Computers & Education 60 (2013) 412–425

Contents lists available at SciVerse ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

A meta-analytic review of the role of instructional support in game-based learning

Pieter Wouters*, Herre van Oostendorp

Institute of Information and Computing Sciences, Utrecht University, P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 2 April 2012
Received in revised form
26 June 2012
Accepted 13 July 2012

Keywords:
Computer games
Serious games
Game-based learning
Cognition
Meta-analysis
Instructional support

ABSTRACT

Computer games can be considered complex learning environments in which players require instructional support to engage in cognitive processes such as selecting and actively organizing/integrating new information. We used meta-analytical techniques to test if instructional support enhances learning in game-based learning ($k = 107$, $N_{adj} = 3675$). We found that instructional support in game-based learning environments improved learning ($d = .34$, $p < .001$). Additional moderator analyses revealed that the learning effect was largest when learning of skills was involved ($d = .62$, $p < .001$) and when the instructional support aimed at the selection of relevant new information ($d = .46$, $p < .001$). Furthermore, we found some evidence for a publication bias since the effect sizes for studies in peer-reviewed journals was significantly higher than for studies in proceedings and unpublished studies (journals: $d = .44$; proceedings: $d = .08$; unpublished: $d = .14$).

1. Introduction

The last decade shows an increasing attention for the use of computer games in learning and instruction, often referred to as serious games or game-based learning (GBL). In this respect serious games are hypothesized to address both the cognitive and affective dimensions of learning (O'Neil, Wainess, & Baker, 2005), enable learners to adapt learning to their cognitive needs and interest and provide motivation for learning (Malone, 1981). Some recent reviews have indeed shown that GBL can be more effective than conventional instruction such as lectures or classroom instruction (e.g, Sitzmann, 2011; Wouters, Van Nimwegen, van Oostendorp, & van der Spek, submitted for publication).

From a cognitive theory perspective it can be argued that GBL often involves complex learning environments in which players – especially novices – can easily become overwhelmed by all the information that has to be processed and consequently refrain from activities that foster learning (Wouters, van der Spek, & van Oostendorp, 2008). The question can be raised whether learners in GBL environments require support to engage in relevant cognitive activities. Such instructional support may appear in several forms such as providing feedback, scaffolding, giving advice et cetera. However, little is known regarding the effect of instructional support in GBL. Earlier reviews and meta-analyses on GBL did not consider instructional support, but focused on the comparison of GBL with traditional instruction methods (e.g., O'Neil et al., 2005; Sitzmann, 2011; Vogel, Vogel, Cannon-Bowers, Muse, & Wright, 2006; Wouters et al., submitted for publication).

The goal of this study is to systematically investigate the role of instructional support in GBL. Mayer (2011) has divided game research into three categories: a value-added approach with the question how specific game features foster learning and motivation; a cognitive consequences approach which investigates what people learn from serious games and a media comparison approach which investigates whether people learn better from serious games than from conventional media. Our meta-analysis will adopt the value-added approach by comparing studies with and without instructional support. Instructional support includes a broad range of techniques and methods that aim at different cognitive activities. Therefore, we will not only consider the main effect of instructional support, but also investigate the effects of several types of instructional support.

In the following sections we will first define GBL. Next, we will describe the theoretical framework we use to investigate the different types of instructional support. The Method section comprises a description of the literature research, the inclusion criteria, the coding of the

* Corresponding author. Tel.: +31 30 253 9219.
E-mail addresses: P.J.M.Wouters@uu.nl, pieterw@cs.uu.nl (P. Wouters).

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2012.07.018

moderator variables and the calculation of effect sizes. The Results section presents the general characteristics of the analysis, the main effects and the effects of the moderator variables. Finally, we will discuss the findings, draw conclusions and depict some avenues for future research.

2. Definition of GBL

Several scholars have provided definitions or classifications of computer games characteristics (Garris, Ahlers, & Driskell, 2002; Malone, 1981; Prensky, 2001). For the purpose of this meta-analysis we describe computer games in terms of being *interactive* (Prensky, 2001; Vogel et al., 2006), based on a set of *agreed rules and constraints* (Garris et al., 2002), and directed toward a clear *goal* that is often set by a *challenge* (Malone, 1981). In addition, games constantly provide *feedback* either as a score or as changes in the game world to enable players to monitor their progress toward the goal (Prensky, 2001). Some scholars contend that computer games also involve a competitive activity (against the computer, another player, or oneself), but it can be questioned if this is a defining characteristic. Of course there are many games in which the player is in competition with another player or with the computer, but in a game like SimCity players may actually enjoy the creation of a prosperous city that satisfies their beliefs or ideas without having the notion that they engage in a competitive activity. In the same vein a narrative or the development of a story can be very important in a computer game (e.g., in adventure games), but again it is not a prerequisite for being a computer game (e.g., action games do not really require a narrative). This definition of GBL would also comprise 'pure simulations' such as SIMQuest (see also www.simquest.nl) that also include an underlying model in which learners can provide input (either changing variables or perform actions) and observe the consequences of their actions (cf. Leemkuil, de Jong, & Ootes, 2000). However, we concur with Jacobs and Dempsey (1993) who argued that task-irrelevant elements are often removed from simulations whereas other elements such as an engaging context are included or emphasized to define a (simulation) game. In GBL the objective of the computer game is not to entertain the player, which would be an added value, but to use the entertaining quality for training, education, health, public policy, and strategic communication objectives (Zyda, 2005).

3. Theoretical framework

From a cognitive perspective instructional support can be implemented in order to overcome the limitations of the human cognitive architecture (Mayer, 2001, 2008; Paas, Renkl, & Sweller, 2003; van Oostendorp, Beijersbergen, & Solaimani, 2008). Two structures are regarded crucial for the processing of information. The first, working memory, has a limited capacity to process information and is often not adequate for learning material when it is complex, multimodal and/or dynamical. Especially for novices the complexity and the dynamic character of instructional material may lead to problems: they do not know what is relevant and therefore focus on the wrong information. The second structure, long term memory, has a virtually unlimited capacity which can serve as added processing capacity by means of schemas, i.e., cognitive structures that can be processed in working memory as a single entity (Kintsch, 1998; Paas et al., 2003).

Based on this cognitive architecture, theories have emphasized several important cognitive processes that are involved in learning. Mayer's cognitive theory of multimedia learning, for example, discerns three types of cognitive processing: the *selection* of relevant information by paying attention to the presented material, mentally *organizing* the new information in a coherent structure and the *integration* of this structure with prior knowledge (Mayer, 2001).

Although the organization and integration of information reflect different cognitive processes, they are closely related and difficult to separate (cf. Moreno & Mayer, 2005). Therefore, we propose that, from a cognitive perspective, instructional support should enable learners to engage in two types of learning processes: (1) the selection of relevant information from the learning material and (2) the active organization of that information and the integration with prior knowledge in long term memory (Mayer & Moreno, 2003).

GBL involves complex learning environments in which it is not obvious that players automatically engage in these basic learning processes. To start with, players can be easily overwhelmed by the plentitude of information, the multimodal presentation of information (sometimes simultaneously on different locations of the screen), the choices players potentially can make, the dynamics of the game and the complexity of the task that has to be performed. This implies that instructional guidance to support players to adequately *select* relevant information and ignore irrelevant information is important, certainly given working memory constraints. Secondly, in computer games players act and see the outcome of their actions reflected in changes in the game world. This may lead to a kind of intuitive learning: players know how to apply knowledge, but they cannot explicate it (Leemkuil & de Jong, 2011). Yet, it is important that learners articulate and explain their knowledge, because it urges them to *organize* new information and *integrate* it with their prior knowledge. Ultimately, this will yield a knowledge base with higher accessibility, better recall and higher transfer of learning (Wouters, Paas, & van Merriënboer, 2008).

4. Method

4.1. Sample and literature search

We started with computer-based searches via GoogleScholar. The search terms we used were: 'Game-based learning', 'serious games', 'educational games', 'simulation games', 'virtual environments', and 'muve'. When necessary these search terms were combined with 'learning', 'instruction', and 'training'. In addition, we investigated the references of previous meta-analyses and reviews on the effectiveness of serious games (Fletcher & Tobias, 2006; Ke, 2009; O'Neil et al., 2005; Sitzmann, 2011; Vogel et al., 2006; Wouters et al., submitted for publication). In order to find unpublished, but relevant studies we also asked researchers and educators within our network of scholars whether they were aware of relevant studies for the meta-analysis. Our meta-analysis covered the period from 1990 to 2012. For the selection process we considered only studies in which GBL complied with our definition. In addition, studies were only selected when GBL with instructional support was compared with GBL without instructional support in a controlled experimental design. Furthermore, the study had to report data or indications that allowed us to calculate or estimate effect sizes (group means and standard deviations, *t*-test, *F* test et cetera). Third, the focus of the study comprised nondisabled participants. Our research located 197 studies of which 29 studies met these inclusion criteria.

4.2. Variables coded from each study

4.2.1. Learning outcome

The effectiveness of instructional support may be contingent on the type of learning that is involved. Several classifications have been proposed for learning outcomes (see Kraiger, Ford, & Salas, 1993; Wouters, van der Spek, & van Oostendorp, 2009). In this meta-analysis we will focus on the cognitive dimension of learning. In the Wouters et al. (2009) classification this dimension is divided into knowledge and cognitive skills. Knowledge refers to encoded knowledge reflecting both text-oriented (e.g., verbal knowledge) and non text-oriented knowledge (e.g., knowledge derived from an image). A cognitive skill pertains to more complex cognitive processes, such as in problem solving when a learner applies factual knowledge and rules to achieve a solution for a (novel) situation. In-game performance (i.e., the performance of players on a task during the game) is regarded as a third type of learning outcome, because it may reflect intuitive learning.

In the studies knowledge was explicitly mentioned as a learning outcome (e.g., Leemkuil, 2006; van der Spek, 2011), called retention (e.g., Moreno & Mayer, 2004) or described as recall (Wouters, van Oostendorp, Boonekamp, & van der Spek, 2011). The most occurring measurement of cognitive skills was transfer (e.g., Barab et al., 2009; Lang & O'Neil, 2008). In the study by Ritterfeld, Shen, Wang, Nocera, and Wong (2009) we used the score on the essay task as a measurement for skills. In-game performance was generally described as an in-game score in the game environment (e.g., in van der Spek (2011) the game score was based on the correct application of a procedure), solution time to solve a problem in the game (Mayer, Mautone, & Prothero, 2002) or the number of problems solved in the game (Van Eck & Dempsey, 2002). From the Leutner (1993) study the scores on the functional knowledge measurement were used.

4.2.2. Type of instructional support

As mentioned before, there is a large variety in instructional support. First we classified instructional support according to the cognitive activities they targeted at. When a type of instructional support (e.g., the availability of background information) was difficult to classify into one of the two categories, the label 'Unknown' was used. Table 1 gives an overview of the different types of instructional support that were mentioned in the studies and how we classified them.

In order to have a more comprehensive understanding we also classified them in 10 groups of instructional support that are regularly used in GBL: reflection, modeling, advice, collaboration, control, narrative elements, modality, feedback, personalization, and others. Table 2 explains the 10 groups of instructional support and how we categorized each type of instructional support into one of these 10 groups. We always checked whether the classification used in the study to describe instructional support was in line with our classification. For example, Leemkuil (2006) uses a form of instructional support which is described as alternative advice, the nature of this support, however, is to foster reflection (by posing questions). Therefore, it was classified under Reflection.

4.2.3. Instructional domain

The studies in our research covered a broad range of domains. The domains Biology, Mathematics, and Economy were coded as such. Because of the low number of comparisons we have coded geology (6 comparisons from only 1 study), electronic circuits/electricity (4 comparisons from 2 studies) and science (1 comparison) as 'Others'. The broad domain General problem solving was defined for studies in which problem solving in a non-specific domain was involved. For example, in some studies the game Space Fortress was used in which visual attention, decision making and psychomotor skills are required to fulfill a complex task (e.g., Day et al., 2007).

4.2.4. Age

The following categories were used: 'Elementary school (children until 11–12 years)', 'Middle/High school (preparatory education from 13 to 17 years)', 'College/University (18–24 years)' and 'Adults (older than 24 years)'.

Table 1
The types of instructional support and their classification in type of cognitive process.

Type of support	Description	Cognitive process
Advice	System-generated hints and suggestions to focus attention	Selection
Adaptivity	Adapt the game to the player's performance	Selection
Assignment	Prompts to investigate relations between variables	Organization/integration
Background info	Retrievable domain-specific knowledge	Unknown
Cues	Signals to cue the attention of the player	Selection
Collaboration	Discussion often aiming at the explication of implicit knowledge	Organization/integration
Choice	Control over irrelevant aspect of learning activities	Unknown
Contextualization	Presenting learning in a meaningful context	Organization/integration
Elaboration	Additional task-related cognitive activities	Organization/integration
Feedback/guidance	Information whether an/or why an answer is correct	Selection
Foreshadowing	Events occurring later are shown briefly ahead to focus attention	Selection
Interactivity	The game is responsive to the player's actions	Organization/integration
JIT options	Options are shown when they are needed to solve a problem	Selection
Modality	The use of the audio channel to limit visual search	Selection
Modeling	Describing how a problem is solved	Selection
Narrative	Storyline that can help players to organize educational material	Organization/integration
Pedagogical agent	An agent gives hints to help the student to learn from the game	Unknown
Personalization	Adapt context to personal interest of the player	Unknown
Pretraining	The player receives domain-specific knowledge in advance	Selection
Process/product goal	Players have a specific goal in order to focus on specific features	Selection
Reflection	Stimulation to think about answers and/or explain these answers	Organization/integration
Surprising events	Unexpected events that trigger an update of the mental model	Organization/integration
Variability	Presenting problems that vary in task features	Organization/integration
Worked examples	Showing how a problem can be solved	Selection

Table 2
A description of the groups of instructional support and how they are related to the types of instructional support.

Groups of instructional support	Specific types of instructional support
1. Reflection Learners are stimulated to think about their answers and (sometimes) explain it to themselves	Reflection, self-explanation, elaboration, assignments
2. Modeling An explication or indication how to solve a problem. The explanation can be given by a peer or expert and can be verbal, animated or graphical.	Different types of scaffolding, modeling, worked examples
3. Advice System-generated information to support the learner to continue in the game (e.g. by focusing attention).	All types of advice whether contextualized, adaptive or not
4. Collaboration Working in groups with discussion, often aiming at the explication of implicit knowledge	Players played in dyads, groups or engaged in group discussion
5. Interactivity Learners make choices in the game in order to solve a problem or to perform a task	Interactivity, learner control and choice of game features
6. Narrative elements A narrative context or manipulation of narrative elements which provides a cognitive framework	Fantasy, rich narrative, foreshadowing, surprising events
7. Modality Textual explanation is provided auditory rather than visually (written)	Modality
8. Feedback Information is given whether an answer or action is correct or not. The feedback can be corrective (correct or not), explanatory (why correct or not).	Feedback, guidance
9. Personalization Ideas, characters, topics and messages are presented in such a way that they have a specific high interest value for the learner/player	Personalization, personalized messages
10. Others A group of instructional support that is difficult to categorize because there are too few comparison	Goal direction, pretraining, background information, cues, adaptivity

4.2.5. Level of realism

Vogel et al. (2006) have divided level of realism into photo-realistic, high-quality cartoonlike pictures, low-quality programmed pictures and unknown. We adapted this taxonomy because we found a number of studies with only a schematic design consisting of lines, number and symbols (see for example Cordova & Lepper, 1996). Fig. 1 provides an illustration of the different categories of levels of realism. We combined low-quality programmed pictures and schematic into the group 'Basic/Schematic' (examples are Cameron & Dwyer, 2005; Cordova & Lepper, 1996). Cartoonlike games were coded as 'Cartoon' (e.g., Blumberg, 2000), likewise (photo)realistic games were coded as 'Realistic' (e.g., McQuiggan, Rowe, Lee, & Lester, 2008). 'Unknown' was used when the level of realism of a game could not be determined.

4.2.5.1. Methodological variables. For each study three methodological variables were coded. To start with, the publication source could be either 'peer-reviewed journal', 'proceedings' and 'unpublished' (e.g., master and doctoral theses, conference papers). Secondly, it was coded whether participants were assigned randomly to the conditions or not. Thirdly, the experimental design of the study – either posttest only or pretest–posttest – was coded.

4.3. Calculation and analysis of the effect sizes

Cohen's *d* was used as indicator for the effect size: the difference on the dependent variable learning between GBL with and without instructional support was calculated and divided by the pooled standard deviation. When the means and/or standard deviations were not available appropriate formulas were used to estimate the effect size based on *t*-test, univariate *F* test (Glass, McGraw, & Smith, 1981), adjusted means, or gain scores (Hedges & Olkin, 1985). Table 3 gives the formulas that were used to calculate the effect sizes:

Effect sizes for studies with small sample bias were corrected (cf. Hedges & Olkin, 1985). When a study reported multiple measurements for a learning outcome we calculated an averaged effect size for these measurements (e.g., the proximal, distal and transfer measurements in the Barab et al. (2009) study were averaged and classified as a skill). When multiple learning outcomes and/or multiple treatment or comparisons groups were used, each pairwise combination of a learning outcome and/or treatment or comparison group was treated as an independent study. To avoid the overrepresentation of studies with multiple pairwise comparisons, the sample size was adjusted to the number of pairwise comparisons. For this purpose we used a procedure to assure that no comparison received an inappropriate weight (see Appendix A for a description of the procedure and an example).

Given the great variation in types of instructional support it is likely that the average effect sizes in the populations varies between the studies. Therefore we used the random-effects model for the main analyses and the moderator analyses with 95% confidence intervals

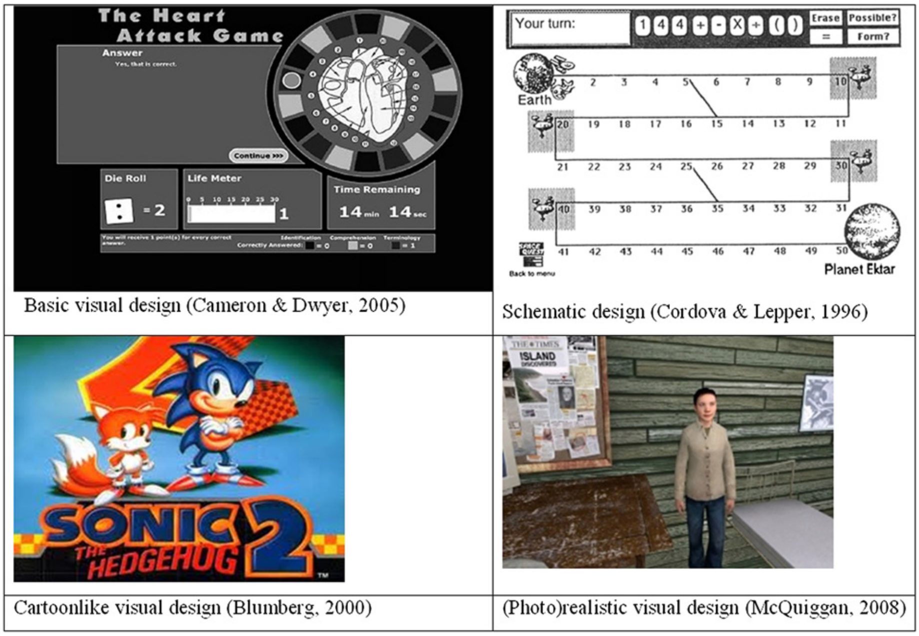

Fig. 1. Examples of studies to illustrate the different categories of levels of realism.

around the weighted mean effect sizes. For all analyses we created a Visual Basic for Applications-program in Excel using the formulas provided by Ellis (2010) and Borenstein, Hedges, Higgins, and Rothstein (2009).

5. Results

We found 29 studies with 107 pairwise comparisons involving 3675 participants. The sample sizes of the studies ranged from 24 to 267 participants. Table 4 presents all included pairwise comparisons with effect sizes and their classification on the non-methodological moderator variables.

The heterogeneity of effect sizes was confirmed by the results of the homogeneity test for learning outcome ($Q_{total} = 175.62$ df $= 106$, $p < .0001$). Therefore, a moderator analysis is justified for learning. For all analyses we set alpha at .05.

5.1. Main effect analysis

We calculated the fail-safe N which is the number of studies averaging null results that has to be retrieved in order to reject the weighted mean effect size. According to Rosenthal (1995) a publication bias is unlikely to occur when the fail-safe N (in our case 12,780) exceeds the suggested threshold of the quintuple of pairwise comparisons plus 10 (Ellis, 2010) which is clearly the case in this review: $12{,}780 > 5^*107 + 10 = 545$.

The weighted mean effect size for instructional support was $d = .34$ ($z = 7.26$, $p < .001$) indicating that learners in GBL environments with instructional support (i.e., generating self-explanation, advice, scaffolding et cetera) outperform those who engage in GBL environments without this instructional support.

5.2. Moderator analysis

The results of the moderator analysis are shown in Table 5. In order to compare the subgroups in the moderator analysis we adopted the z-testing method for random effects with separate estimates of between-study variance (see Borenstein et al., 2009). When a moderator variable comprised more than two categories the Holm-Bonferroni procedure was used to adjust the critical p-value to control for the Type 1 error (cf. Ginns, 2005). In Holm's sequential version, the results of the Bonferroni tests are ordered from the smallest p-value to the one with the highest p-value. The test with the lowest p-value is then tested first with a Bonferroni correction involving all tests. The second test applies a Bonferroni correction involving one test less. This procedure continues for the remaining tests (Abdi, 2010).

Table 3
Formulas to calculate effect sizes.

Data	Formula
M, SD given	Cohen's d
T test with df	$d = (2^*t)/\sqrt{\text{df}}$
F test with df error	$d = (2^*\sqrt{F})/\sqrt{\text{df(error)}}$
Adjusted means with SD and n	Cohen's d
Gain scores with SD and n	Cohen's d
Mean difference with SE and n	$T = M_{diff}/$ SE, then $d = t^*\sqrt{(n_e + n_c)/(n_e^* n_c)}$

Table 4
Studies and pairwise comparisons of GBL with and without instructional support.

Study		Total *N*	Adj. *N*	*d*	Learning outcome	Comparison	Cognitive process	Category of instructional support	Domain	Age	Level of realism
Adams et al. (2012)	Exp. 2	114	114	.19	Mixed	Narrative	Org./integr.	Narrative	Science	Col./univ.	Real.
Arthur, Day, Bennet, McNelly, and Jordan (1997)		89	89	−.12	In-game perf.	Dyads	Org./integr.	Collaboration	PS	Adults	Schema
Barab et al. (2009)[a]		24	24	.34	Skills	Dyads	Org./integr.	Collaboration	Biol.	Col./univ.	Real.
Blumberg (2000)											
	Grade 2	45	45	.10	In-game perf.	Goal direction	Selection	Others	PS	Elem. school	Cartoon
	Grade 5	47	47	.43	In-game perf.	Goal direction	Selection	Others	PS	Elem. school	Cartoon
Cameron and Dwyer (2005)[b]											
		212	160	.29	Knowledge	Resp. feedback	Selection	Feedback	Biol.	Col./univ.	Schema
		208	156	.57	Knowledge	Elab. feedback	Selection	Feedback	Biol.	Col./univ.	Schema
Conati and Zhao (2004)		20	20	.70	Skills	Pedagogical agent	Unknown	Others	Math	H/M school	Schema
Cordova and Lepper (1996)[c]											
		56	9.3	.49	Skills	Choice	Unknown	Control	Math	Elem. school	Schema
		56	9.3	.26	In-game perf.	Choice	Unknown	Control	Math	Elem. school	Schema
		70	16.3	1.23	Skills	Fantasy	Org./integr.	Narrative	Math	Elem. school	Schema
		70	16.3	.07	In-game perf.	Fantasy	Org./integr.	Narrative	Math	Elem. school	Schema
		56	9.3	1.20	Skills	Personalization	Unknown	Personalization	Math	Elem. school	Schema
		56	9.3	.38	In-game perf.	Personalization	Unknown	Personalization	Math	Elem. school	Schema
Day et al. (2007)[d]		47	47	.14	In-game perf.	Collaboration	Org./integr.	Collaboration	PS	Adults	Schema
Inkpen, Booth, Klawe, and Upitis (1995)[e]											
	Study 1	56	44	.25	In-game perf.	Dyads	Org./integr.	Collaboration	PS	Elem. school	Schema
	Study 1	30	18	.15	In-game perf.	Dyads	Org./integr.	Collaboration	PS	Elem. school	Schema
	Study 1	40	28	−.21	In-game perf.	Dyads	Org./integr.	Collaboration	PS	Elem. school	Schema
	Study 1	26	14	−.21	In-game perf.	Dyads	Org./integr.	Collaboration	PS	Elem. school	Schema
	Study 2	265	164.5	.01	In-game perf.	Dyads	Org./integr.	Collaboration	PS	Elem. school	Schema
	Study 2	267	166.5	.39	In-game perf.	Dyads	Org./integr.	Collaboration	PS	Elem. school	Schema
Johnson and Mayer (2010)											
	Exp. 1	23	23	1.10	Skills	Selection reflection	Org./integr.	Reflection	Science	Col./univ.	Schema
	Exp. 2	54	41	.65	Skills	Self reflection	Org./integr.	Reflection	Science	Col./univ.	Schema
	Exp. 2	53	40	−.07	Skills	Selection reflection	Org./integr.	Reflection	Science	Col./univ.	Schema
Lang and O'Neil (2008)											
		78	59	.56	Skills	JIT worked example	Selection	Modeling	PS	Col./univ.	Unknown
		78	58.5	.06	Skills	Worked example	Selection	Modeling	PS	Col./univ.	Unknown
Leutner (1993)[f]											
	Exp.1 No adv.	32	12	.29	Knowledge	Pretraining	Selection	Others	PS	H/M school	Schema
	Exp.1 No adv.	32	12	−.69	In-game perf.	Pretraining	Selection	Others	PS	H/M school	Schema
	Exp.1 No prt.	32	12	.61	Knowledge	Adaptive advice	Selection	Advice	PS	H/M school	Schema
	Exp.1 No pretrain.	32	12	−.96	In-game perf.	Adaptive advice	Selection	Advice	PS	H/M school	Schema
	Exp. 2	38	19	.83	Knowledge	Adaptive advice	Selection	Advice	PS	Col./univ.	Schema
	Exp. 2	38	19	−.53	In-game perf.	Adaptive advice	Selection	Advice	PS	Col./univ.	Schema
	Exp. 3 No info	40	15	.23	Knowledge	Adaptive advice	Selection	Advice	PS	H/M school	Schema
	Exp. 3 No info	40	15	.18	In-game perf.	Adaptive advice	Selection	Advice	PS	H/M school	Schema
	Exp. 3 No adv.	40	15	.19	Knowledge	Background info	Unknown	Others	PS	H/M school	Schema
	Exp. 3 No adv.	40	15	.46	In-game perf.	Background info	Unknown	Others	PS	H/M school	Schema
Leemkuil (2006)[g]											
	Chap. 5	28	9.3	−.15	Knowledge	Added assignment	Org./integr.	Reflection	Econ.	Col./univ.	Cartoon
	Chap. 5	28	9.3	.34	Skills	Added assignment	Org./integr.	Reflection	Econ.	Col./univ.	Cartoon
	Chap. 5	28	9.3	−.30	In-game perf.	Added assignment	Org./integr.	Reflection	Econ.	Col./univ.	Cartoon
	Chap. 6	29	9.7	−.70	Knowledge	Advice	Selection	Advice	Econ.	Col./univ.	Cartoon
	Chap. 6	29	9.7	.14	Skills	Advice	Selection	Advice	Econ.	Col./univ.	Cartoon
	Chap. 6	29	9.7	−.11	In-game perf.	Advice	Selection	Advice	Econ.	Col./univ.	Cartoon
	Chap. 7	181	65.5	.24	Knowledge	Alternative advice	Org./integr.	Reflection	Econ.	Col./univ.	Cartoon
	Chap. 7	181	65.5	.18	In-game perf.	Alternative advice	Org./integr.	Reflection	Econ.	Col./univ.	Cartoon
	Chap. 7	185	67.5	.11	Knowledge	Advice	Selection	Advice	Econ.	Col./univ.	Cartoon

(*continued on next page*)

Table 4 (*continued*)

Study		Total *N*	Adj. *N*	*d*	Learning outcome	Comparison	Cognitive process	Category of instructional support	Domain	Age	Level of realism
	Chap. 7	185	67.5	.25	In-game perf.	Advice	Selection	Advice	Econ.	Col./univ.	Cartoon
Mayer et al. (2002)											
	Exp. 1	28	28	–.18	In-game perf.	Modeling	Selection	Modeling	Science	Col./univ.	Unknown
	Exp. 2	53	37.3	.32	In-game perf.	Strategic modeling	Selection	Modeling	Science	Col./univ.	Unknown
	Exp. 2	48	36.3	.56	In-game perf.	Pictorial modeling	Selection	Modeling	Science	Col./univ.	Unknown
	Exp. 2	38	31.3	.69	In-game perf.	Str.& pic. modeling	Selection	Modeling	Science	Col./univ.	Unknown
	Exp. 3	73	36.5	.99	Skills	Pictorial modeling	Selection	Modeling	Science	Col./univ.	Unknown
	Exp. 3	73	36.5	.84	In-game perf.	Pictorial modeling	Selection	Modeling	Science	Col./univ.	Unknown
McQuiggan et al. (2008)		116	116	–.32	Skills	Narrative	Org./integr.	Narrative	Biol.	Col./univ.	Real.
Meluso, Zheng, Spires, and Lester (2012)		100	100	.00	Knowledge	Collaboration	Org./integr.	Collaboration	Biol.	Elem. school	Real.
Moreno and Mayer (2002)											
	Exp. 1	89	44.5	.83	Knowledge	Modality	Selection	Modality	Biol.	Col./univ.	Schema
	Exp. 1	89	44.5	1.31	Skills	Modality	Selection	Modality	Biol.	Col./univ.	Schema
	Exp. 2 desktop	28	14	1.52	Knowledge	Modality	Selection	Modality	Biol.	Col./univ.	Schema
	Exp. 2 desktop	28	14	1.75	Skills	Modality	Selection	Modality	Biol.	Col./univ.	Schema
	Exp. 2 HMD	28	14	2.62	Knowledge	Modality	Selection	Modality	Biol.	Col./univ.	Schema
	Exp. 2 HMD	28	14	2.48	Skills	Modality	Selection	Modality	Biol.	Col./univ.	Schema
Moreno and Mayer (2004)											
		48	24	.76	Knowledge	Pers. messages	Unknown	Personalization	Biol.	Col./univ.	Schema
		48	24	1.61	Skills	Pers. messages	Unknown	Personalization	Biol.	Col./univ.	Schema
Moreno and Mayer (2005)[f]											
	Exp. 1	105	26.3	.43	Knowledge	Guidance	Selection	Feedback	Biol.	Col./univ.	Schema
	Exp. 1	105	26.3	1.10	Skills	Guidance	Selection	Feedback	Biol.	Col./univ.	Schema
	Exp. 1	105	26.3	–.08	Knowledge	Reflection	Org./integr.	Reflection	Biol.	Col./univ.	Schema
	Exp. 1	105	26.3	.13	Skills	Reflection	Org./integr.	Reflection	Biol.	Col./univ.	Schema
	Exp. 2	71	17.8	.16	Knowledge	Interactivity	Org./integr.	Control	Biol.	Col./univ.	Schema
	Exp. 2	71	17.8	–.70	Skills	Interactivity	Org./integr.	Control	Biol.	Col./univ.	Schema
	Exp. 2	71	17.8	.08	Knowledge	Reflection	Org./integr.	Reflection	Biol.	Col./univ.	Schema
	Exp. 2	71	17.8	.73	Skills	Reflection	Org./integr.	Reflection	Biol.	Col./univ.	Schema
	Exp. 3	78	23.4	.45	Knowledge	Reflection	Org./integr.	Reflection	Biol.	Col./univ.	Schema
	Exp. 3	78	23.4	.39	Skills	Reflection	Org./integr.	Reflection	Biol.	Col./univ.	Schema
	Exp. 3 w/ PR	31	7.8	–.03	Knowledge	Interactivity	Org./integr.	Control	Biol.	Col./univ.	Schema
	Exp. 3 w/ PR	31	7.8	.13	Skills	Interactivity	Org./integr.	Control	Biol.	Col./univ.	Schema
	Exp. 3 w/o R	31	7.8	.11	Knowledge	Interactivity	Org./integr.	Control	Biol.	Col./univ.	Schema
	Exp. 3 w/o R	31	7.8	.34	Skills	Interactivity	Org./integr.	Control	Biol.	Col./univ.	Schema
Moreno, Mayer, Spires, and Lester (2001)											
	Exp. 3	38	19	.82	Knowledge	Participation	Org./integr.	Control	Biol.	Col./univ.	Schema
	Exp. 3	38	19	.74	Skills	Participation	Org./integr.	Control	Biol.	Col./univ.	Schema
	Exp. 4	64	32	.79	Knowledge	Modality	Selection	Modality	Biol.	Col./univ.	Schema
	Exp. 4	64	32	1.04	Skills	Modality	Selection	Modality	Biol.	Col./univ.	Schema
	Exp. 5	79	39.5	.67	Knowledge	Modality	Selection	Modality	Biol.	Col./univ.	Schema
	Exp. 5	79	39.5	1.56	Skills	Modality	Selection	Modality	Biol.	Col./univ.	Schema
Prislin, Jordan, Worchel, Semmer, and Shebilske (1996)[h]		127	127	.14	In-game perf.	Group discussion	Org./integr.	Collaboration	PS	Adults	Schema
Ritterfeld et al. (2009)											
		50	25	.03	Knowledge	Interactivity	Org./integr.	Control	Biol.	Col./univ.	Real.
		50	25	–.05	Skills	Interactivity	Org./integr.	Control	Biol.	Col./univ.	Real.
Sandberg, Wielinga, and Christoph (2012)											
		46	23	.13	Knowledge	Modeling	Selection	Modeling	Econ.	Col./univ.	Cartoon
		46	23	.81	Skills	Modeling	Selection	Modeling	Econ.	Col./univ.	Cartoon
Shebilske, Regian, Arthur, and Joran (1992)[h]		40	20	.69	In-game perf.	Dyads	Org./integr.	Collaboration	PS	Adults	Schema

Table 4 (*continued*)

Study		Total *N*	Adj. *N*	*d*	Learning outcome	Comparison	Cognitive process	Category of instructional support	Domain	Age	Level of realism
Shebilske, Goettl, Corrington, and Day (1999)[h]		40	40	.62	In-game perf.	Task-related elaboration	Org./integr.	Reflection	PS	Adults	Schema
Van der Meij et al. (2011)		45	45	.25	Knowledge	Dyads	Org./integr.	Collaboration	Econ.	Col./univ.	Cartoon
Van Eck and Dempsey (2002)											
	Comp.	35	25.5	−.35	Skills	Context. advice	Selection	Advice	Math	H/M school	Schema
	Comp.	51	17.5	−.10	In-game perf.	Context. advice	Selection	Advice	Math	H/M school	Schema
	No comp.	29	24.0	.79	Skills	Context. advice	Selection	Advice	Math	H/M school	Schema
	No comp.	48	14.5	.54	In-game perf.	Context. advice	Selection	Advice	Math	H/M school	Schema
van der Spek (2011)											
	Chap. 5	27	6.7	.02	Knowledge	Visual cues	Selection	Others	PS	Col./univ.	Real.
	Chap. 5	27	6.7	.35	In-game perf.	Visual cues	Selection	Others	PS	Col./univ.	Real.
	Chap. 5	27	7	−.84	Knowledge	Audio cues	Selection	Others	PS	Col./univ.	Real.
	Chap. 5	27	6.8	.02	In-game perf.	Audio cues	Selection	Others	PS	Col./univ.	Real.
	Chap. 6a	56	9.3	.05	Knowledge	Problem complexity	Selection	Others	PS	Col./univ.	Real.
	Chap. 6a	56	9.3	−.09	In-game perf.	Problem complexity	Selection	Others	PS	Col./univ.	Real.
	Chap. 6a	56	9.3	.40	Knowledge	Option complexity	Selection	Others	PS	Col./univ.	Real.
	Chap. 6a	56	9.3	.66	In-game perf.	Option complexity	Selection	Others	PS	Col./univ.	Real.
	Chap. 6b	28	14	−.04	Knowledge	Adaptivity	Selection	Others	PS	Col./univ.	Real.
	Chap. 7	41	13.7	.05	Knowledge	Surprising events	Org./integr.	Narrative	PS	Col./univ.	Real.
	Chap. 7	41	13.7	.16	In-game perf.	Surprising events	Org./integr.	Narrative	PS	Col./univ.	Real.
Wouters et al. (2011)		29	29	.26	Knowledge	Foreshadowing	Org./integr.	Narrative	Biol.	Col./univ.	Real.
Worchel, Shebilske, Jordan, and Prislin (1997)		51	51	.00	In-game perf.	Collaboration	Org./integr.	Collaboration	PS	Adults	Schema

[a] A mean of distal, proximal and open-ended transfer task.
[b] Composite score.
[c] Several measure for game performance, '% problems solve correctly' was selected.
[d] Better game performance across six sessions.
[e] The four comparisons in Study 1 involved 4 different types of dyads (e.g., mixed sexes, same sexes etc.).
[f] Several conditions per experiment were used for multiple types of instructional support. Sample size was adjusted (see Appendix A).
[g] Several measures for game performance. 'Competence in marketing, RD and production' was selected.
[h] In-game performance on test session.

Table 5
Moderator analysis comparing GBL with and without instructional support.

	d	SE	*v*	*k*	N_{adj}	95% Confidence interval	
						Lower	Upper
Learning outcome							
Knowledge	.33***	.07	.00	36	1200	.20	.98
Skills	.62***	.12	.02	32	915	.37	1.83
In-game performance	.19***	.05	.00	38	1446	.08	.55
Mixed	–	–	–	1	114	–	–
Instructional support classified on cognitive process							
Selection	.46***	.08	.01	51	1582	.31	1.38
Organization/integration	.16**	.05	.00	48	1929	.07	.46
Unknown	.78***	.16	.04	8	163	.38	2.31
Instructional support classified on group of support							
Reflection	.29**	.10	.01	15	438	.11	.87
Modeling	.46***	.12	.01	10	370	.23	1.37
Advice	.13	.11	.01	16	358	−.09	.37
Collaboration	.14*	.06	.00	14	958	.01	.40
Control	.13	.14	.02	12	223	−.13	.39
Narrative	.11	.16	.03	7	319	−.21	.34
Modality	1.24***	.17	.03	10	288	.90	3.67
Feedback	.49***	.13	.02	4	369	.23	1.44
Personalization	1.06***	.28	.08	4	115	.50	3.13
Others	.15	.13	.02	15	239	−.10	.45
Domain							
Biology	.59***	.11	.01	35	1261	.38	1.76
Math/arithmetic	.40*	.16	.02	11	172	.10	1.19
General problem solving	.17**	.06	.00	37	1304	.06	.50
Economics	.16	.10	.01	13	414	−.04	.46
Others	.40**	.12	.02	11	524	.16	1.19
Age							
Elementary school	.19*	.08	.01	15	697	.04	.57
Middle/High school	.18	.14	.02	13	210	−.10	.52
College/university	.41***	.06	.00	73	2374	.29	1.20
Adults	.18	.12	.01	6	394	−.06	.53
Level of realism							
Schematic	.45***	.07	.00	66	2242	.32	1.34
Cartoonlike	.18*	.09	.01	15	506	.00	.52
(Photo)realistic	.01	.08	.01	18	603	−.15	.04
Unknown	.47	.13	.02	8	324	.21	1.38
Methodological variables							
Publication source							
Peer-reviewed journal	.44***	.06	.00	76	2543	.32	1.31
Proceedings	.08	.11	.01	9	585	−.14	.24
Unpublished	.14	.09	.01	22	547	−.03	.41
Randomization							
Yes	.36***	.05	.00	95	3037	.26	1.08
No	–	–	–	1	45	–	–
Unknown	.22**	.08	.01	11	593	.06	.65
Design							
Posttest only	.56***	.08	.01	42	1740	.40	1.67
Pre–posttest	.16**	.05	.00	59	1878	.07	.48
Unknown	−.17	.27	.07	6	57	−.69	−.49

Note: d = weighted mean effect size (*$p < .05$; **$p < .01$; ***$p < .001$); SE = standard error of the effect size; v = variance of the effect size; k = number of pairwise comparisons; N = sum of the sample sizes of each pairwise comparison.

5.2.1. Learning outcome

The measurement of the learning outcome moderator reveals significant improvement in knowledge ($d = .33$, $p < .001$), skills ($d = .62$, $p < .001$) and in-game performance ($d = .19$, $p < .001$). Comparisons based on the Holm-Bonferroni procedure show that the effect of instructional support is larger for skills than for in-game performance ($z_{\text{skills-in-game performance}} = 3.17$, $p < .001$) and knowledge ($z_{\text{skills-knowledge}} = 2.03$, $p < .05$). In addition, the effect of instructional support is larger for knowledge that for in-game performance ($z_{\text{knowledge-in-game performance}} = 1.68$, $p = .05$). In short, players benefit from instructional support for each type of learning outcome, but the effect is largest for skills and knowledge.

5.2.2. Type of instructional support

The analysis according to the first classification reveals that instructional support for selection ($d = .46$, $p < .001$) and organization/integration ($d = .14$, $p < .01$) improve learning. However, relative to organization/integration the effect for selection is much larger ($z_{\text{selection-organization/integration}} = 3.56$, $p < .001$).

The results of the analysis of the second classification show that instructional support classified as reflection ($d = .29$, $p < .01$), modeling ($d = .46$, $p < .001$), collaboration ($d = .14$, $p < .05$), modality ($d = 1.24$, $p < .001$), feedback ($d = .49$, $p < .001$) and personalization ($d = 1.06$, $p < .001$) all improve learning. The other classifications did not reach statistical significance (advice: $d = .13$; control: $d = .13$; narrative: $d = .11$; others: $d = .15$, all $p > .05$). Comparisons based on the Holm-Bonferroni procedure show that modality improved learning more in comparison to collaboration ($z_{\text{modality-collaboration}} = 5.99$, $p < .0001$), advice ($z_{\text{modality-advice}} = 5.42$, $p < .0001$), control ($z_{\text{modality-control}} = 5.06$, $p < .0001$), others ($z_{\text{modality-others}} = 5.03$, $p < .0001$), reflection ($z_{\text{modality-reflection}} = 4.78$, $p < .0001$), narrative ($z_{\text{modality-narrative}} = 4.74$, $p < .0001$), modeling ($z_{\text{modality-modeling}} = 3.73$, $p < .0001$) and feedback ($z_{\text{modality-feedback}} = 3.45$, $p < .001$). Likewise, personalization improved learning more than collaboration ($z_{\text{personalization-collaboration}} = 3.17$, $p < .001$) and advice ($z_{\text{personalization-advice}} = 3.06$, $p < .01$). All other comparisons show no differences (all $p > .05$). In short, instructional support that facilitates learners in selecting relevant information improves learning more than instructional support that stimulates organization/integration of new information. Especially modeling, modality, personalization and feedback appear to be effective techniques.

It should be noted that the results are influenced by confounding effects: all modeling comparisons used the same age group (students), all modality and feedback comparisons were in the same domain (biology) and with the same age group (students).

5.2.3. Instructional domain

The analysis revealed moderate effect sizes for biology, mathematics, general problem solving and others (resp. $d = .59$, $p < .001$; $d = .40$, $p < .05$, $d = .17$, $p < .01$ and $d = .40$, $p < .01$). The effect of instructional support in economics did not reach statistical significance ($p > .1$). Comparisons based on the Holm-Bonferroni procedure between the domains make clear that instructional support in biology improves learning more than in general problem solving ($z_{\text{biology-general ps}} = 3.54$, $p < .001$) and economics ($z_{\text{biology-economics}} = 3.01$, $p = .001$). All other comparisons show no differences (all $p > .05$). In short, the results show that instructional support can be effective in GBL in several domains.

It should be noted that the results are influenced by confounding effects: all biology comparisons were with the same age group (students) and all math/arithmetic comparisons had the same level of realism (schematic).

5.2.4. Age

From the four age levels (elementary school, middle/high School, college/university and adults), instructional support was only effective for elementary school and college/university (resp. $d = .19$, $p < .05$; $d = .41$, $p < .001$). For middle/high school ($d = .18$, $p > .05$) and adults ($d = .18$, $p > .05$) the effect of instructional support was not statistically significant. Comparisons based on the Holm–Bonferroni procedure between the age levels show no statistical differences (all $p > .05$).

5.2.5. Level of realism

From the levels that were defined (schematic/textual, cartoonlike, realistic and unknown) we found that instructional support improved learning in games with a schematic/textual ($d = .45$, $p < .001$) and a cartoonlike ($d = .18$, $p < .05$) design. The effect size for realistic games was not significant ($d = .01$, $p > .1$). Comparisons based on the Holm-Bonferroni procedure between the levels of realism reveal that schematic/textual designs improved learning more than realistic ($z_{\text{schematic-realistic}} = 4.11$, $p = .0001$) and cartoonlike designs ($z_{\text{schematic-cartoonlike}} = 2.45$, $p < .01$). In short, instructional support in basic/schematic designs seems to improve learning more than in cartoonlike and (photo)realistic designs.

It should be noted that most of the comparisons in the cartoonlike as well as the realistic levels of realism use the same computer game (cartoonlike: 13 of 15 comparisons use the simulation game KMQuest, realistic: 17 of 23 comparisons used the game Code Red: Triage).

5.2.6. Methodological moderators

Turning to the methodological moderators we see that only studies in peer-reviewed journals report higher learning gains for GBL. Comparisons of the different publication sources show that the reported effect size for peer-reviewed journals is larger than for unpublished studies and studies in proceedings (resp. $z_{\text{journals-unpublished}} = 2.90$, $p < .01$ and $z_{\text{journals-proceedings}} = 2.81$, $p < .01$). With respect to random assignment to the serious game group and comparison groups we found only one study that reported a non random design, while 4 studies involving 11 comparisons failed to include information on the random assignment of participants. The experimental design of the study (posttest only: $d = .56$ vs. pretest–posttest design: $d = .16$) had an impact, since studies with a posttest only design reported higher effect sizes than studies using a pre–posttest design ($z_{\text{posttest only–preposttest}} = 4.21$, $p < .0001$). In short, methodological factors such as publication source and experimental design moderate the reported effect size.

6. Discussion

In this meta-analysis we consider GBL as a complex learning environment. From a cognitive perspective on learning this implies that without instructional support players may use their limited cognitive capacity for ineffective activities (e.g., focusing on irrelevant information) at the expense of activities that contribute to learning (e.g., reflection). The weighted mean effect size of $d = .34$ can be qualified as medium and it confirms that the use of instructional support in GBL can improve learning. These results are consistent with other research into the role of instructional support in complex learning environments. With regard to multimedia learning environments various analyses have shown that well-designed instructional support (e.g., segmentation, pretraining) improves learning, because they help learners to refrain from ineffective use of cognitive capacity (Mayer & Moreno, 2003; van Oostendorp et al., 2008; Wouters et al., 2008). The findings also connect well with recent research on discovery learning with which GBL has a commonality in the active exploration of the learning environment. Several reviews have shown that, at least for novices, unstructured discovery learning can be made more effective (and efficient) when instructional support is implemented (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; De Jong & van Joolingen, 1998; Kirschner, Sweller, & Clarke, 2006; Mayer, 2004).

In subsequent moderator analyses we also looked in more detail to the effects of instructional support in GBL. These analyses reveal that instructional support that facilitates learners in selecting relevant information is more effective than instructional support that stimulates organizing and integrating new information. This is also reflected in the analysis of different types of instructional support: especially modeling (showing which information is important in order to solve a problem and how to solve a problem), modality (the use of the audio channel for verbal explanations to limit visual search) and feedback (information whether and/or why an answer is correct) are effective techniques to support learners in selecting relevant information. It seems more difficult to implement instructional support in such a way that learners are prompted to actively engage in the organization and integration of new information. The only exception is reflection ($d = .29$) in which learners are explicitly asked to think about their actions or the answer they have given. Instructional support that less explicitly stimulates the organization and integration of information such as narrative elements and collaboration seem to be less effective (resp. $d = .11$ and $d = .14$).

In game genres such as adventure games and role playing games, narratives play an important role (Prensky, 2001). Narratives may serve as a cognitive framework which makes a task meaningful, allow individuals to explicate ideas, to interpret events and to guide actions (Jonassen, 1999). For example, compared with expository text or newspaper items, stories yield better recall, generate more inferences, and are more entertaining (Graesser, Singer, & Trabasso, 1994). Yet, in games narrative elements were not found to improve learning. From a cognitive perspective it can be argued that an engaging narrative may distract learners from the learning material and, given working memory constraints, withheld them from cognitive activities that yield learning (Adams, Mayer, MacNamara, Koenig, & Wainess, 2012). Indeed it is possible that learners will use too much of their cognitive capacity for processing the narrative information that is not directly related to the learning content. Several approaches can be used to overcome this problem. First, align the narrative theme of the game with the instructional goals (Adams et al., 2012). Second, design GBL in such a way that game mechanics, game content (the narrative) and learning content are seamlessly integrated. Some evidence for the feasibility of this approach comes from Habgood and Ainsworth (2011) who have reported higher learning gains and motivation when arithmetic was integrated with the game mechanics and content of a 3-D adventure game. The third approach states that a narrative consists of a series of events (e.g., an initiating event, exposition, complication, climax and resolution) and that the deliberate manipulation of the order of these events may trigger cognitive activities that can contribute to learning (cf. Hoeken & van Vliet, 2000). The role of the manipulation of narrative events in games is still unexplored, but research on texts provides a promising prospect. Hoeken and van Vliet (2000) manipulated the order of story events in a detective story in order to investigate how curiosity, suspense and surprise influenced the cognitive and affective processing of the story. Of these manipulations, the introduction of surprise was most effective in terms of recall of story information and appreciation of the story. Kintsch (1980) has explained the cognitive effect of surprise by arguing that readers are forced to reassess the preceding events (cognitively organized in a mental model) in order to fit in the unexpected surprising event. Some evidence for the feasibility of this proposition in GBL comes from van der Spek (2011) who found that specifically designed unexpected events caused that learners could not rely anymore on the procedure that they had learned and forced them to rethink the procedure and develop another solution. The results showed that unexpected events yielded a higher level of deep knowledge without a decline in the reported engagement.

Also the implementation of collaboration does not improve learning. A rationale for collaboration is that it is supposed to stimulate learners to explicate their knowledge (Van der Meij, Albers, & Leemkuil, 2011). Theoretically, collaboration should foster the articulation of knowledge but playing a game in a dyad does not guarantee that the player will discuss about issues such as the characteristics of the task. Indicative is the analysis of the discussion protocols by Van der Meij et al. (2011) which showed that much discussion involved superficial game features such as movements in the game. In addition, the authors propose that scripted collaboration in which partners are assigned a specific role or task may improve learning (see also Hummel et al., 2011; Weinberger, Ertl, Fisher, & Mandl, 2005). We agree that scripted collaboration and computer games can be a promising combination. It would also be interesting to investigate if the integration of scripted collaboration in a narrative gives an added value.

Several scholars have argued that without instructional support in GBL players are more likely to learn to play the game (i.e., in-game performance) rather than learn domain-specific knowledge and skills (Ke, 2009; Leutner, 1993). In a study with a role playing simulation game in which the player has to act as a farmer Leutner (1993) observed that players without any instructional support learned how to play the game, but only acquired limited domain-specific knowledge such as facts, rules and principles. Players with instructional support, on the other hand, acquired much more domain-specific knowledge, but learned to play the game in a limited degree. This connects well with the observation in this meta-analysis that instructional support is more effective for the acquisition of knowledge and skills than for improving in-game performance. The implication is that instructional support in GBL is desirable when the objective is that players explicitly acquire knowledge and skills.

There is some controversy whether GBL fosters higher-order thinking such as required in skills more than the acquisition of knowledge. Recent reviews have shown that GBL is not more effective for skills than for knowledge acquisition. Sitzmann (2011) compared the effect of simulation games on knowledge (declarative) and skill-based (procedural) outcome, but did not find differences. In our own review we also found that GBL was not more effective for knowledge than for skills (Wouters et al., submitted for publication). Other scholars have suggested that GBL is more effective for skills than for knowledge (Dempsey, Rasmussen, & Lucassen, 1996; Ke, 2009). Although this controversy seems inconclusive, the results of this meta-analysis show that GBL extended with instructional support is more beneficial for skills than for knowledge acquisitions. Assuming that learning skills is cognitively more demanding than acquiring knowledge, a possible explanation may be that the application of skills in already complex learning environments is more likely to cause overload or ineffective use of the cognitive system which can be avoided or optimized by appropriate instructional support (e.g., by supporting players to focus on relevant information).

The observation that instructional support in GBL with a schematic level of realism is more effective than a cartoonlike or realistic level of realism corroborates the findings of Vogel et al. (2006) and our own meta-analysis regarding the effect of serious games (Wouters et al., submitted for publication). From the perspective of learning, there is no argument to opt for (photo)realistic and cartoonlike visual designs that involve more technological expertise (and thus time and money) in comparison to schematic visual designs. It would be interesting to further categorize studies that we call (photo)realistic into 3-D (photo)realistic, 2-D (photo)realistic and virtual reality games and to investigate how these levels of realism moderate learning. This would be particularly interesting when these new levels of realism are related to specific domains and types of knowledge. For example, are 3-D and virtual reality game environments more effective for learning a medical triage (the classification of victims), than a plain 2-D (photo)realistic game environment?

6.1. Limitations and directions for future research

Scholars have different views on what studies to include in a meta-analysis, varying from a broad sample with different study characteristics coded to a restricted sample that meets specific criteria. In this meta-analysis we have chosen a broad focus including not only studies conducted in controlled laboratory settings, but also studies that took place in a classroom setting. At the same time we have tried to further qualify the weighted mean effect size of the analysis with a number of moderators such as the methodological quality of the studies (randomization or not, pretest–posttest or posttest only). The moderator 'Experimental design' indicates that the robustness of the experimental design plays a role: studies in which the prior knowledge of the learners is considered in the statistical analysis show a significant lower effect size in favor of instructional support than designs using a posttest only design. This is in contrast with other meta-analyses regarding game-based learning (Sitzmann, 2011; Wouters et al., submitted for publication). The question is whether this can be attributed to the experimental design only or to other factors as well. Although a pretest–posttest design is more robust, we believe posttest only studies are valid in meta-analyses as long as the participants are randomized across the several experimental groups in the study. In our meta-analysis randomization was used in almost all studies. A confound with another moderator is a possible explanation for the larger effect size for the posttest only studies. For example, a closer examination of the results show that posttest only studies are relatively more often conducted in biology which has a very high effect size in favor of instructional support (see Table 5). When we remove the biology studies from the meta-analysis, the difference between pretest–posttest and posttest only becomes smaller (resp. $d = .16$ and $d = .31$, but the difference is not significant). More research is required to investigate whether the difference is truly attributable to the experimental design or to confounds with other variables.

Also the publication source moderates the results. Although our calculation of the fail-safe N showed that a publication bias was not likely to occur, this conclusion is refuted by the publication source moderator. For this meta-analysis the implication may be that the observed effects are only valid when peer-reviewed journal studies are taken into account. Our findings are in line with those of Sitzmann (2011) on simulation games. The conclusion is that researchers who review the literature (a meta-analysis, but also more qualitative reviews) should include methodological variables such as the publication source and the experimental design in their report in order to better understand how the selection of studies has influenced the results of the review.

Some of the characteristics of this meta-analysis should be kept in mind when interpreting the results. To start with, we used a broad definition of instructional support which may imply that it is difficult to generate and interpret the weighted mean effect size. We have tried to overcome this by adopting a random-effects model and a further qualification of the effect of instructional support by using two different perspectives. Firstly, we have classified instructional support according to the cognitive processes that they aim at (selection vs. organization/integration). In some cases this may be arbitrary. For example, we have classified 'surprising events' as a type of instructional support that triggers players to engage in organizing and integrative activities. At the same time a surprising event may also focus a player's attention to relevant information. In these cases we are guided by our idiosyncratic assessment of the most important cognitive process as described in the study, but we realize that other valid choices are possible as well. Another point that has to be resolved is the strong effect size for the 'Unknown' category (instructional support that could not be classified under one of the types of cognitive processes). This is largely attributable to the fact that all four comparisons of the group Personalization with a $d = 1.06$ are classified as Unknown. Secondly, we have classified the different types of instructional support that belong together in groups. For example all instructional support that provided system-generated hints, suggestions or alerts was grouped as Advice (see Table 2).

This review considers the moderator variables separately. For example, we investigated the effect of different types of instructional support on learning in general, but we do not know how these different types of instructional support influence the different types of learning outcome such as knowledge, skills and in-game performance. This means that we have no understanding of the interactions between the moderator variables. Although we did not find a general effect of Advice on learning, it is possible that we will find different effects for Advice on knowledge, skills and in-game performance. More research is required on the possible interaction between moderators.

Finally, the list of moderator variables that we used is not comprehensive. There are other moderator variables that may potentially influence the effectiveness of instructional support in GBL. We did not involve these variables because they were not in line with the focus of the meta-analysis (for example instructional support that aims to improve learning by increasing the motivation which will make learners spent more time and effort on learning) or because too few studies were available (e.g., the impact of instructional support when delayed testing is taken into account, or the issue whether GBL is used solely or in combination with other instructional methods).

Summarized, the use of instructional support in GBL can be effective, especially when learning skills is involved and when the support enables the learners to select relevant information and thus prevent overloading the cognitive system. The impact of instructional support on organizing and integrating the new information with prior knowledge was limited. In order to foster such organizational and integrative cognitive activities we recommend more research on instructional support that connects well with the game play (e.g. scripted collaboration and integrated narrative elements). Also for designers of GBL the results of the meta-analysis yield some guidelines. They should focus on the learning content and less on the visual design. When a narrative is important, consider to integrate it with the learning content. Apply instructional support to foster cognitive skills and the acquisition of knowledge rather than to stimulate in-game performance.

Acknowledgment

This research is funded by the Netherlands Organization for Scientific Research (project number 411-10-902).

This research is also supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie).

Appendix A

In the formula $N_{adjusted} = ([N_{experimental\ group}/a] + [N_{comparison\ group}/b])/c$ the letter '*a*' means the number of comparison groups (GBL without instructional support), '*b*' the number of experimental groups (GBL with instructional support) and '*c*' means the number of dependent variables. For example, in chapter 7 Leemkuil (2006) uses two experimental groups (alternative advice, which we classified as Reflection, and normal advice) and two learning outcomes (knowledge and in-game performance). This results in four pairwise comparisons with in total 266 participants. The number of dependent variables is 2 ($c = 2$), each experimental group is only used once ($a = 1$) and the comparison condition is used twice ($b = 2$).

Pairwise comparison	$N_{Experimental\ group}$	$N_{Comparison\ group}$	Dependent variable	Formula	Adjusted *n*
Advice–no advice	85	100	Game	([85/1] + [100/2])/2	67.5
Advice–no advice	85	100	Knowledge	([85/1] + [100/2])/2	67.5
Alt. advice–no advice	81	100	Game	([81/1] + [100/2])/2	65.5
Alt. advice–no advice	81	100	Knowledge	([81/1] + [100/2])/2	65.5
				Total *N*	266

References[1]

Abdi, H. (2010). Holm's sequential Bonferroni procedure. In N. Salkind (Ed.), *Encyclopedia of research design* (pp. 1–8). Thousand Oaks, CA: Sage.

*Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*, 235–249.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*, 1–18.

*Arthur, W., Day, E. A., Bennet, W., McNelly, T. L., & Jordan, J. A. (1997). Dyadic versus individual training protocols: loss and reacquisition of a complex skill. *Journal of Applied Psychology, 82*, 783–791.

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

*Blumberg, F. C. (2000). The effects of children's goals for learning on video game performance. *Journal of Applied Developmental Psychology, 21*, 641–653.

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

*Cameron, B., & Dwyer, F. (2005). The effect of online gaming, cognition and feedback type in facilitating delayed achievement of different learning objectives. *Journal of Interactive Learning Research, 16*, 243–258.

*Conati, C., & Zhao, X. (2004). Building and evaluating intelligent pedagogical agent to improve the effectiveness of an educational game. In *Proceedings of the 9th international conference on intelligent user interfaces (IUI '04)* (pp. 6–13). New York, NY, USA: ACM, doi:10.1145/964442.964446.

*Cordova, D. L., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*, 715–730.

*Day, E. A., Boatman, P. R., Kowolik, V., Espejo, J., McEntire, L. E., & Sherwin, R. E. (2007). Collaborative training with a more experienced partner: remediating low pretraining self-efficacy in complex skill acquisition. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*, 1132–1148.

De Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179–201.

Dempsey, J. V., Rasmussen, K., & Lucassen, B. (1996). Instructional gaming: implications for instructional technology. In *Proceedings of the annual meeting of the Association for Educational Communications and Technology*, Nashville, TN.

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

Fletcher, J. D., & Tobias, S. (2006). *Using computer games and simulations for instruction: A research review* (pp. 1–14). Paper in the proceedings of the Society for Applied Learning Technology Meeting. Orlando, FL: New Learning Technologies.

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

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331.

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

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395.

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

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

Hoeken, H., & van Vliet, M. (2000). Suspense, curiosity, and surprise: how discourse structure influences the affective and cognitive processing of a story. *Poetics, 26*, 277–286.

Hummel, H. G. K., van Houcke, J., Nadolski, R. J., van der Hiele, T., Kurvres, H., & Löhr, A. (2011). Scripted collaboration in serious gaming for complex learning: effects of multiple perspectives when acquiring water management skills. *British Journal of Educational Technology, 42*, 1029–1041.

*Inkpen, K., Booth, K. S., Klawe, M., & Upitis, R. (1995). Playing together beats playing apart, especially for girls. In *Proceedings of computer support for collaborative learning '95 (CSCL)*, Bloomington, Indiana. Retrieved 27.09.11, from http://www.cs.ubc.ca/nest/egems/reports/koricscl95.doc.

Jacobs, J. W., & Dempsey, J. V. (1993). Simulation and gaming: fidelity, feedback and motivation. In J. V. Dempsey, & G. C. Sales (Eds.), *Interactive instruction and feedback*. Englewood Cliffs, New Jersey: Educational Technology Publications.

*Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior, 26*, 1246–1252.

Jonassen, D. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.). *Instructional design theories and models: A new paradigm of instructional theory, Vol. II* (pp. 215–239). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Kintsch, W. (1980). Learning from text, levels of comprehension, or: why anyone would read a story anyway? *Poetics, 9*, 87–98.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Kirschner, P. A., Sweller, J., & Clarke, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

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*, 311–328.

*Lang, J., & O'Neil, H. (2008). *The effect of presenting just-in-time worked examples for problem solving in a computer game*. Paper presented at the American Educational Research Association, New York, US.

*Leemkuil, H. H. (2006). *Is it all in the game? Learner support in an educational knowledge management simulation game*. Unpublished doctoral dissertation, Universiteit Twente, The Netherlands.

[1] References marked with an asterisk indicate studies included in the meta-analysis.

Leemkuil, H., & de Jong, T. (2011). Instructional support in games. In S. Tobias, & D. Fletcher (Eds.), *Computer Games and Instruction* (pp. 353–369). Charlotte, NC: IAP.
Leemkuil, H., de Jong, T., & Ootes, S. (2000). *Review of educational use of games and simulations (IST-1999–13078 deliverable D1)*. Enschede, The Netherlands: University of Twente.
*Leutner, D. (1993). Guided discovery learning with computer-based simulation games: effects of adaptive and non-adaptive instructional support. *Learning & Instruction, 3*, 113–132.
Malone, T. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4*, 333–369.
Mayer, R. (2001). *Multimedia learning*. New York: Cambridge University Press.
Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning?: the case for guided methods of instruction. *American Psychologist, 59*, 14–19.
Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*, 760–769.
Mayer, R. E. (2011). Multimedia learning and games. In S. Tobias, & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 281–305). Greenwich, CT: Information Age Publishing.
*Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*, 171–185.
Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52.
*McQuiggan, S., Rowe, J., Lee, S., & Lester, J. (2008). *Story-based learning: the impact of narrative on learning experiences and outcomes*In *Intelligent tutoring systems, lecture notes in computer science, Vol. 5091*. Berlin, Heidelberg: Springer Berlin, (pp. 530–539).
*Meluso, A., Zheng, M., Spires, H. A., & Lester, J. (2012). Enhancing 5th graders' science content knowledge and self-efficacy through GBL. *Computers & Education, 59*, 497–504.
*Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: role of methods and media. *Journal of Educational Psychology, 94*, 598–610.
*Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*, 165–173.
*Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology, 97*, 117–128.
*Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents. *Cognition and Instruction, 19*, 177–213.
O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: evidence from the computer games literature. *The Curriculum Journal, 16*, 455–474.
van Oostendorp, H., Beijersbergen, M. J., & Solaimani, S. (2008). Conditions for learning from animations. In *Proceedings of the 8th international conference of the learning sciences* (pp. 438–445). International Society of the Learning Sciences.
Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist, 38*, 1–5.
Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.
*Prislin, R., Jordan, J. A., Worchel, S., Semmer, F. T., & Shebilske, W. L. (1996). Effects of group discussion on acquisition of complex skills. *Human Factors, 38*, 404–416.
*Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Wong, W. L. (2009). Multimodality and interactivity: connecting properties of serious games with educational outcomes. *Cyber Psychology & Behavior, 12*, 691–697.
Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*, 183–192.
*Sandberg, J. A. C., Wielinga, B. J., & Christoph, L. H. (2012). The role of prescriptive models in learning. *Computers & Education, 59*, 839–854.
*Shebilske, W. L., Goettl, B. P., Corrington, K., & Day, E. A. (1999). Interlesson spacing and task-related processing during complex skill acquisition. *Journal of Experimental Psychology, 5*, 413–437.
*Shebilske, W. L., Regian, W., Arthur, W., & Joran, J. (1992). A dyadic protocol for training complex skills. *Human Factors, 34*, 369–374.
Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489–528.
van der Spek, E. D. (2011). *Experiments in serious game design: A cognitive approach*. Unpublished doctoral dissertation, Universiteit Utrecht, The Netherlands.
*Van der Meij, H., Albers, E., & Leemkuil, H. (2011). Learning from games: does collaboration help? *British Journal of Educational Technology, 42*, 655–664.
*Van Eck, R., & Dempsey, J. (2002). The effect of competition and contextualized advisement on the transfer of mathematics skills in a computer-based instructional simulation game. *Educational Technology, Research & Development, 50*, 23–41.
Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: a meta-analysis. *Journal of Educational Computing Research, 34*, 229–243.
Weinberger, A., Ertl, B., Fisher, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*(1), 1–30.
*Worchel, S., Shebilske, W. L., Jordan, J. A., & Prislin, R. (1997). Competition and performance on a computer-based complex perceptual-motor task. *Human Factors, 39*, 410–416.
*Wouters, P., van Oostendorp, H., Boonekamp, R., & van der Spek, E. D. (2011). The role of game discourse analysis and curiosity in creating engaging and effective serious games by implementing a back story and foreshadowing. *Interacting with Computers, 23*, 329–336.
Wouters, P., Paas, F., & van Merriënboer, J. J. M. (2008). How to optimize learning from animated models: a review of guidelines based on cognitive load. *Review of Educational Research, 78*, 645–675.
Wouters, P. J. M., van der Spek, E. D., & van Oostendorp, H. (2008). Serious games for crisis management: what can we learn from research on animations? In A. Maes, & S. Ainsworth (Eds.), *Exploiting the opportunities: Learning with textual, graphical and multimodal representations* (pp. 162–165) Tilburg: Earli.
Wouters, P., van der Spek, E. D., & van Oostendorp, H. (2009). Current practices in serious game research: a review from a learning outcomes perspective. In T. M. Connolly, M. Stansfield, & L. Boyle (Eds.), *Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices* (pp. 232–255). Hershey, PA: IGI Global.
Wouters, P., van Oostendorp, H., van Nimwegen, C., & van der Spek, E. D. A meta-analysis of the cognitive and motivational effects of serious games, submitted for publication.
Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38*(9), 25–32.