

You cannot do this alone! Increasing task interdependence in cooperative educational videogames to encourage collaboration

Steve Nebel¹  · Sascha Schneider¹ · Maik Beege¹ · Franziska Kolda² · Valerie Mackiewicz² · Günter Daniel Rey¹

Published online: 26 January 2017

© Association for Educational Communications and Technology 2017

Abstract Complex, multimedia software such as educational videogames offer a wide range of elements to modify learner behavior. The adjustment of such software might support learning, especially in complex settings like collaborative or cooperative scenarios. Coming from a theoretical background of educational psychology, our experiment seeks to implement the “jigsaw strategy” within educational videogames. We conducted an experiment with sixty participants to compare conditions with or without increased task interdependence through the jigsaw strategy (i.e., the distribution of game elements and essential information). The participants had to rebuild a house from the 1894 novel *Effi Briest* within an adjusted version of the “sandbox” game *Minecraft*. The results show increased play performance, and learning outcomes with increased task interdependence. We conducted mediator and moderator analysis, which revealed a strong impact of play performance on learning outcomes. Additional analyses of mental effort, cognitive load, and efficiency allowed for deep insights into the playing and learning process. These insights enrich current theories about collaboration, mental strain, and the working memory effect and highlight the applicability of collaborative mechanics within educational videogames.

Keywords Collaboration · Educational videogames · Teaching literacy · Jigsaw strategy · Interdependence · Learning efficiency

✉ Steve Nebel
steve.nebel@phil.tu-chemnitz.de

¹ E-Learning and New Media, Faculty of Humanities, Technische Universität Chemnitz, Straße der Nationen 12, 09111 Chemnitz, Germany

² Media Communication at the Technische Universität Chemnitz, Chemnitz, Germany

Introduction

Collaboration in game environments came to the attention of researchers again when off the shelf games such as *World of Warcraft* (2004) paradigmatically showed that cooperation can enhance the appeal of a game, and can indeed serve as a substantial game feature. Collaboration itself, as “a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem” (Teasley and Roschelle 1993, p. 235), is not a new concept, and has been the subject of investigations in the educational context before. Several issues have arisen as a result of such collaboration. For example, if students collaborate without additional guidance, they proceed in a simple step-by-step manner, rather than in an abstract, “planful” way (King 2008). Additional challenges to initiate collaboration arise with “collaboration lethargy” (Azadegan and Hartevelde 2014), which is the lack of a natural tendency to collaborate. Slavin (1980), in contrast, observed that if users see chances to increase their own rewards by increasing group performance, they might support each other voluntarily.

The potential of videogames can be harnessed to address these issues, as they consist of various elements that can be manipulated to regulate collaborative behavior. For example, the way to foster cooperation in the first place can be derived from massive multiplayer online roleplaying games such as *World of Warcraft* (2004). Players fulfill certain exclusive roles, while game mechanics (e.g., healing or guarding) are distributed among them. Players thus cooperate to beat the game; they appreciate this task distribution as a core game mechanism. Furthermore, it seems natural that a warrior would fulfill different tasks than a sorceress. Compared to learning tasks, this offers new possibilities, since (for example) distributing a calculator, pen, and paper to players to solve a calculation task might appear arbitrary and unnatural to them. Therefore, this research tries to connect traditional educational strategies (i.e. collaboration) with new mechanisms within educational videogames. Additionally, educational methods and strategies need to be frequently updated, as the technological circumstances change drastically. In this vein, new additions arise, as for example, the cooperation of players around the world becomes a common phenomenon. Furthermore, this research aims to reveal further insights into the process, by using new approaches out of the field of cognitive psychology.

Literature review

Cooperation and collaboration

Collaborative videogames are games in which players work together to achieve shared goals (Azadegan and Hartevelde 2014); the differences between collaboration and cooperation emerge when looking more closely at these goals. Whereas cooperative games include tasks where players have neither completely opposed nor coincident goals, collaborative scenarios include tasks where every player has exactly the same goal (Azadegan and Hartevelde 2014). Another way to distinguish between cooperation and collaboration is the analysis of the distribution of labor. In cooperative settings, users are able to split the task and often tend to work alone, while collaborative settings result in users working together (Dillenbourg 1999; Dillenbourg et al. 1995; Huber and Huber 2008). It should be noted, however, that this difference is not stable, since a temporary cooperative behavior might also occur within collaborative settings. Thus, because the terms are often used

interchangeably (Janssen et al. 2010), it seems more plausible to regard them as a continuum rather than two exclusive categories.

Slavin (1980) introduced three basic dimensions to describe this complex process in more detail. He used task structure, reward structure, and authority structure to explore different facets of cooperation. Using these categories, learning environment developers can manipulate and optimize the “collaborativeness” of their content more specifically. For example, the reward structure can be altered to reward the team as a whole, or the task structure can be modified to prevent a single player from finishing the work alone. More specifically, the latter is called task interdependency. Furthermore, the collaborative experience is anchored differently within videogames. Azadegan and Hartevelde (2014) identified simple archetypes in their review of sixty-two collaborative games. They revealed that the majority of games did not include collaboration very deeply within the gameplay. However, their results indicated the broad variety of collaborative tasks within videogames. Bearing in mind the limited number of studies that have addressed cooperation in educational videogames (Ke and Grabowski 2007), there is also the need for specific experimental approaches to be derived.

Impacts of collaborative or cooperative learning

Collaborative and cooperative mechanics demonstrate a large variety of impacts on learning. For example, learning and collaborating in small groups can outperform individual learning and classroom learning (Lou et al. 1996). It can have a positive effect on cognitive-, process-, affective-, attitude-, and persistence-related outcomes (Lou et al. 2001; Springer et al. 1999; Sung and Hwang 2013). Learning and collaborating in small groups may also lead to increased enjoyment, interest (Plass et al. 2013), learning (Fu et al. 2009), and task performance (Johnson et al. 1986; Ke and Grabowski 2007). Such learning might also have a positive influence on error rates (Mullins et al. 2011), self-esteem, and group cohesiveness (Slavin 1980).

When educational videogames use cooperative approaches, several positive game experiences (e.g.: fun, absorption, feeling of competence, empathy and involvement with others, Oksanen 2013) and an increased tendency to play the game in the future (Plass et al. 2013) can be observed. Cooperation might foster a tendency for master goal over performance goal orientation (Stevens 2008); cooperative actions might also be slower and more carefully produced (Staiano et al. 2012). These effects have drawn the interest of several educational researchers, and have elicited explanations from many different perspectives. For example, students are more likely to explain things to one another in cooperative settings. Providing this information is related to increased performance (Dillenbourg et al. 1995), since learners have to verbalize their knowledge and thus elaborate on it (Mullins et al. 2011). This process can therefore stimulate high-order cognitive processes (King 2008). This could further reinforce why conceptual knowledge (rather than procedural knowledge) might be better suited for collaborative settings (Mullins et al. 2011), and could explain why effects differ between explainers and recipients (Janssen et al. 2010; King 2008).

Another approach can be taken using cognitive load theory (Sweller 1988; Sweller et al. 1998), which focuses on the effective use of mental resources. The task imposes intrinsic load (difficulty), germane load (schema creation) and extraneous load (presentation and orientation). Individual load can decrease as the load induced by the learning task is divided across a larger “reservoir” of cognitive capacity among the collaborating learners. The learners form an effective information-processing system, benefiting from the

“distribution advantage” (Kirschner et al. 2011a, b). The necessary “transaction costs” might be useful (germane load) or harmful (extraneous load); demanding collaborative learning environments could interfere with schema construction due to limited working memory (Janssen et al. 2010). This system of trade-offs between transaction activities and distribution advantages is called the “collective working memory effect (CWME)” (Kirschner et al. 2011a, b).

Several moderating and mediating factors can be determined after analyzing different approaches. For example, the smaller the group, the larger the effects, especially when compared to competition (Johnson et al. 1981). Furthermore, the effects are influenced by group heterogeneity (Lou et al. 2001), individual prerequisites, and task features (Dillenbourg et al. 1995). The positive effects of small group learning with computer technology are also larger when tutorials are provided, general ability level is relatively low, and cooperative group learning strategies are employed rather than general encouragement or individual strategies (Lou et al. 2001). More specifically, if tasks are overly simple (Kirschner et al. 2011b; Lou et al. 2001), closed, or controlled, they are not suitable for collaboration (Kirschner et al. 2008).

Asynchronous or distributed collaborative learning has to be supported with various communication tools for technical implementation (Leemkuil et al. 2003). Even if every factor has been taken into account, social complications might still occur. For example, the task may be assigned to the group member with the strongest resources in the specific area (Huber and Huber 2008), and not to the member who would benefit most. “Social loafing” can also occur (e.g., Diziol et al. 2010), resulting in lowered learning outcomes. Addressing these challenges, educational videogames offer a wide range of functionality and creative room for rules and different types of gameplay that might improve cooperation and, subsequently, learning.

The collaboration paradox

One popular method among teachers in cooperative settings is the jigsaw strategy (Aronson 1997; Diziol et al. 2010; Huber and Huber 2008; Slavin 1980), in which information and tasks are distributed among participants, resulting in a broken up main task that needs to be solved together to be completed. Referring again to Slavin’s terminology, the jigsaw strategy can be further described with an increased positive task and information interdependence (Johnson et al. 1998; Slavin 1980). This approach offers potential, as groups where every member contributes to group success (and there is some form of group reward) outperform groups without these mechanisms (Lou et al. 1996). Educational videogames can develop these mechanisms further, as the game designer can create gameplay that cannot be solved alone, thus increasing task independency without the teacher explicitly distributing the tasks. This might be counterintuitive, since preventing certain players from using parts of the game mechanisms might appear non-user-friendly. Additionally, strong knowledge interdependency might prevent a common ground and inhibit sufficient use of core concepts (Deiglmayr and Schalk 2015). For example, a basic understanding of mathematics might be useful before collaborating on complex physics problems. Therefore, researchers must carefully implement this “collaboration paradox” (i.e., methods of hindering collaboration, subsequently fostering collaboration, Azadegan and Hartevelde 2014), and should review its impact.

The present experiment

The present study seeks to systematically manipulate the level of cooperation within two different group constellations, and to evaluate the level of cooperation's impact on learning, cooperation in play, cognitive load, efficiency, and play experience. We created two group conditions, representing different forms of cooperation. The first represents the *voluntary cooperation* (VC) group, where every member had access to the same information and game elements, and the instructor encouraged them to cooperate in order to solve the task. (they could fulfill the task alone, so it was not essential to collaborate to reach the goal.) The second group represents the *increased task interdependence* (ITI) collaboration condition, where every participant only had access to certain game elements and information; collaboration thus was essential for the ITI group.

We focused on small groups, as they reduce the diversity of views and knowledge; large groups tend to prevent sufficient participation from every member (Lohman and Finkelstein 2000). We chose an even number of group members, as groups of three tend to behave competitively, while pairs behave more cooperatively (Dillenbourg et al. 1995). After analyzing the literature on group size [four to six (AbuSeileek 2012), three to five (Lou et al. 2001), two to four (D. W. Johnson et al. 1998) and five (Kooloos et al. 2011)], we decided to form groups of four members for our experiment.

The aim of the experiment was to answer several research questions on the advantages of increased task interdependence in educational videogames. As cooperative learning is influenced from changes within task, reward or authority structure (Slavin 1980), we posited that players in the ITI group should work together more frequently and show more collaborative behavior. Cooperation, in this context, means increased interactions between participants, and should be indicated by overall speaking time and given explanations within the group. Thus, we formulated the following hypothesis:

H1 Students in the ITI group will cooperate more than students in the VC group.

In order to validate whether our manipulation (and the resulting cooperative behavior) would increase performance (in this context, performance should be indicated by the amount of criteria met for the building task) as predicted (e.g., R. T. Johnson et al. 1986; Ke and Grabowski 2007; Lou et al. 1996), we formulated a second hypothesis:

H2 Increased task interdependence will increase performance.

The resulting variation should subsequently increase learning as well. Regarding our experiment, learning will be operationalized with a cloze test and a spatial orientation task. In order to verify the assumptions from previous research (e.g., Fu et al. 2009; Sung and Hwang 2013), we postulated the following hypothesis:

H3 Students in the ITI group will exceed students in the VC group in learning tasks.

In order to gain a better indication of the quality of the cognitive schemas participants have acquired, the performance measures should be accompanied by cognitive load measures (Janssen et al. 2010). Following the assumptions of the "collective working memory effect"—CWME (Kirschner et al. 2011a, b), we assumed that the individual cognitive load would decrease among members of the ITI group. More specifically, we operationalize cognitive load as the sum of intrinsic load (IL), germane load (GL) and extraneous load (EL). We also had to differentiate between the subjective reactions to imposed tasks (i.e., cognitive effort) and the overall necessary demands (i.e., mental load)

the participants faced (Manzey 1998). In other words, mental effort represents personal characteristics, whereas mental load reflects task characteristics (Krell 2015). In contrast to the subjective load, we expected the overall mental effort to increase in the ITI condition, as the participants had additional interactions with their peers as part of the aforementioned collaboration paradox. Therefore, we formulated two opposing hypotheses:

H4a Students in the ITI group will show a lower cognitive load than students in the VC group and

H4b Students in the ITI group will have to invest more mental effort than students in the VC group.

In addition, because criteria measuring mental strains in combination with learning results can indicate the quality of learning in terms of the efficiency of schema acquisition (Kirschner et al. 2011b), we wanted to combine the assumptions from the previous hypotheses. Efficiency would be addressed as effort/cognitive load in relation to performance. Thus, we assumed that the ITI condition would create more learning efficiency regarding cognitive load, and less learning efficient regarding mental effort. In order to check this assumption, we postulated the following hypothesis:

H5 Increased task interdependence is more learning process efficient regarding cognitive load, and less learning process efficient regarding mental effort.

Methods

Participants

A total of 60 senior-class students were recruited from secondary schools in Saxony, Germany for this study. Participants were divided into collaborating groups of four students. One group had to be excluded due to problems with the gaming software during the investigation; thus 56 students (14 groups) were included in the final analysis. The voluntary cooperation (VC) condition contained six groups ($n = 24$), while the increased task interdependence (ITI) condition contained eight groups ($n = 32$). The gender ratio was balanced (51.8% male), and the ages of the participants varied from 15–20 years, with a mean of 16.88 years ($SD = 0.97$). Only ten students didn't like playing videogames at all, and only one participant liked to play educational videogames. Most of the students preferred action games ($n = 25$) and strategy games ($n = 21$). Fifty-seven percent of the participants ($n = 32$) had already played the game *Minecraft* (2011), which we used as a learning environment in this study. Students preferred playing *Minecraft* at home ($n = 27$) to playing it at school, and most played online with friends ($n = 20$). Only three students played *Minecraft* more than 2 h per week, and many had no or limited experience with the game ($n = 22$). Furthermore, neither prior experience with *Minecraft* $t(55) = 0.98$; $p = 0.70$ or age $t(55) = 2.63$; $p = 0.50$ differed between the experimental groups. One student was excluded from the cognitive load analysis due to an incomplete questionnaire.

Materials and design

We used the game *Minecraft* (2011) to create a learning environment. *Minecraft* is an “open-world” game, where players in a first-person perspective can interact with the world

by destroying and setting blocks (for a more extensive description see: Nebel, Schneider, & Rey, 2016). The game contains a creativity mode that allows players to build anything, from small areas to giant worlds. The game is distributed worldwide and first became popular, amongst other things, because of its multiplayer mode: several players can interact and collaborate, creating buildings and landscapes complementarily. Therefore *Minecraft* was ideal for building a suitable environment for the current investigation. Figure 1 illustrates the area in which the participants played together.

The basic skills of reading performance include, in addition to reading itself, mastering reading tasks such as seeking and memorizing information, reflecting, and interpreting (Baumert et al. 2001). This study aimed for literacy teaching beyond simply knowing the text: we hoped the participants would have fun with reading, understanding, imagining, and using their creativity. These are central aspects of literature education (Nickel-Bacon 2006). More specifically, our research addressed learning outcomes including the content of a text and a mental model of the components found within the text. Therefore a literary subject was particularly suitable for the present study.

The learning material consisted of two texts (an “arrangement text” and a “material text,” Table 1) written by the researchers, based on the 1894 realist novel *Effi Briest* (Fontane 2015). The arrangement text described the disposal of the property of the Briest family in detail, while the material text described the shape, size, and color of the objects. Although some works of the *literacy realism* and especially the first chapter of *Effi Briest* offers detailed descriptions of the environment, the original text had to be adapted to the capabilities of *Minecraft*. This included more precise explanations, and a few amendments. An example of an alteration was a swing in the original text, which could not be built in *Minecraft* and so was replaced by a playground sandbox. Therefore, the participants only worked with our texts, not with the original novel.

Texts were optically separated into three parts, similar in length, to support participants in their time management. Furthermore, we added a square to the ground to give the students an idea of the size of the house (Fig. 1); we showed them a note with the most important controls and translations for building materials to support the students. The study



Fig. 1 Collaborative environment created with *Minecraft*. Letters indicate cardinal directions (in German; east is “O”)

Table 1 Text example

Original novel	Arrangement text	Material text	Cloze test
[...] Some twenty paces further, and parallel to the wing of the house, there ran a churchyard wall, entirely covered with small-leaved ivy, except that place where an opening had been made for an little white iron gate. [...]	[...] parallel to the wing of the house, there ran a churchyard wall, with a gate in the middle. Along the wall, plants grow in every direction and the wall is not very high [...]	[...] The churchyard wall is made out of two layers of moss stone and a gate is a brown gate. Additionally, small-leaved ivy was growing along the wall. [...]	[...] Some twenty paces further, and parallel to the wing of the house, there ran a _____ wall, entirely covered with small-leaved ____, except that place where an opening had been made for an little brown iron gate. [...]
<i>//comment: excerpt from chapter one, translated from German</i>	<i>//comment: the exact location of the gate was added, to help deciding where to build it. Additionally, the participants were informed that the wall was not really high, to prevent excessive wall constructions</i>	<i>//comment: Since Minecraft does not provide white gates, the type of gate had to be modified. Additionally, stone and ivy are two different items within Minecraft, thus, they had to be mentioned separately</i>	<i>//comment: In order to stay consistent, the color of the gate was adjusted. The remaining text was as accurate as possible</i>

was pilot-tested to determine its difficulty, usability, tutorial, and time exposure. Five groups of students received the text about the Briest house and had to adequately build it in *Minecraft*. In order to save time, structural details were removed and the tutorial was adjusted.

We used a between-subject design to vary cooperation. Participants were randomly assigned to the surveyed groups within their school class. Groups were randomly allocated either to the VC condition or to the ITI condition, resulting in an unequal number of groups within the experimental conditions. In the former, players could solve the task alone, and cooperation was not mandatory; in the latter, however, cooperation was necessary to complete the game successfully. More specifically, in the VC condition, we gave every group member both texts and made every building material available. In the ITI condition, we handed out the arrangement text to two participants and the materials text to the two remaining players. Thus, they needed to exchange information in order to build something accurately, as none of them knew where or how to build something simultaneously without talking to other group members.

This application of the jigsaw strategy was enhanced even further, since we split the building materials among the members of the ITI condition. Thus, we utilized access to game elements as a simple method to substantially increase task interdependence without significantly restricting playability. For example, in order to build the house's backyard, the players had to build a small pond with a boat and a sandpit enclosed by wooden planks. The necessary materials were distributed among the four players. Thus, every player could dig a hole for the pond or the sand box, but only one player could fill it in with the appropriate material. Players with the boat also needed to coordinate their workflow with the other players, as they could only progress with their game element once the others were finished. We analyzed the text closely to ensure a sufficient interdependency in every part

of the building task, and carefully planned which players received which elements; this resulted in roughly nine exclusive materials per player.

Tasks

Because a common knowledge base is an important factor (Bopp 2006), the players of every group had to master a tutorial that taught them the controls and basic gameplay mechanics of *Minecraft*. They learned how to move, climb, and build and destroy blocks. The players were guided through seven different stations, where they had to complete different tasks. We thus ensured that even unexperienced players were able to complete the given task (Fig. 2).

The following experimental task was to read the given text and to try to reconstruct the property as described there. The participants were encouraged to rebuild the garden and the house as accurately as possible within 45 min, which was equal to one lesson. Our experimental manipulation started at this point, since we had distributed building materials and information texts differently. We separated the players after they had finished the tutorial in order to organize the distribution, and forced them to access the final building area through the gates via the four cardinal directions (Fig. 1). Every student had a chest that included specific building materials on these exclusive paths. In addition to these materials, we granted every player access to *dirt blocks* to fill accidentally created holes, *ladders* to reach every space, and a *building tool* to destroy blocks more efficiently. Since we used the additional modified *MinecraftEDU* (2013) program, we were able to prevent the players from (1) building in places we did not want them to go, (2) destroying essential game elements, or (3) escaping the building area.

Measurements

Because it can be challenging to analyze collaboration (e.g., Peppler et al. 2013) and to distinguish between collaboration and cooperation in text analysis (Dillenbourg et al. 1995), we chose to measure simple and comparable indicators of cooperation, and discarded differences between collaborative and cooperative behavior. More specifically, increased cooperation might show increased times of between-subject interactions and

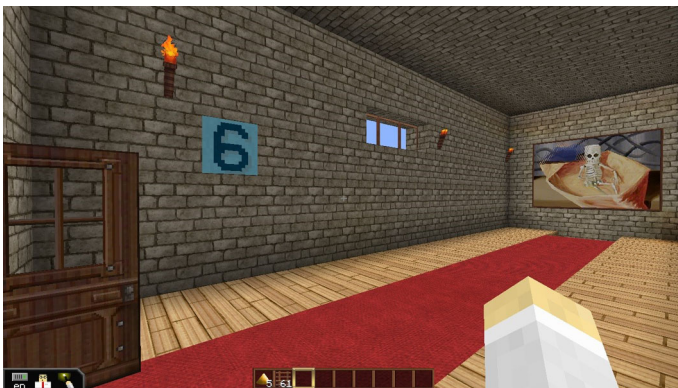


Fig. 2 Station six in the guided tutorial, with the task: “Dig your way through the red floor blocks to advance.”

higher numbers of explanations that were given. For this, we measured the overall speaking time (in seconds) of every group. More specifically, a maximum of 2700" (45') could be achieved, and as the recordings allowed no exact individual identification of every player, a group value was formed. Although this value might include other content than cooperative behavior, such as verbal conflict, we assume that it should at least serve as a simple indication of social interaction which is the necessary foundation of cooperation. Furthermore, two independent experts who were not familiar with the experimental conditions counted explanations (e.g., one player explaining another player how to build or how to interpret the text) by using audio records of the test sessions. The values that the experts counted were combined into the explanatory score ($\alpha = 0.88$). The performance measure addressed the whole group rather than individuals as well, as this is considered to be a more valid approach within collaborative scenarios to access effects like distributed cognition and other emergent effects (Dillenbourg 1999). More specifically, we analyzed the text and created a list with 51 criteria, describing in detail what the accurately finished building task should look like. For example, we checked if the house had a flat roof, or if the players placed two chairs on the porch. If the criterion was met, the group was rewarded one point; if not, the group received none.

With our game, we tried to teach the content of the first pages of *Effi Briest*, an understanding of how the area might have looked and a mental model of the environment. Thus, we used two different tests to measure learning performance, addressing different levels of elaboration of the learning material. More specifically, to measure simple factual learning, we used a "cloze test", using parts of the first page of the original text of *Effi Briest* with sixteen gaps that the participants had to fill in. To stay consistent with our learning material, we changed words that were also different in the arrangement text. The gaps were composed of single words, and we ignored the orthography in the analysis. We used spatial orientation questions to measure "schemata elaborateness". Our scale consisted of fourteen questions created by the researchers, which referred to participants' sense of direction. After our pre-test, we chose seven questions with the best distributions (averages scores located close to the center of the scale) for the final test. For example, we asked (translated from German): "You leave the house through the door in the side wing and walk toward the round bed. You look to your right; what do you see?"

To access our hypothesis regarding cognitive load, we also used a scale provided by Eysink, de Jong, Berthold, Kolloffel, Opfermann, and Wouters (2009). This scale included one question addressing intrinsic load (IL), one question addressing germane load (GL), three questions addressing extraneous load (EL), and a final question measuring overall load (OL). This final overall load item addressed effort; thus, we referred to it as the "mental effort" item, and used the scale to verify our assumptions about mental effort, as well. To access efficiency, we used the standardized values of mental effort, and the combined scores of the cognitive load scale. Furthermore, we used the standardized values of the two learning measures as learning performance indicators. In contrast to the approach of Van Gog and Paas (2008), we measured both effort indicators directly after the learning phase; thus we can only indicate *learning process* efficiency, but not *outcome* efficiency. This was a necessary variation, as we were more interested in the former (Kirschner et al. 2011a). The difference between the effort and performance values served as the basis to calculate the efficiency measure (Van Gog and Paas 2008). Lower values indicate a decreased efficiency, whereas high values can be interpreted as increased efficiency:

$$Efficiency = \frac{zP_{test} - zE_{test}}{\sqrt{2}}$$

Finally, we conducted a questionnaire, gathering the demographic data of the participants and their experiences with computer games, in particular with *Minecraft*. We asked whether the participants had played *Minecraft* before, and if so, how much experience they had. In our analysis that followed, we combined these two questions into our 0–5 “prior *Minecraft* experience” score, ranging from “no experience” to “very experienced” with *Minecraft*, respectively.

Procedure

The experiment took place in computer labs in schools, with the exception of four groups that were tested in the computer labs of our university due to a lack of technical equipment at the school. The laboratories were prepared so that four players within one experimental condition could sit side by side. All paper materials were arranged in front of each computer station. Only eight students were invited for each test run in order to minimize distracting effects. Each group was also separated by a spatial distance ranging from 4 to 6 m, as well as visual covers. Seats were allocated by drawing lots. Afterward, both groups were told to follow the experimenters’ instructions: navigating through the menu pages, choosing character names, and selecting predefined characters, which took about 5 min.

Once all participants had joined the game, they were instructed to complete the tutorial, which took about 15–20 min. Students who finished their tutorial were instructed to wait until all players were finished. In the next step, students were instructed to open the second “world.” For this, students in the voluntary cooperation group had to open a different world than the students in the increased task interdependence group in order to allocate the participants to both experimental conditions. In each condition, students had to follow one of four paths that led to their own inventory chests. All instructions concerning task and time of the experiment were provided at this point; students were also allowed to use their paper materials and to chat with their teammates. Students were allowed to build their worlds by giving a starting signal; the rest of the time was signaled every 15 min. Experimenters did not intervene, except for students who were unable to solve problems on their own. For example, when a student did not manage to reach a certain point within the world. The central building task, however, was not influenced. After 45 min of building time, all participants were instructed to immediately stop their movements within the game and switch off their monitors. Students were handed the questionnaire materials and instructed to complete each question. Once students finished their questionnaires, which took about 15–20 min, they were allowed to leave the room. Overall, the whole experiment lasted about 90 min.

Results

The data will be interpreted according to Cohen’s (1988) standards regarding effects sizes. More specially, values below $d = 0.5$ ($\eta_p^2 = 0.06$) will be interpreted as a small effect and a value below $d = 0.8$ ($\eta_p^2 = 0.14$) as a medium effect. Larger effects sizes will be

interpreted as strong effects. Based on the type of measure (group vs. individual), different calculations can be conducted. More specifically cooperation and performance will be analyzed on a group level, whereas learning, cognitive factors and efficiency will be interpreted on an individual level. With this approach, the five postulated hypothesis are analyzed (Table 2). Additional mediator and moderator analysis will shed more light on this complex topic.

H1 Students in the ITI group will cooperate more than students in the VC group.

The first hypothesis was checked by one independent *t* test. On average, participants in the ITI group ($M = 2171.25$; $SD = 469.60$) spent more time on interactions than the VC group ($M = 1964.00$; $SD = 332.29$). However, this difference is not significant $t(12) = 0.92$; $p = 0.38$. We conducted another independent *t* test to address differences in explanations. We could observe a significant result $t(12) = 3.20$; $p = 0.008$, representing a strong effect (Cohen 1988): $d = 1.75$. The comparison of means ($M_{VC} = 0.17$; $SD = 0.41 < M_{ITI} = 1.9$; $SD = 1.25$) supports the assumed direction. Subsequently, H1 can be supported for explanations, not for interaction time.

H2 Increased task interdependence will increase performance.

We conducted an analysis of covariance (ANCOVA) with “task interdependence” as the between-subject factor (considering “prior Minecraft knowledge” as the covariate), and “performance scores” as the dependent measure. All pre-defined test assumptions

Table 2 Data analysis and results

Step	Hypothesis	Analysis technique	Supported
Cooperation	Students in the ITI group will cooperate more than students in the VC group	Overall speaking time (in seconds)	No
		Given explanations	Yes
Performance	Increased task interdependence will increase performance	Group performance was compared with a list with 51 criteria	No (although a strong effect could be detected)
Learning	Students in the ITI group will exceed students in the VC group in learning tasks	Cloze test with parts of the original text	Yes
		Spatial orientation questions	Yes
Cognitive load and mental effort	Students in the ITI group will show a lower cognitive load than students in the VC group	CL as a sum of intrinsic load, germane load and extraneous load	No
	Students in the ITI group will have to invest more mental effort than students in the VC group	ME as one item (OL) within the CL questionnaire	Yes
Efficiency	Increased task interdependence is more learning process efficient regarding cognitive load, and less learning process efficient regarding mental effort	Calculated, based on CL in relation to performance measure	Yes
		Calculated, based on ME in relation to performance measure	No

were met, including Levene’s test $F(1, 12) = 0.34; p = 0.57$. We found a tendency for “performance” $F(1, 11) = 4.03; p = 0.07$; with a large effect of $\eta_p^2 = 0.27$. The follow-up estimated means corrected for “prior Minecraft knowledge” showed that students in the ITI group ($M = 33.07; SD = 9.53$) showed higher performance scores than students in the VC group ($M = 22.58; SD = 9.57$). As a result, H2 cannot be supported, although a strong effect could be detected.

H3 Students in the ITI group will exceed students in the VC group in learning tasks.

We conducted a multivariate analysis of covariance (MANCOVA) to check this hypothesis. “Prior Minecraft knowledge” scores served as the covariate, and learning scores of the cloze and schemata elaborateness tasks as the dependent measures. All pre-defined test assumptions were met, Box’s $M(3, 283014.368) = 1.179; p = 0.63$. We found a significant main effect Wilk’s $\Lambda = 0.77; F(2, 52) = 7.71; p = 0.001$; representing a strong effect $\eta_p^2 = 0.23$.

For this, follow-up ANCOVAs were conducted for each dependent variable. The test assumptions were met for the cloze task, including Levene’s test $F(1, 54) = 0.66; p = 0.42$. This ANCOVA showed a significant effect $F(1, 53) = 9.88; p = 0.003$ and a large effect size ($\eta_p^2 = 0.16$). Estimated means revealed that students with increased task interdependence ($M = 7.84; SD = 3.82$) performed better on this test format than students without increased task interdependence ($M = 4.59; SD = 3.82$). The schemata elaborateness score met the test assumptions, including Levene’s test $F(1, 54) = 0.03; p = 0.87$. Analysis showed a significant effect $F(1, 53) = 7.91; p = 0.007$, and a medium to high effect size ($\eta_p^2 = 0.13$). Corrected means revealed that the ITI group ($M = 1.74; SD = 1.00$) outperformed their experimental counterparts ($M = 0.97; SD = 1.00$). Subsequently, H3 can be fully supported.

We also conducted a follow-up mediation analysis—according to (Preacher and Hayes 2008; Warner 2008), using the multiple mediation procedure for SPSS (Hayes 2008)—in order to check if the learning scores were directly affected by task interdependence, or if this connection was mediated by the players’ performances. Each dependent variable, cloze task score, and schemata elaborateness score was checked for its mediated influences; we calculated standardized coefficients for this. We could see that interdependence was a significant predictor of performance $\beta = 0.53; t(54) = 4.64; p < 0.001$, and that performance was a significant predictor of cloze task score $\beta = 0.62; t(54) = 4.85; p < 0.001$. Interdependence, however, was no longer a significant predictor of the cloze task scores after controlling for the mediator “performance”: $\beta = 0.04; t(54) = 0.28; p = 0.78$. This result reflects a full mediation (Fig. 3). Approximately 39% of the variance in the variable cloze task scores was accounted for by the predictors ($R^2 = 0.386$). We

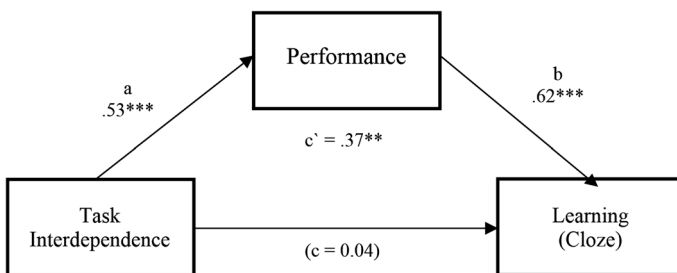


Fig. 3 Mediation by performance

tested the indirect effect by using a bootstrap estimation approach with 20,000 samples (Shrout and Bolger 2002). The indirect coefficient was significant: $\beta = 0.37$; 99% CI [0.72, 4.51].

We conducted a second mediation analysis with the same predictors, and schemata elaborateness scores as the dependent variable. The results showed that performance was not a significant predictor of schemata elaborateness $\beta = 0.09$; $t(54) = 0.61$; $p = 0.54$. We tested the indirect effect using the bootstrap approach with 20,000 samples (Shrout and Bolger 2002); no significant indirect correlation could be shown $\beta = 0.29$; 95% CI [-0.28, 0.63]. According to these results, we also checked for a moderation, as they are characterized by interactions that might have prevented mediation (Cohen et al. 2013). In order to test for linear moderation, we calculated an additional *predictor 1* (Performance \times Test condition) and calculated regressions that were in line with standardized coefficients for a model that included performance, interdependence, and *predictor 1* (Baron and Kenny 1986; Warner 2008).

The predictors accounted for approximately 21% of the variance among the schemata elaborateness scores ($R^2 = 0.21$). Regression analysis showed significant coefficients for interdependence $\beta = 1.28$; $t(4) = 2.95$; $p = 0.005$, performance $\beta = 0.42$; $t(54) = 2.12$; $p = 0.039$, and *predictor 1* $\beta = -1.22$; $t(54) = -2.41$; $p = 0.019$. According to (Baron and Kenny 1986), these results confirmed the moderation of “performance” on schemata elaborateness (Fig. 4).

H4a Students in the ICI group will show a lower cognitive load than students in the VC group.

We conducted a MANOVA, with interdependence as the between-subjects factor; “prior Minecraft knowledge” as the covariate; and intrinsic load (IL), extraneous load (EL), and germane load (GL) as dependent measures. Although all pre-defined test assumptions were met Box’s $M(6, 17066.421) = 6.75$; $p = 0.39$, we found no significant main effect Wilk’s $\Lambda = 0.952$; $F(3, 50) = 0.85$; $p = 0.48$; $\eta_p^2 = 0.05$. Statistically, null hypotheses can be accepted for an effect size of $f = 0.40$, because of sufficient power ($1 - \beta = 0.96$ for $\alpha = 0.05$) and H4a cannot be supported.

H4b Students in the ITI group will have to invest more mental effort than students in the VC group.

We conducted an ANCOVA for this hypothesis, with interdependence as the between-subjects factor, “prior Minecraft knowledge” as the covariate, and mental effort as the dependent measure. All pre-defined test assumptions were met: $F(1, 54) = 1.42$; $p = 0.24$.

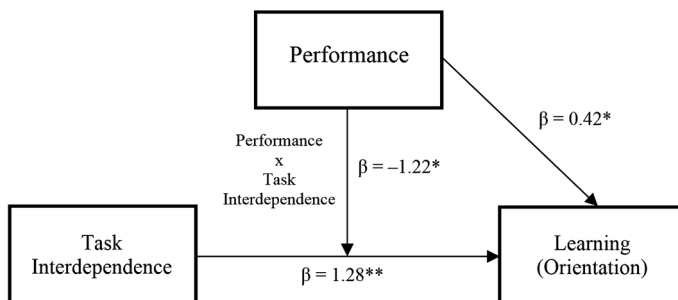


Fig. 4 Moderation by performance

The results showed a significant difference $F(1, 53) = 6.88; p = 0.011$, with a medium to high effect size ($\eta_p^2 = 0.12$). The corrected means revealed that students in the increased task interdependence group ($M = 4.63; SD = 1.71$) reported a higher amount of invested mental effort than students in the voluntary group ($M = 3.34; SD = 1.71$). Therefore, H4b is supported.

H5 Increased task interdependence is more learning process efficient regarding cognitive load, and less learning process efficient regarding mental effort.

In order to estimate if increased task interdependence within videogames was less effort- and more load- efficient, we calculated efficiency scores (according to Van Gog and Paas 2008) for both cognitive variables (overall cognitive load and mental effort) and both learning scores (cloze and schemata elaborateness tasks). We conducted a MANCOVA for cognitive load process efficiency and mental effort process efficiency on the cloze task, with “interdependence” as the between-subjects factor and “prior Minecraft knowledge” as the covariate. All pre-defined test assumptions were met Box’s $M(3, 404025.151) = 2.01; p = 0.57$; we found a significant main effect Wilk’s $\Lambda = 0.88; F(2, 51) = 3.49; p = 0.038; \eta_p^2 = 0.12$. For this step, we conducted follow-up ANCOVAs for each dependent variable (Table 3). The test assumptions were met for cognitive load efficiency $F(1, 53) = 0.24; p = 0.63$; the ANCOVA revealed a significant effect, indicating higher efficiency within the ITI condition. The test assumptions were met for mental effort efficiency $F(1, 53) = 0.002; p = 0.97$, but analysis showed no significant effect. Statistically, null hypotheses can be accepted for an effect size of $f = 0.40$, because of sufficient power ($1 - \beta = 0.82$ for $\alpha = 0.05$).

We conducted a MANCOVA for cognitive load process efficiency and mental effort process efficiency on the schemata elaborateness, with “interdependence” as the between-subjects factor and “prior Minecraft knowledge” as the covariate. All pre-defined test assumptions were met Box’s $M(3, 404,025.151) = 3.36; p = 0.36$; we found a significant main effect Wilk’s $\Lambda = 0.88; F(2, 51) = 3.58; p = 0.035; \eta_p^2 = 0.12$. For this step, we conducted follow-up ANCOVAs for each dependent variable (Table 3). The test assumptions were met for cognitive load efficiency $F(1, 53) = 1.55; p = 0.22$; the ANCOVA revealed a significant effect, indicating higher efficiency within the ITI condition. The test assumptions were met with mental effort efficiency $F(1, 53) = 0.05; p = 0.82$, but again, analysis showed no significant effect. Statistically, null hypotheses can be accepted for an effect size of $f = 0.40$, because of sufficient power ($1 - \beta = 0.82$ for

Table 3 Impact of ITI on efficiency measures

Variable	ITI		VC		$F(1,52)$	p	95% CI		
	M	SD	M	SD			LL	UL	η_p^2
CL efficiency cloze*	0.23	0.95	-0.31	0.95	4.36	0.042	0.02	1.07	0.08
ME efficiency cloze	-0.02	0.99	-0.05	0.95	0.08	0.78	-0.45	0.59	0.001
CL efficiency schemata elaborateness *	0.18	0.77	-0.28	0.77	4.78	0.033	0.04	0.88	0.08
ME efficiency schemata elaborateness	-0.03	0.95	0.02	0.91	0.003	0.95	-0.52	0.49	<0.001

CI Sidak-corrected confidence interval for mean differences, CL cognitive load, ME mental effort, LL lower limit, UL upper limit

* $p < 0.05$; Estimated means with prior Minecraft knowledge = 1.945

$\alpha = 0.05$). Subsequently, H5 can be supported for cognitive load efficiency, not for mental load efficiency.

Discussion

The implementation of ITI did increase the overall time of interactions between the participants, although only on a descriptive level. However, the manipulation increased social interactions significantly. Furthermore, the ITI groups outperformed the VC groups showing a large effects size. However, presumably because of the low number of groups, only a tendency could be discovered. In contrast, the learning tests revealed a more detailed picture. The implementation of ITI did increase learning in both schemata elaborateness and cloze tests, and we used mediator and moderator analysis to further enrich these results. A full mediation by performance could be observed in the cloze test, and a negative moderation by performance could be measured in the schemata elaborateness tasks. In the analysis of cognitive load that followed, we observed no difference, and no major difference should have occurred, bearing in mind the test of the null hypothesis. In contrast, the VC condition members reported significantly lower mental effort than the participants in the ITI groups. Furthermore, the combination of cognitive and learning measures revealed an increased efficiency in cognitive load and both learning outcomes within the ITI condition, and no difference in mental effort. Again, no large differences should occur, as investigated by our null hypothesis test.

Analyses of the results of interaction times and explanation scores, showed further evidence of an explanation of collaborative effects in learning settings, through social, speech-based interactions. The results of learning and performance show our successful advancement of the jigsaw strategy, and one of Slavins' (1980) categories in educational videogames. Because the players were forced to collaborate, their group task performance increased. This, and their increased interactions, subsequently fostered individual learning outcomes. Our analysis that followed further illuminated the interactions between these factors. The full mediation for the comparably simple cloze scores by performance indicated a simple relationship. As the players achieved more, they also processed more text, and thus internalized more words.

The negative moderation by "performance" on the impact of our group variation on schemata elaborateness indicated a more complex scenario. Although the ITI resulted in an overall positive effect, this impact seemed to be reduced as performance increased, or the negative effect of the VC was reduced. This could be explained by the processing of the content. Because players in the ITI condition could not work independently, they had to work on elements of the material they might not have intended to, because someone else needed their assistance. As the overall performance increased, and more parts of the building tasks were finished (since everybody worked on every part), this positive effect was reduced.

The analysis of mental effort and cognitive load that followed helped to understand the effects of our manipulation more deeply. Since we could rule out any large differences regarding cognitive load, and were able to measure a lowered mental effort within the VC condition, we could now discriminate between those two concepts in a collaborative setting. We could not detect the CWME using our measures. In contrast, the participants needed to invest more effort while interacting with their peers, thus leading to a higher mental effort, as predicted by theory. The results could indicate that the cognitive load

measures that were focused on the specific task (building something) neither benefited from nor were harmed by our manipulation, whereas the mental effort question addressing subjective efforts (which were induced through cooperating), revealing a significant difference. Furthermore, because the CWME might have occurred in both groups, without a control group with no cooperation we cannot detect its effects. Finally, just because something was elaborate (i.e., mental effort), it did not necessarily have to be more difficult (intrinsic or germane load), inconvenient (extraneous load), or complex (intrinsic or germane load).

After we combined cognitive and learning variables into the efficiency construct, we could derive even more insights that can further describe these concepts. In contrast to the increased mental effort through ITI, we could not find significant differences in mental effort efficiency; by our null hypothesis tests, we can assume that no major differences should occur. In contrast to the negative impact of ITI on mental effort, we can record that this increased mental effort did not particularly harm the learning outcomes, thus shedding a more positive light on the increased mental effort in our experiment. Because we found a positive impact of ITI on cognitive load efficiency, we can also assume that collaboration did not affect the cognitive load level, but rather how it was used to enable learning.

This is an important addition to the CWME effect, as it indicates that participants could depend on their peers to manage the load more efficiently. For example, players could leave one task to other players and focus on one specific task at a time, thus increasing the quality of this specific task. As such, they might not report a decreased load, but would increase the quality of their learning schemata one at a time. While this behavior could also occur within the VC group, the ITI may have fostered its frequency; and since other peers relied on the tasks, they may have helped to judge whether or not a task had been completed sufficiently in order to prevent overly exact work from being done. This might have served as an additional feedback mechanism. Players might also criticize others' insufficient work. As a result, the players in the ITI condition may have fostered efficient working and, subsequently, learning.

Implications

Our main practical implication can be drawn from the positive effects of our manipulation on learning and group performance, and the effects on player interaction. We did manage to create an ITI without disturbing playing or cognition, and subsequently showed a promising method for increasing collaboration in educational videogames: a simple mechanism that might also be applied to non-digital games. Furthermore, this highlights the potential for different digital applications such as cMOOCs (Massive open online courses focusing on creation and cooperation). On the theoretical side, we transferred an educational principle that was not invented with games in mind to a new pedagogical medium. This is important, since other studies have highlighted the difficulties of crossing that gap, or discovering diverse impacts within new settings.

In addition, because of our analysis of learning, group performance, and mental demands, we can now provide deeper insights into collaboration's effects on these factors. We not only showed that the introduction of ITI works, but also *how* it works. We have successfully demonstrated that by enhancing group performance with ITI, one could enhance simple learning outcomes. By analyzing complex processes such as building orientation schema, we also discovered that although the overall load might not be

affected, the efficiency might differ, which influences the learning outcomes. Thus, we provide another interesting approach to CWME theory by indicating a further mechanism for collaboratively optimizing cognitive load within ITI tasks; we have also provided evidence that an increased mental effort might not be harmful in an ITI task, thus indicating a balance of mental demands and benefits that were induced by our manipulation. This should be carefully analyzed to provide optimal outcomes.

Limitations

There is limited comparability when addressing cognitive effects while discussing collaboration, because of the various forms of collaboration (Dillenbourg 1999) and intrapersonal factors (Janssen et al. 2010; Peterson and Janicki 1979). For example, limited prior experience with cooperating dampens any positive effects of small group learning (D. W. Johnson et al. 1998; Lou et al. 2001). In addition, since we gathered the cognitive load as an overall score, it is impossible to draw any implications on the specific processes (Janssen et al. 2010), although they might affect learning and cooperation differently. Furthermore, the measurement of cognitive load itself remains a controversial topic within the scientific community and different measurements might have led to other insights. In addition, deeper analyses are required to determine the nature of increased mental effort (Kirschner et al. 2011b). Additionally, a multi-level analysis might be more appropriate for future studies, to identify effects within the group or the individual more precisely.

Further limiting factors within our analysis are the nature of the moderator and mediator analysis, which depended on very sensible regressions that were influenced by small changes within the test groups. Thus, there is the need for reevaluations with more participants, especially since our schemata elaborateness score might have shown a few “floor” effects. Additionally, correlative analysis are not casual analysis and should always be interpreted with caution.

Certain experimental conditions also limited our results. For example, some participants put more effort into decorating the house than in completing the task; also, *Minecraft* offers a lot of room for creativity, which led some experienced players to try to craft their own materials. We noticed this happening and instead encouraged them to focus on the given task; in the interests of not delaying the building process, we told them that crafting their own materials was not necessary. The students in the ITI condition also tended to share their materials by dropping them on certain locations in the play area so that other group members would have direct access to the dropped items. We did not intend for this exchange to happen. Players were then able to build, for instance, parts of the garden without the assistance of the other group members. They did have to cooperate to obtain the other materials, however, so we did not prevent this behavior. Additionally, this aspect of searching the most efficient solution within a collaborative task (e.g. desire lines; Myhill 2004), could be an important section for future research. Especially within the huge potentials but often complicated tasks within virtual environments.

Future directions

When discussing our theoretical basis, we should bear in mind that explanations are only one form of beneficial interaction in cooperative settings (others include, for example: elaborating, asking thought-provoking questions, argumentation, King 2008; which have

not been addressed in the analysis); different interactions thus should be recorded as well. Additionally, as no detailed recordings regarding the exchange of information or the participants' motivation were made, this experiment leaves some room for future research. We can draw a similar conclusion for the CWME, which requires further process data to be analyzed in detail. Additionally, further research is needed to explore the complex relationships between effort, efficiency and cognitive load. Regarding our assumptions on how peers might have influenced the learning process, further analyses and experiments are needed. These could benefit from the theories of co-regulation and socially shared regulation (Hadwin, Järvelä, & Miller, 2011). In addition, besides using ITI within different scenarios, the other categories of Slavin's approach, such as reward structure and authority structure (Slavin 1980), should also be manipulated within educational videogames. Furthermore, additional ways to use educational videogames to transfer literacy knowledge could be explored.

Finally, the students' reactions to participating in this study were mostly positive—especially among the male participants—while several female students were originally rather unenthusiastic. After the implementation, we received only positive feedback about our study, both from males and females. The students thanked us, were interested in the ideal constructed property, and mentioned that they gladly would have played longer. The students' positive attitude might be particularly important to consider, bearing in mind that the goal of reading lessons goes beyond teaching the ability to read: it also strives to understand and engage students within the topic. As mentioned within the aims of this experiment: literacy teaching beyond simply knowing the text. Having fun with reading, understanding, imagining, and using creativity, we demonstrated a simple but fruitful approach to create elaborated schema of descriptions within literature and a simple way to optimize this. Additionally, with traditional measures like the cloze test, we could demonstrate that besides enjoyment and creativity, the knowledge of the underlying literature itself was promoted.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- AbuSeileek, A. F. (2012). The effect of computer-assisted cooperative learning methods and group size on the EFL learners' achievement in communication skills. *Computers & Education*, 58(1), 231–239.
- Aronson, E. (1997). *The jigsaw classroom: Building cooperation in the classroom*. Glenview: Scott Foresman & Company.
- Azadegan, A., & Hartevelde, C. (2014). Work for or against players: On the use of collaboration engineering for collaborative games. *Foundations of Digital Games*.
- 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(6), 1173.
- Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie *PISA 2000* (pp. 15–68): Springer: New York.
- Bopp, M. (2006). Didactic analysis of digital games and game-based learning. In M. Pivec (Ed.), *Affective and emotional aspects of human-computer interaction: Game-based and innovative learning approaches* (Vol. 1, pp. 8–37). Amsterdam: IOS Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates, Inc.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. London: Routledge.

- Deiglmayr, A., & Schalk, L. (2015). Weak versus strong knowledge interdependence: A comparison of two rationales for distributing information among learners in collaborative learning settings. *Learning and Instruction, 40*, 69–78. doi:10.1016/j.learninstruc.2015.08.003.
- Dillenbourg, P. (1999). What do you mean by collaborative learning? *Collaborative-learning: Cognitive and Computational Approaches, 1*, 1–19.
- Dillenbourg, P., Baker, M. J., Blaye, A., & O'Malley, C. (1995). The evolution of research on collaborative learning. *Learning in Humans and Machine*, 189–211
- Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review, 22*(1), 89–102.
- Eysink, T. H., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *American Educational Research Journal, 46*, 1107–1149. doi:10.3102/0002831209340235.
- Fontane, T. (2015). Effi Briest: OUP Oxford.
- Fu, F.-L., Wu, Y.-L., & Ho, H.-C. (2009). An investigation of cooperative pedagogic design for knowledge creation in Web-based learning. *Computers & Education, 53*, 550–562. doi:10.1016/j.compedu.2009.01.004.
- Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of self-regulation of learning and performance, 30*, 65–84.
- Hayes, A. F. (2008). INDIRECT. Retrieved from <http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html>
- Huber, G., & Huber, A. (2008). Structuring group interaction to promote thinking and learning during small group learning in high school settings. In R. Gillies, A. Ashman, & J. Terwel (Eds.), *The teacher's role in implementing cooperative learning in the classroom* (Vol. 8, pp. 110–131). New York: Springer.
- Janssen, J., Kirschner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load approaches. *Educational Psychology Review, 22*(2), 139–154.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: The Magazine of Higher Learning, 30*(4), 26–35. doi:10.1080/00091389809602629.
- Johnson, R. T., Johnson, D. W., & Stanne, M. B. (1986). Comparison of computer-assisted cooperative, competitive, and individualistic learning. *American Educational Research Journal, 23*, 382–392. doi:10.3102/00028312023003382.
- Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89*, 47. doi:10.1037/0033-2909.89.1.47.
- Ke, F., & Grabowski, B. (2007). Gameplaying for maths learning: cooperative or not? *British Journal of Educational Technology, 38*, 249–259. doi:10.1111/j.1467-8535.2006.00593.x.
- King, A. (2008). Structuring peer interaction to promote higher-order thinking and complex learning in cooperating groups. In R. Gillies, A. Ashman, & J. Terwel (Eds.), *The teacher's role in implementing cooperative learning in the classroom* (Vol. 8, pp. 73–91). New York: Springer.
- Kirschner, P. A., Beers, P. J., Boshuizen, H. P., & Gijsselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior, 24*(2), 403–420.
- Kirschner, F., Paas, F., & Kirschner, P. A. (2011a). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology, 25*(4), 615–624.
- Kirschner, F., Paas, F., Kirschner, P. A., & Janssen, J. (2011b). Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learning and Instruction, 21*(4), 587–599.
- Kooloos, J. G., Klaassen, T., Vereijken, M., Van Kuppeveld, S., Bolhuis, S., & Vorstenbosch, M. (2011). Collaborative group work: Effects of group size and assignment structure on learning gain, student satisfaction and perceived participation. *Medical Teacher, 33*(12), 983–988.
- Krell, M. (2015). Evaluating an instrument to measure mental load and mental effort using Item Response Theory. *Science Education Review Letters, 2015*
- Leemkuil, H., De Jong, T., De Hoog, R., & Christoph, N. (2003). KM QUEST: A collaborative Internet-based simulation game. *Simulation & Gaming, 34*(1), 89–111.
- Lohman, M., & Finkelstein, M. (2000). Designing groups in problem-based learning to promote problem-solving skill and self-directedness. *Instructional Science, 28*(4), 291–307. doi:10.1023/A:1003927228005.
- Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*(3), 449–521.
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423–458.

- Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. *Ergebnisse und Anwendungen der Psychophysiologie. Enzyklopädie der Psychologie*, 100, 799–864.
- MINECRAFT. (2011). [Developed by Mojang]: Mojang.
- MINECRAFTEDU. (2013). [Developed by Teacher Gaming].
- Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning*, 6(3), 421–443.
- Myhill, C. (2004). Commercial success by looking for desire lines. In M. Masoodian, S. Jones, & B. Rogers (Eds.), *Computer Human interaction: 6th Asia Pacific Conference, APCHI 2004, Rotorua, New Zealand, June 29–July 2, 2004. Proceedings* (pp. 293–304). Springer Heidelberg.
- Nebel, S., Schneider, S., & Rey, G. D. (2016). Mining learning and crafting scientific experiments: a literature review on the use of minecraft in education and research. *Journal of Educational Technology & Society*, 19(2), 355–366.
- Nickel-Bacon, I. (2006). Positionen der Literaturdidaktik–Methoden des Literaturunterrichts. Ein heuristischer Explikationsversuch für die empirische Grundlagenforschung. *Empirische Unterrichtsforschung in der Literatur und Lesedidaktik. Ein Weiterbildungsprogramm*.
- Oksanen, K. (2013). Subjective experience and sociability in a collaborative serious game. *Simulation & Gaming*, 1046878113513079.
- Peppler, K., Danish, J. A., & Phelps, D. (2013). Collaborative gaming: Teaching children about complex systems and collective behavior. *Simulation & Gaming*, 1046878113501462.
- Peterson, P. L., & Janicki, T. C. (1979). Individual characteristics and children's learning in large-group and small-group approaches. *Journal of Educational Psychology*, 71(5), 677.
- Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., et al. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of Educational Psychology*, 105, 1050–1066. doi:10.1037/a0032688.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422.
- Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research*, 50(2), 315–342.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21–51.
- Staiano, A. E., Abraham, A. A., & Calvert, S. L. (2012). Competitive versus cooperative exergame play for African American adolescents' executive function skills: short-term effects in a long-term training intervention. *Developmental Psychology*, 48, 337. doi:10.1037/a0026938.
- Stevens, R. (2008). Cooperative Learning and Literacy Instruction in Middle Level Education. In R. Gillies, A. Ashman, & J. Terwel (Eds.), *The teacher's role in implementing cooperative learning in the classroom* (Vol. 8, pp. 92–109). New York: Springer.
- Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63, 43–51.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285. doi:10.1016/0364-0213(88)90023-7.
- Sweller, J., van Merriënboer, J. G., & Paas, F. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296. doi:10.1023/A:1022193728205.
- Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. *Computers as Cognitive Tools*, 229–258.
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, 43(1), 16–26. doi:10.1080/00461520701756248.
- Warner, R. M. (2008). *Applied statistics: From bivariate through multivariate techniques*. Thousand Oaks, CA: Sage.
- WORLD OF WARCRAFT. (2004). [Developed by Blizzard Entertainment]: Vivendi & Activision Blizzard.

Steve Nebel is Doctoral student at the chair of E-Learning and New Media, Faculty of Humanities, Technische Universität Chemnitz, Germany. M.A. in the field of media communication.

Sascha Schneider is Doctoral student at the chair of E-Learning and New Media, Faculty of Humanities, Technische Universität Chemnitz, Germany. M.A. in the field of educational technology and further education.

Maik Beege is Doctoral student at the chair of E-Learning and New Media, Faculty of Humanities, Technische Universität Chemnitz, Germany. M.Sc. in the field of psychology.

Franziska Kolda is Undergraduate student of media communication at the Technische Universität Chemnitz, Germany.

Valerie Mackiewicz is Undergraduate student of media communication at the Technische Universität Chemnitz, Germany.

Günter Daniel Rey is Head of chair E-Learning and New Media, Faculty of Humanities, Technische Universität Chemnitz, Germany.