through Science Laboratory Experimentation. *Science Education*, *96*(1), 21–47.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes evidence from the computer games literature. *The Curriculum Journal*, *16*(4), 455-474.

O'Neill, G. (2009). *Guide to Taxonomies of Learning outcomes* [PDF document]. Retrieved from http://www.ucd.ie/t4cms/taxonomies3.pdf

Papastergiou, M. (2009). Exploring the potential of computer and video games for health and physical education: A literature review. *Computers & Education*, *53*, 603–622.

*Perry, J., Meir, E., Herron, J. C., Maruca, S., & Stal, D (2008). Evaluating Two Approaches to Helping College Students Understand Evolutionary Trees through Diagramming Tasks. *CBE—Life Sciences Education*, *7*, 193–201.

Pierfy, D. A. (1977). Comparative simulation game research-stumbling blocks and stepping stones. *Simulation and Games, 8*, 255-268.

Poppenk, J., Moscovitch, M. & Köhler, S. (2010). Revisiting the novelty effect: When familiarity, not novelty, enhances memory. Journal of Experimental Psychology Learning: Memory and Cognition*, 36*(5), 1321-30.

*Predavec, M. (2001). Evaluation of E-Rat, a computer-based rat dissection, in terms of student learning outcomes. *Journal of Biological Education, 35*(2), 75-80.

Prensky, M. (2001). *Digital Game-Based Learning*. New York: McGraw-Hill Education.

*Pyatt, K. & Sims, R. (2007). Learner performance and attitudes in traditional versus simulated laboratory experiences. In *Proceedings of the Ascilite*: *ICT: Providing choices for learners and learning* (pp. 870-879), Singapore, Singapore.

*Pyatt, K. & Sims, R. (2012). Virtual and Physical Experimentation in Inquiry-Based Science Labs: Attitudes, Performance and Access. *J Sci Educ Technol*, *21*, 133–147.

Quinn, C. (2015). *Making the Distinction Between Serious Games and Gamification – Guest Post*. Retrieved from http://www.litmos.com/blog/articles/ making-the-distinction-between-serious-games-and-gamification-guest-post-by-clark-quinn

Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: A review of recent research. *Simulation & Gaming*, *23*(3), 261-276.

Renken, M. D. & Nunez, N. (2013). Computer simulations and clear observations do not guarantee conceptual understanding. *Learning and Instruction*, *23*, 10–23.

Rieber, L. P., Smith, L., & Noah, D. (1998). The value of serious play. *Educational Technology, 38*(6), 29-37.

Roh, K. H. & Park, H.-A. (2010). A Meta-Analysis on the Effectiveness of Computer-Based Education in Nursing. *Healthc. Inform. Res.*, *16*(3), 149-157.

Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59*(2), 464-8.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118,* 183–192.

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

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education, 58,* 136–153.

Salthouse, T. A. & Tucker-Drob, E. M. (2008). Implications of short term retest effects for the interpretation of longitudinal change. *Neuropsychology*, *22*(6), 800–811.

Scanlon, E., Morris, E., di Paolo, T., & Cooper, M. (2002). Contemporary approaches to learning science: technologically-mediated practical work. *Studies in Science Education*, *38*(1), 73-114.

Shaffer, D.W., Squire, K.D., Halverson, R., & Gee, J.P. (2004). Video Games and the Future of Learning. *The Phi Delta Kappan, 87*(2), 104-111.

Schenker, J. D. (2007). *The effectiveness of technology use in statistics instruction in higher education: A meta-analysis using hierarchical linear modeling* (Doctoral dissertation, Kent State University, Ohio, USA).

*Scoville, S.A. & Buskirk, T.D. (2007). Traditional and Virtual Microscopy Compared Experimentally in a Classroom Setting. *Clinical Anatomy*, *20*, 565–570.

*Shegog, R., Lazarus, M. M., Murray, N. G., Diamond, P. M., Sessions, N., & Zsigmond, E. (2012). Virtual Transgenics: Using a Molecular Biology Simulation to Impact Student Academic Achievement and Attitudes. *Res Sci Educ*. *42,* 875–890.

Sibbald, B. & Roland, M. (1998) Understanding controlled trials: why are randomised controlled trials important? *British Medical* Journal, *316*, 201.

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

Sitzmann, T., Kraiger, K., Stewart, D. et Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology*, *59*, 623–664.

Stobart, S. & Chau, P. (2002). Postgraduate student perspectives of using an on-line learning environment. *Turk. Online J. Distance Edu*., *3*(3).

Sullivan, G. M. & Feinn, R. (2012). Using Effect Size—or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279-282.

*Sun, K.-T., Lin, C.-L., & Wang, S.-M. (2010). A 3-D virtual reality model of the Sun and and the Moon for e-learning at elementary schools. *International Journal of Science and Mathematics Education*, *8*(4), 689-710.

*Sun, K.-T., Lin, Y.-C., & Yu, C.-J. (2008). A study on learning effect among different learning styles in a Web-based lab of science for elementary school students. *Computers & Education*, *50*, (2008), 1411–1422.

*Taghavi, S. E. & Colen Jr, C. (2009). Computer Simulation Laboratory Instruction vs. Traditional Laboratory Instruction in Digital Electronics. *Journal of Information Technology Impact*, *9*(1), 25-36.

*Tan, J. & Biswas, G. (2007, March). Simulation-Based Game Learning Environments: Building and Sustaining a Fish Tank. In *Proceedings of the DIGITEL 2007, The First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning*, (pp. 73-80), Jhongli, Taiwan.

*Tarekegn, G. (2009). Can computer simulations substitute real laboratory apparatus? Lat. Am. J. Phys. Educ., 3(3), 506-517.

*Tarng, W., Chang, M.-Y., Ou, K.-L., Yu, K.-F., & Hsieh, K.-R. (2012). The Development of a Virtual Farm for Applications in Elementary Science Education. International Journal of Distance Education Technologies, 10(2), 1-16.

*Tatli, Z. & Ayas, A. (2013). Effect of a Virtual Chemistry Laboratory on Students' Achievement. *Educational Technology & Society*, *16*(1), 159–170.

Tennyson, R. D. & Jorczak, R. L. (2008). A conceptual framework for the empirical study of instructional games. In H. F. O'Neil & R. S. Perez (Eds.), *Computer games and team and individual learning* (pp. 39-54). Oxford, UK: Elsevier.

Thomas, R., Cahill, J., & Santilli, L. (1997). Using an interactive computer game to increase skill and self-efficacy regarding safer sex. *Health Education and Behavior*, *24*(1), 71-86.

Timmerman, C. E. & Kruepke, K. A. (2006). Computer-Assisted Instruction, Media Richness, and College Student Performance. *Communication Education*, *55*(1), 73-104.

Tobias, S. & Fletcher, (2007). What Research Has to Say about Designing Computer Games for Learning. *Educational Technology*, *47*(5), 20-29.

*Trundle, K. C. & Bell, R. L. (2010). The use of a computer simulation to promote conceptual change: A quasi-experimental study. *Computers & Education*, *54*, 1078–1088.

*Tsai, S.-L., Chai, S.-K., Hsieh, L.-F., Lin, S., Taur, F.-M., Sung, W.-H., & Doong, J.-L. (2008). The Use of Virtual Reality Computer Simulation in Learning Port-A Cath Injection. *Advances in Health Sciences Education*, *13*, 71–87.

*Tüysüz, C. (2010). The Effect of the Virtual Laboratory on Students' Achievement and Attitude in Chemistry. *International Online Journal of Educational Sciences, 2*(1), 37-53.

*Ulen, S., Cagran, B., Slavinec, M., & Gerlic, I. (2014). Designing and Evaluating the Effectiveness of Physlet-Based Learning Materials in Supporting Conceptual Learning in Secondary School Physics. *J Sci Educ Technol*, *23*, 658–667.

*Ünlü, Z. K. & Dökme, I. (2011). The effect of combining analogy-based simulation and laboratory activities on Turkish elementary school students' understanding of simple electric circuits. *The Turkish Online Journal of Educational Technology, 10*(4), 320-329.

VanSickle, R. L. (1986). Quantitative Review of Research on Instructional Simulation Gaming: A Twenty-Year Perspective. *Theory and Research in Social Education*, *XIV*(3), 245-264.

*Virvou, M., Katsionis, G., et Manos, K. (2005). Combining Software Games with Education: Evaluation of its Educational Effectiveness. *Educational Technology & Society*, *8*(2), 54-65.

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*(3), 229-243.

Waldrop, M. M. (2015). The Science of Teaching Science. *Nature*, *523*, 272-274.

Warren, S. J., Dondlinger, M. J., & Barab, S. A. (2008). A MUVE towards PBL writing: Effects of a digital learning environment designed to improve elementary school writing. *Journal of Research on Technology in Education*, *41*(1), 113–140.

Webster, J. & Hackley, P. (1997). Teaching effectiveness in technology-mediated distance learning. *Academy of Management Journal*, *40*, 1282-1309.

*White, B. Y. (1984). Designing Computer Games to Help Physics Students Understand Newton's Laws of Motion. *Cognition and Instruction*, *1*(1) 69-108.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology*, *75*, 315-321.

Wieman, C. E. (2012). *Applying New Research to Improve Science Education. Issues in Science and Technology*, XXIX(1). Retrieved from http://issues.org/29-1/carl/

Wiesner, T. F. & Lan, W. (2004). Comparison of Student Learning in Physical and Simulated Unit Operations Experiments. *Journal of Engineering Education*, *93*(3), 195-204.

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., Orvis, K. L., & Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming, 40,* 217-266.

Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of educational research, 72*(2), 131-175.

*Wong,W. L., Shen, C., Nocera, L., Carriazo, E., Tang, F., Bugga, S., … Ritterfeld, U. (2007, June). Serious video game effectiveness. In *Proceedings of the international conference on Advances in computer entertainment technology,* (pp. 49–55), Salzburg, Austria.

Wouters, P., van der Spek, E., & van Oostendorp, H. (2009). Current practices in serious game research: A review from a learning outcomes perspective. In T. Connolly, M. Stansfield, & L. Boyle (eds.), *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices* (pp. 232-250).

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*, 249–265.

*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*, 178–187.

*Yang, K.-Y. & Heh, J.-S. (2007). The Impact of Internet Virtual Physics Laboratory Instruction on the Achievement in Physics, Science Process Skills and Computer Attitudes of 10th-Grade Students. *J Sci Educ Technol*, *16*, 451–461.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, … Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research, 82*(1), 61–89.

*Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: An effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted* Learning, *23*, 120–132.

*Zacharia, Z. C. & Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the Physics by Inquiry curriculum: The case of undergraduate students' conceptual understanding of heat and temperature. *Am. J. Phys*., *76*(4-5).

*Zacharia, Z. C. & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction*, *21*, 317-331.

Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of Experimenting with Physical and Virtual Manipulatives on Students' Conceptual Understanding in Heat and Temperature. *Journal of Research in Science Teaching*, *45*(9), 1021-1035.

Table 1. *Past meta-analyses and some meta-syntheses that examined the impact of serious games on learning achievement*

Year	Authors	Number ofo primary studies	Main conclusion with regard to learning achieved with serious games versus more conventional instruction
1981	Dekkers and Donatti	93	Games as effective as more conventional
1986	vanSickle	42	Games more effective to a small degree
1992	Randel, Morris, Wetzel, and Whitehill	68	Games more effective for retention, but not for immediate learning gain
1999	Lee	19	Games combined with conventional more effective than games alone
2006	Vogel, Vogel, Cannon-Bowers, Bowers, Muse and Wright	32	Games more effective than conventional for cognitive gain outcomes
2011	Sitzmann and Ely	65	Games more effective than conventional for declarative and procedural knowledge acquisition and retention
2013	Wouters, van Nimwegen, van Oostendorp and van der Spek	38	Games more effective for knowledge learning and retention
2015	Clark, Tanner-Smith and Killingsworth	69	Games more effective for cognitive learning outcomes

Table 2. *Coding for each moderator variable*

Moderator	**Codes**
Subject area	1 = physics; 2 = chemistry; 3 = life sciencescience; 4 = astronomy or Earth science; 5 = not specified
Grade level	1 primary; 2 = secondary; 3 = college or higher; 4 = not specified
Duration of intervention	1 = less than 1 week; 2 = between 1 and 4 weeks; 3 = more than 4 weeks; 4 = not specified
Activity level of the comparison group	1 = mostly passive instruction, such as listening to a lecture or to a presentation, reading a textbook, watching a video; 2 = mostly active instruction, such as hands-on practice, discussion, exercises, problem solving; 3 = not specified
Ludic content	1 = high: serious game contains at least one characteristic commonly seen in board games or video games, such as rolling a virtual dice, moving pegs around a board, striving to make the list of top scorers, playing the role of a character in a fantasy world, or shooting foreign objects; 2 = low: serious game does not contain any of these characteristics; 3 = not specified
Level of realism	1 = schematic; 2 = cartoon-like; 3 = photorealistic; 4 = not specified
Level of user control	1 = learner has control over content; sequence or pace; 2 = learner has no control over any of these elements; 3 = not specified
Randomization	1 = learners randomly assigned to groups; 2 = learners not randomly assigned to groups; 3 = not specified
Experimental design	1 = post-test only design; 2 = pretest and post-test design; 3 = not specified
Year of publication	1 = 2010 to present; 2 = between 2000 and 2009; 3 = before 2000
Publishing status	1 = published in a peer-reviewed journal; 2 = not published in a peer-reviewed journal; 3 = not specified

Table 3. *Coding for each moderator variable and each study*

N	First author	Year	SA	GL	DOI	ALCG	LC	LR	LUC	R	ED	YP	PS
1	Akçay	2006	2	3	1	1	2	4	3	3	2	2	1
2	Akpan	2002	3	2	1	2	2	3	2	1	2	2	1
3	Anderson	2013	1	2	2	2	1	2	1	2	2	1	1
4	Annetta	2009	3	2	1	1	1	2	1	2	1	2	1
5	Arici	2008	3	1	2	1	1	2	1	1	2	2	2
6	Barab	2009	3	3	1	1	1	2	1	2	1	2	1
7	Barnea	1999	2	2	2	1	2	1	3	2	2	3	1
8	Baser	2010	1	3	2	2	2	1	1	2	2	1	1
9	Baxter	1999	4	2	1	1	2	1	1	1	2	3	1
10	Bayrak	2008	1	3	3	2	2	1	1	1	2	2	1
11	Boothby	2009	3	2	1	2	2	2	1	2	2	2	2
12	Bozkurta	2010	1	3	4	3	3	1	3	1	2	1	1
13	Brom	2011	3	2	1	1	1	2	1	1	1	1	1
14	Chini	2010	1	3	1	2	2	2	1	2	2	1	2
15	Civelek	2014	1	3	2	1	2	2	1	2	1	1	1
16	Climent-Bellido	2003	2	3	3	2	2	1	1	2	1	2	1
17	Cross	2004	3	3	1	2	2	3	1	1	1	2	1
18	Darrah	2014	1	3	3	2	2	3	1	1	2	1	1
19	Dobson	2009	3	3	4	2	2	2	2	1	1	2	1
20	Elangovan	2013	3	2	2	1	2	3	2	2	2	1	1
21	Engum	2003	3	3	1	2	2	3	1	1	2	2	1
22	Farrokhnia	2010	1	3	3	2	2	2	3	1	2	1	1
23	Finkelstein	2005	1	3	1	2	2	2	1	2	2	2	1
24	Frederick	2013	5	3	3	2	2	4	3	2	1	1	1
25	Gelbart	2009	3	3	2	3	2	1	2	2	2	2	1
26	Harris	2008	3	1	2	1	1	3	1	2	2	2	2
27	Hawkins	2013	2	3	1	2	2	4	1	1	2	1	1
28	Herga	2014	2	1	4	3	3	2	3	1	1	1	1
29	Huppert	2002	3	2	2	3	2	1	1	2	2	2	1
30	Husmann	2009	3	3	3	2	2	3	1	3	1	2	1
31	Jaakkola	2008	1	1	1	2	2	2	1	1	2	2	1
32	Jaakkola	2011	1	1	1	2	2	2	1	1	2	2	1
33	Kerr	2004	2	2	2	3	2	3	2	3	2	2	1
34	Keyser	2010	3	2	2	2	2	3	1	2	2	1	2
35	Kiboss	2004	3	2	2	1	2	2	2	1	2	2	1
36	Kim	2002	3	3	1	1	2	3	2	1	2	2	1
37	Kim	2006	4	1	1	1	2	3	1	2	2	2	1
38	Kinzie	1993	3	2	1	2	2	3	1	1	2	3	1
39	Klahr	2003	1	1	1	2	2	3	1	1	2	2	1
40	Klahr	2007	1	2	1	2	1	1	2	1	2	2	1

N	First author	Year	SA	GL	DOI	ALCG	LC	LR	LUC	R	ED	YP	PS
41	Kollöffel	2013	1	3	3	2	2	3	1	1	2	1	1
42	Kozhevnikov	2013	1	3	1	2	2	3	1	1	2	1	1
43	Krippendorf	2005	3	3	3	2	2	3	3	2	1	2	1
44	Lalley	2010	3	2	1	2	2	3	1	2	2	1	1
45	Limniou	2007	2	3	1	2	2	1	1	1	1	2	1
46	Liu	2011	1	2	3	3	2	2	1	2	1	1	1
47	Mamo	2011	4	3	1	2	1	3	1	1	2	1	1
48	Marszalek	1999	3	2	2	2	2	1	1	1	2	3	2
49	Martinez	2011	1	3	2	2	2	3	1	1	1	1	1
50	Merchant	2013	2	3	3	2	1	3	1	2	2	1	1
51	Montgomery	2008	3	3	1	1	1	3	3	2	2	2	2
52	Myneni	2013	1	3	1	2	2	1	1	2	2	1	1
53	Olympiou	2012	1	3	3	2	2	3	1	1	2	1	1
54	Perry	2008	3	3	4	1	2	1	2	2	2	2	1
55	Predavec	2001	3	3	1	2	2	3	1	2	1	2	1
56	Pyatt	2007	2	2	1	2	2	3	1	1	1	2	2
57	Pyatt	2012	2	2	1	2	2	3	1	2	1	1	1
58	Scoville	2007	3	3	1	2	2	3	2	1	1	2	1
59	Shegog	2012	3	2	1	3	2	3	1	1	2	1	1
60	Sun	2008	5	1	3	2	2	2	1	1	2	2	1
61	Sun	2010	4	1	2	2	2	1	2	2	2	1	1
62	Taghavi	2009	1	3	3	2	2	1	3	3	2	2	1
63	Tan	2007	3	1	1	1	2	1	1	1	1	2	2
64	Tarekegn	2009	1	3	1	2	2	2	1	1	2	2	1
65	Tarng	2012	3	1	2	3	1	2	1	2	2	1	1
66	Tatli	2013	2	2	3	2	2	3	1	1	2	1	1
67	Trundle	2010	4	3	3	2	2	2	2	1	2	1	1
68	Tsai	2008	3	3	2	2	2	3	1	1	2	2	1
69	Tüysüz	2010	2	2	3	1	2	2	2	3	2	1	1
70	Ulen	2014	1	2	2	1	2	1	1	2	2	1	1
71	Ünlü	2011	1	1	2	2	2	1	1	1	2	1	1
72	Virvou	2005	4	1	1	3	1	2	1	2	2	2	1
73	White	1984	1	3	4	3	1	1	1	2	2	3	1
74	Wong	2007	3	3	1	1	1	2	1	1	2	2	2
75	Wrzesien	2010	3	1	1	1	1	3	1	1	2	1	1
76	Yang	2007	1	2	3	3	2	2	1	2	2	2	1
77	Zacharia	2007	1	3	3	2	2	3	1	1	2	2	1
78	Zacharia	2008	1	3	3	2	2	2	1	1	2	2	1
79	Zacharia	2011	1	3	3	2	2	2	1	3	2	1	1

Table 4. *Main effects for declarative knowledge learning, knowledge retention and procedural knowledge learning comparing serious games with more conventional instruction*

Learning outcome	***d***	**95% CI**	***k***	***N***	***QT***
Declarative knowledge	0.34*	[0.25, 0.43]	65	7 354	296.17*
Knowledge retention	0.31*	[0.10, 0.52]	8	1 025	5.28
Procedural knowledge	0.41*	[0.11, 0.71]	7	556	7.12

* Significant at the $p < .05$ threshold.

Table 5. *Resume of moderator analyses*

Moderator	***F***	***p***	d_1	d_2	d_3
Subject area	0.179	.911	0.38	0.24	0.33
Grade level	6.064	.004*	0.33	0.73	0.20
Duration of intervention	4.227	.022*	0,64	0.52	0.39
Activity level of the comparison group	2.883	.064	0.43	0.19	
Ludic content	0.249	.781	0.33	0.36	
Level of realism	*1.854*	*.179*	0.39	0,34	0.26
Level of user control	3.582	.012*	0.45	0.21	
Randomization	1.741	.194	0.28	0.36	
Experimental design	0.808	.608	0.35	0.32	
Year of publication	6.993	.002*	0.42	0.16	0.16
Publishing status	3.198	.038*	0.36	0.11	

* *S*ignificant at the $p < .05$ threshold.