Ruipérez-Valiente, J. A. & Kim, Y. J. (2020). Effects of Solo vs. Collaborative Play in a Digital Learning Game on Geometry: Results from a K12 Experiment. Computers & Education, 159, 104008.

https://www.sciencedirect.com/science/article/pii/S0360131520302062

DOI: 10.1016/j.compedu.2020.104008

Effects of Solo vs. Collaborative Play in a Digital Learning Game on Geometry: Results from a K12 Experiment

José A Ruipérez-Valiente[a,b,*], Yoon Jeon Kim[a]

[a] *Massachusetts Institute of Technology, Playful Journey Lab, Cambridge (MA, USA)*
[b] *University of Murcia, Faculty of Computer Science, Murcia (Spain)*

Abstract

Digital games for learning are one of the most prominent examples of the use of technologies in the classroom, where numerous studies have presented promising results among children and adolescents. However, scarce evidence exists regarding different ways of implementing games within the classroom and how those affect students' learning and behaviors. In this study we explore the effect that collaboration can have in digital gameplay in a K12 context. More specifically, we have designed a 2x2 experimental study in which high school first year students participated in solo or collaborative gameplay in pairs, solving puzzles of diverse difficulty, using *Shadowspect*, a digital game on geometry. Our main results, computed by applying learning analytics on the trace data results, suggest that students playing solo had higher in-game engagement and solved more puzzles, while students collaborating were less linear in their pathways, skipping more tutorial levels and were more exploratory with *Shadowspect* features. These significant differences that we observe in solo and collaborative gameplay call for more experimentation around the effect of having K12 students collaborate on digital tasks, so that teachers can take better decisions about how to implement these practices in the classrooms of the future.

Keywords: Games for learning, collaborative task solving, learning analytics,

*Corresponding author
Email addresses: `jruiperez@um.es` (José A Ruipérez-Valiente), `yjk7@mit.edu` (Yoon Jeon Kim)

K12 education

2010 MSC: 00-01, 99-00

1. Introduction

Around the world, educational technology is being introduced slowly into the classroom and holds the potential to have a transformative effect on the educational landscape Baumöl & Bockshecker (2017). One of the most prominent examples is the use of digital games for learning De Freitas (2006). Numerous authors Prensky (2003); Shaffer (2006); Gee (2003) have expressed that well designed digital games represent outstanding opportunities for learning through an enjoyable experience. Playing video games is essential part of young generation's daily life Buckingham & Willett (2013). A recent report on media consumption Ofcom (2019), builds profile snapshots by age, estimating that 40% of kids 3-4 years old play games for nearly 6 hours a week, 66% of kids 5-7 for 9 hours a week, 81% of kids 8-11 for around 10 hours a week and 77% of kids 12-15 for 12 hours a week. Many studies provide ample evidence that using video games with children and adolescents can support various skills and life outcomes such as academic domain-specific learning including science Li (2013); Council et al. (2011) and math Divjak & Tomić (2011); Starkey (2013), executive functions Homer et al. (2018), involvement in real life sports Adachi & Willoughby (2016) or for acquiring health knowledge Baranowski et al. (2016).

While many teachers report a positive attitude towards games being used in K12 classrooms and believe that they can improve learning and curriculum, the actual number of teachers who are implementing digital games in their curriculum is contrarily low De Grove et al. (2012); Pastore & Falvo (2010). One of the main implementation barriers reported by teachers is uncertainty and limited literature about how to actually effectively implement games in the classroom Watson & Yang (2016); An & Cao (2017), and thus there is a necessity to support evidence-based decision making regarding how different teachers' decisions impact students' experiences when implementing games in classrooms.

One strong attraction for teachers to use games in classroom is because of the belief that it is a powerful tool to foster collaboration among students and develop 21st Century skills such as collaborative problem solving Kim & Shute (2015). In this experimental case study, we explore what the effect of having students play a game solo vs. with others is, with the ultimate goal of supporting teachers' implementation decisions for game-based curriculum. A few studies compared solo vs. social play in commercial gaming contexts Arellano et al. (2016); Kaye & Bryce (2014), but little work is done in educational games, especially using in-game analytics. Given that learning games are frequently introduced as classroom activities Squire (2005), the existence of barriers to effectively implementing game-based learning Watson & Yang (2016); An & Cao (2017) and the importance of collaboration to prepare younger generations for the future UNESCO Bangkok (2016), educators would have more confidence in using games in classrooms if they knew what to expect from having students play these games alone or collaboratively and the potential implications for their learning.

We organize the rest of the manuscript in the following sections. Next Section 2 reviews related literature and Section 3 presents the current case study. Section 4 details the methods of the study, including an overview of *Shadowspect*, the context, experimental design, data and metrics. Section 5 presents the results and in Section 6 we discuss those results. Finally, in Section 7 we finalize with conclusions and future follow-up ideas.

2. Literature Review

2.1. Collaboration in Game-Based Learning

There are two related yet subtly different concepts for collaboration in the context of game-based learning. First, *collaborative game*, refers to a specific type of game that has explicit game mechanics that require two or more players to work on joint tasks or quests Berland & Lee (2011). Second, *collaborative gameplay*, is a broader category that describes non-competitive gameplay where two or more players socially engaging while playing together with a common objective in mind. Note that this can involve various forms; playing a collaborative game, collaborating with each other while playing a single-player game, or sharing a single device while playing a single-player game.

Numerous authors have brought up the potential of collaborative gameplay Voulgari & Komis (2008); Marfisi-Schottman & George (2014); Wendel et al. (2010, 2012); Papastergiou (2008). For example, collaborative learning can help students prepare for future job skills Hamalainen (2008), to improve the flow state through collaboration in game-based learning Admiraal et al. (2011) or facilitate peer learning by sharing skills or knowledge otherwise not available if they would be playing alone Wendel et al. (2010). Some of these characteristics have been attributed to commercial Massively Multiplayer Online Games as well Voulgari & Komis (2008).

There is mixed evidence, however, regarding how different choices, i.e., solo vs. collaborative, would influence the learners' experience and ultimately what they learn from the experience. In general, collaborative gameplay have positive effects in terms of learners' experience and affective state. For instance, a study (N = 302) performed a retrospective rating of the flow experience and post-gameplay mood on solo and collaborative game play experiences through a questionnaire, and found a significantly higher positive mood in social gameplay Kaye & Bryce (2014). Another study explored differences in physiological signals, measured via heart rate, in solo, competitive and collaborative games, and reported that while competition evoked tense behaviors in players, collaboration generated more relaxed and positive situations Arellano et al. (2016). However, simply playing together does not mean that players will productively collaborate. For example, Shih and colleagues conducted a cases study with the digital problem-solving game *William Adventure*, where they observed that the results of collaborative activities were importantly dependent on collaboration strategies and models, but in any case were positive from a cognitive viewpoint Shih et al. (2010).

Few studies also investigated the effect of collaborative gameplay on learning.

Ke and colleagues compared (N = 120) 5th graders playing a math game solo vs. collaborative, and found no impact of this social aspect on learning outcomesKe & Grabowski (2007). Similarly, another study (N = 58) compared middle school students playing the arithmetic game *FactorReactor* in three modes: solo, competitive, and collaborative. The findings suggested that both social modes, competitive and collaborative, increased the situational interest more than solo play, but it was the competition mode that enabled more in-game learning Plass et al. (2013).

In summary, the interplay between these social possibilities for game-based learning requires further investigation to provide better insights in terms of how these choices influence learners' behaviors in the game as well as learning outcomes. It remains unclear why educators should encourage collaborative gameplay and how this choice could influence players' experience and learning. In addition, previous studies largely relied on external measures (e.g., questionnaires or post-test) rather than using in-game behavioral measures, therefore, offer a limited view regarding how learner engaged with the game.

2.2. Interplay of Collaboration with Difficulty

While collaboration can be a means to promote active learning when the learner finds the task difficult van Drie et al. (2005), task difficulty also affects the effectiveness of collaborative learning. For example, in a study where five-year-olds were told to complete easy or hard puzzles, either alone or with a partner Arterberry et al. (2007); as an additional condition, half the children were told that their work would be graded and the other half were not. The results showed that when interacting with easy puzzles, performance was better working with a partner than alone, and when working under the evaluation condition. The role of difficulty is of particular interest to educational games. For example, Gee (2010) argues that players learn best in games that offer properly organized problems that push them toward the outer limits of what Gee calls their "region of competence." When students play a game in classrooms, it is often natural for more competent students to share their tips or to observe students seeking help from others. Given difficulty is a key game element that affects players' experience and performance, we were interested in investigating how difficulty of the game levels would interact with collaborative gameplay.

3. Current Study

To address the gap identified in the literature review, we conducted a 2x2 factorial experiment to investigate the influence of gameplay style (solo or collaborative gameplay) when using the geometry game *Shadowspect*, while also varying puzzle complexity with the second factor. The gameplay style condition entangles either having students play solo or collaboratively in pairs. The puzzle complexity condition involves one set of puzzle levels slightly more difficult than the other. All conditions include a set of tutorial levels that are the same to establish a common baseline ability. Our initial hypothesis is that the social learning experience in the collaborative gameplay could lead to additional ways to interact and explore a digital game Bressler et al. (2018). More precisely, we hypothesized that learners would interact with more features of the game when playing collaboratively, but would also demonstrate more off-task behaviors due to this exploration. Moreover, we expect an interplay between the collaborative gameplay style and task difficulty, since the added coordination between the dyad members could be a handicap to solving a simple puzzle, while having two pairs of eyes looking at how to solve a complex puzzle could be useful, when compared to individual work in easy and complex tasks respectively.

In order to answer our overarching question regarding the effect of individual vs. collaborative gameplay in how students engaged with the game, we will follow a data-driven approach by implementing game learning analytics Freire et al. (2016) using the trace data that learners generated when interacting with *Shadowspect* in each experimental condition. The concept of engagement that we measure is related to the degree of activity or attention someone gives to certain tasks over some period of time Martey et al. (2014); Ruiperez-Valiente et al. (2020), which can be linked to their intrinsic interest, curiosity, and motivation Chapman (1997). We establish the following research question (RQ) to answer through this experiment:

RQ What is the effect of solo vs. collaborative gameplay in terms of:

1. Tutorial and puzzle level completion.
2. Level pathways.
3. In-game engagement metrics.

4. Methods

4.1. Shadowspect Overview

Shadowspect[1] is a 3D digital educational game. *Shadowspect* lies within the category of puzzle games (like the *Witness* or *Bridge Builder*). *Shadowspect* has clearly-defined goals, rules, obstacles for the players to overcome and provides only intrinsic rewards (satisfaction for getting the right answer). Each puzzle presents a number of silhouette views for a 3D figure where each figure is built by using a series of 3D geometric primitives such as cubes, cylinders, spheres, cones or pyramids. The player then selects a set of geometric primitives to recreate a 3D figure by using the 3D game environment.

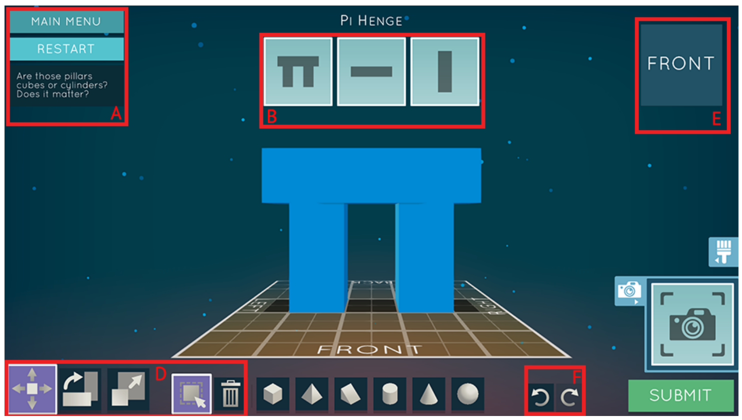
Figure 1 shows two example puzzles in *Shadowspect*, with some zones delimited by red parallelograms and a letter to facilitate the following description. When students enter a puzzle, they see a description (A) and they receive a set of silhouettes (B) from different views that represent the figure they need to build. Students can create (C) cubes, pyramids, ramps, cylinders, cones and spheres. Additionally, some puzzles might set up constraints, such as using a maximum number of objects, or a maximum number of shapes of each type. Learners can use several tools (D) to achieve in-game goals, moving, rotating, and scaling shapes around the stage to match the silhouettes provided. Additionally, they can delete and select multiple shapes at the same time. Students change the

[1] Playable version online at `https://shadowspect.org/`

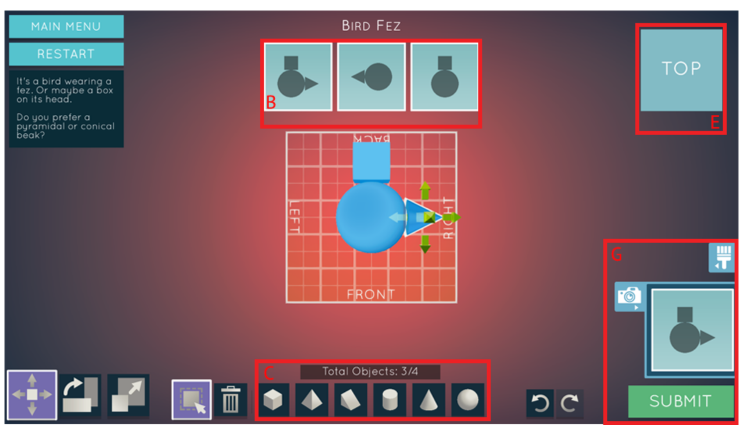
camera view (E) to see the figure they are building from different perspectives and then use the ‘Snapshot’ functionality (G) to generate the silhouette from the current view. Snapshots can help them know if their shapes match any of the solution silhouettes. Finally, they can submit (G) their current shapes, and the system will evaluate if their solution is correct and provide them with feedback.

Shadowspect allows players to build 3D figures, developing their geometric, dimensional, and spatial reasoning skills while having an enjoyable experience. In addition, *Shadowspect* is designed explicitly as a formative assessment tool to measure math content standards, thus teachers can use it in their core math curriculum. In addition to the content standards, the game also measures cognitive and noncognitive skills such as spatial reasoning, creativity, and persistence. During the case study, students interacted with a *Shadowspect* version which had 9 tutorial levels and 12 puzzle levels. These levels have been implemented by the game designer in collaboration with a geometry expert to map the levels with the following four Geometry Common Core State Standards:

- **MG.1**: Use geometric shapes, their measures, and their properties to describe objects (e.g., modeling a tree trunk or a human torso as a cylinder).
- **GMD.4**: Identify the shapes of two-dimensional cross-sections of three-dimensional objects, and identify three-dimensional objects generated by rotations of two-dimensional objects.
- **CO.5**: Given a geometric figure and a rotation, reflection, or translation, draw the transformed figure using, e.g., graph paper, tracing paper, or geometry software. Specify a sequence of transformations that will carry a given figure onto another.
- **CO.6**: Use geometric descriptions of rigid motions to transform figures and to predict the effect of a given rigid motion on a given figure; given two figures, use the definition of congruence in terms of rigid motions to decide if they are congruent.



(a)



(b)

Figure 1: Examples of two puzzles in *Shadowspect* with red parallelograms delimiting different game functionalities.

Factors		**Play**	
		Solo	*Collaborative*
Difficulty	*Easy*	`SoloEasy` (n=31, n_{5min}=26)	`CollEasy` (n=16, n_{5min}=12)
	Hard	`SoloHard` (n=22, n_{5min}=19)	`CollHard` (n=21, n_{5min}=17)

Table 1: The four conditions of the experiment and the associated label where n indicates the number of students that engaged with *Shadowspect* within the condition, and n_{5min} the number of students that were active within the game for at least 5 minutes.

In addition to the puzzle mode, *Shadowspect* also includes a sandbox mode, which does not have any particular objective or finishing condition besides letting students build whatever they want, similar to *Minecraft* or *TinkerCAD*.

4.2. Context and Experimental Design

The experiment took place during a two-day event at a high school in Massachusetts. This event sought to expose students to diverse technologies and activities that could encourage them to develop future-ready skills and educational opportunities in STEM fields like computer science. All of the ninth graders were divided in groups of around 15-20 students by randomly assigning them to a group. Eight of these groups participated in this experiment.

We designed a full factorial experiment with two factors where each factor has two levels Mee (2009), hence we have a 2x2 factorial design with a total of four conditions. The operationalization of the experiment was cohort-based, since the assignment of students to groups was randomized, which should avoid potential cohort bias. We will also use the labels `Solo` or `Collaborative` for all cohorts that played under that condition, no matter the difficulty condition.

Each session was 75 minutes long, and we had a total of eight sessions. Therefore, we administered each one of the conditions to two sessions total. In each one of the sessions, at least one researcher was present to guarantee that the method was administered properly, and there were also one or two proctors from the school helping to facilitate the sessions smoothly. Each session was organized as follows:

- **Presentation of *Shadowspect*** (5 minutes). We explained to the students what *Shadowspect* is and what we were going to do during the session. Then we went through one tutorial level explaining the controls of the game and objectives, so that the learning curve was less abrupt. All of this was conducted through the reading of a script, to guarantee that we provided the same information in all sessions. The only difference was that in the `Solo` condition, we encouraged students to play alone, while in the `Collaborative` condition we encouraged them to play in pairs. The desks in the classroom in the `Solo` condition were arranged apart from each other, while in the `Collaborative` condition desks were arranged in 4-person workstations. Students sat in these desks following the order of arrival to the session without intervention from the researchers or proctors.

- **Interaction with *Shadowspect*** (50 minutes). During this time, students were free to interact with the game. We encouraged them to start interacting with the tutorial levels, and to finish them before moving to the puzzle levels, but we did not force them to do so. The tutorial encompassed a series of levels designed for a first-time player to learn all of the functionalities and concepts of *Shadowspect* in a scaffolded succession of levels. The puzzle set was comprised of standalone levels that assume that the student knows the different features of the game. It is important to point out again that in all conditions the set of tutorial levels was the same, after which there was a difference of difficulty between the puzzle sets of the `Easy` and `Hard` conditions. The difficulty of levels was assessed by the game designer of *Shadowspect* and empirical data obtained through playtesting. Finally, the instructors in the session helped students to understand how to play *Shadowspect* but did not solve puzzles for students.

- **Stop playing and career talk** (15 minutes). We asked students to stop playing and we helped them connect the different technologies and applications of *Shadowspect* with career choices in the areas of computer science and technology. We took some questions from students.

- **Fill out reflections** (5 minutes). Before finishing the sessions, students filled out a short survey about how they felt about the activity.

4.3. Data Collection

Because *Shadowspect* was designed as a game-based assessment system, any interactions that students had with the game are automatically stored as detailed data. This also allowed us to reconstruct the gameplay process that students underwent to solve each puzzle. There is a hard cutoff of data based on the time when students were supposed to stop playing (see previous Subsection 4.2), as some motivated students kept playing while the instructor was talking or played at home after the session.

We did not keep track of the exact number of students that were in the classroom within each session, but instead we used the game data to count them. It was also the case that a small number of students decided not to engage with the activity. Because their participation was entirely voluntary, the research team did not force them play the game. As part of the initial data cleaning and pre-processing, we established a minimum of 5 active minutes (i.e., `active_time` is defined in next Subsection) interacting with *Shadowspect* to include that user/dyad into the final analysis. Table 1 shows the different conditions and labels assigned to each one of them, with n indicating the number of students that interacted with Shadowspect within that condition, and n_{5min}, which is the number of students that were active for at least 5 minutes with Shadowspect within that condition. Then, the final numbers of the experiment included 26 and 19 individual students for the `SoloEasy` and `SoloHard` conditions respectively, and 12 and 17 dyads for the `CollEasy` and `CollHard` conditions respectively. In global numbers, these users triggered more than 60,000 in-game events, used the game for a total of 41 hours, and solved almost 600 levels.

4.4. Metrics

The process and steps that we followed to design and select the in-game metrics to evaluate the experiment were as follows:

1. Formation of a team between *Shadowspect* designers, learning analytics researcher and assessment scientist, that had been involved in the development of the game and its mechanics.
2. Team brainstorm to design metrics that can respond to the research questions established as part of the experiment.
3. Technical implementation of metrics through applying data mining techniques to the game trace data.

We use the following metrics to compare the effect of the different conditions in how students engaged with the game:

- `n_events`: Total number of events triggered within the game (every action performed by a student in *Shadowspect* is recorded as an event).
- `active_time`: Amount of active time in minutes establishing an inactivity threshold of 60 seconds (i.e., if the time between two events is above 60 seconds, the user is considered to be inactive during that time and that time is omitted from the computation).
- `n_started`: Number of levels that were started.
- `n_completed`: Number of levels that were completed.
- `p_abandon`: Percentage between the number of finished and started levels i.e., $100 * (n\_completed/n\_started)$.
- `avg_time_completed`: Average time per level completed i.e., $active\_time/n\_completed$ in minutes. Denotes how fast they are able to solve levels.
- `avg_actions_time`: Average number of actions performed by a student per minute; higher values indicate higher degrees of fluency with the game.

- `median_start_to_exit`: Represents the median of the number of minutes between accessing a level and exiting a level without being able to solve the level. It is a measure of how long they persist when trying to solve a level.

- `n_different_events`: Total number of different events triggered by the student from the total of 25 different events.

- `n_click_nothing`: We log the clicks performed by students in the screen that did not trigger any functionality (i.e., a click in no particular element or button of the game) and this metric measures the number of those clicks. We do this for two reasons, 1) for potential user interface debugging and more importantly 2) to analyze when students might engage in erratic random clicking around the interface.

- `n_paint_events`: There is a paint functionality that allows students to paint shapes in eight different colors. Although this feature is more oriented towards students playing in sandbox mode, it is still enabled during puzzle mode because we are interested in tracking when students go off the main task of solving the puzzles. This metric measures the number of paint events triggered by a student.

We note that some of these metrics might be computed separately for tutorial and puzzle levels in different sections of the manuscript, with the objective of further deepening into a differential analysis.

5. Results

The results section presents comparison between groups and conditions to respond the research questions. In order to find if a difference on a quantitative variable is statistically significant we report the difference between means of each group, Student's t-Test with its p-value, and the effect size via Cohen's d.

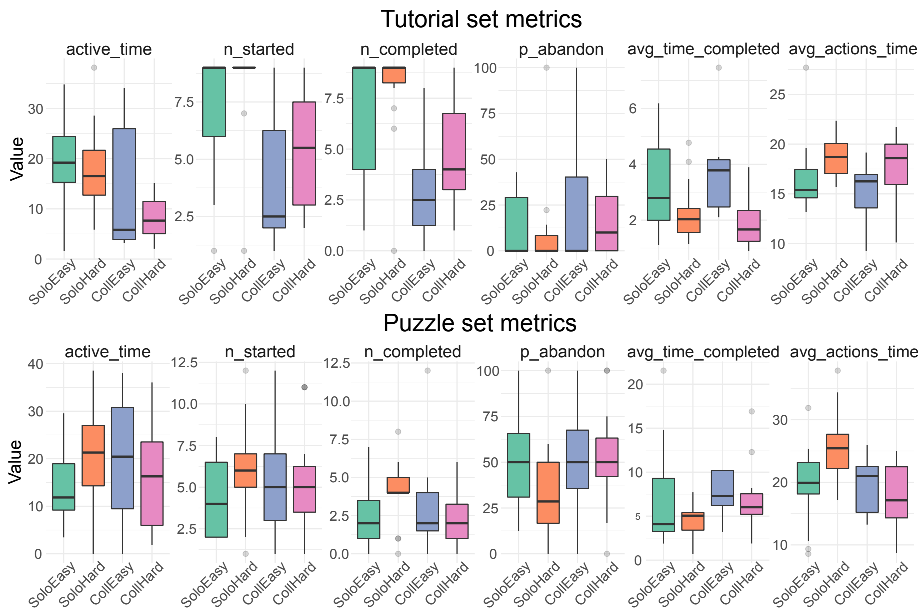


Figure 2: Boxplot visualization with the metrics separated by level type (tutorial on the top and puzzles on the bottom) and condition.

5.1. Overview of Metrics for Each Condition

This first subsection presents results divided by each one of the four conditions in the experiment using a series of in-game metrics. Moreover, in this section these metrics are presented separately for tutorial and puzzle levels due to a twofold rationale: 1) The tutorial puzzles are heavily scaffolded for the purpose of helping players learn how to play the game 2) Since the tutorial levels are exactly the same for all conditions, we can use level completion and fluency metrics in the tutorial to determine the initial game fluency of each cohort.

The set of metrics that we present are as follows: First, `active_time` to measure the overall time invested, then `n_started`, `n_completed` and `p_abandon` to measure the specific interaction with puzzles, and lastly `avg_time_completed` and `avg_actions_time` to infer a level of ability or game fluency. Figure 2 presents a boxplot visualization with all of these metrics separated by condition and type of level.

5.1.1. Initial Gaming Ability

We expected that the `Solo` or `Collaborative` play styles would have a significant effect on how students interact with the game. However, we first needed to investigate if all four groups were equivalent in terms of their initial gaming ability. Therefore, we performed a 1 by 1 comparison for each play mode condition separately to measure each group's gaming ability with this specific game. To perform the inference, we use the `n_completed`, `avg_time_completed` and `avg_actions_time` on the tutorial levels.

First, if we compare `SoloEasy` and `SoloHard` we find that `SoloHard` resolved more tutorial levels with an average of 8.11 compared to 6.7 ($t = 1.7, p = 0.09, d = 0.53$), solved these levels faster with an avg time completed of 2.28 minutes per level compared to 3.2 ($t = 2.4, p = 0.02, d = 0.72$), and performed more actions per minute with an avg actions time of 18.7 compared to 16.33 ($t = 3, p = 0.004, d = 0.91$).

Analogously, once we compare `CollEasy` and `ColdHard` we observe that `ColdHard` resolved more tutorial levels with an average of 4.8 compared to 3.2 ($t = 1.4, p = 0.17, d = 0.6$), solved these levels faster with an `avg_time_completed` of 1.97 minutes per level compared to 3.8 ($t = 3, p = 0.01, d = 1.43$), and performed more actions per minute with an `avg_actions_time` of 17.78 compared to 15.2 ($t = 1.97, p = 0.06, d = 0.84$).

In summary, this indicates that in both `Solo` and `Collaborative` play modes, the t-tests and effect size suggest that the cohort that received the `Hard` condition have a statistically significant higher initial level of game fluency based on how they performed in the tutorial levels.

5.1.2. Effect of the Difficulty Condition

We reported in the previous subsection a statistically significant higher game fluency for both conditions that received the harder set of puzzle levels. In this subsection, we perform a similar comparison, but on the puzzle set metrics, in order to measure if the difficulty condition had any effect.

For the `Solo` condition, when we compare the number of completed puzzle levels, we see that actually SoloHard completed on average 4.1 puzzle levels compared to 2.4 puzzle levels for SoloEasy ($t = 2.3, p = 0.02, d = 0.78$). The number of actions per minute for SoloHard is still higher with an avg_actions_time of 25.6 compared to 19.7 ($t = 3.09, p = 0.004, d = 1.04$), but now avg_time_completed is not significant anymore.

For the Collaborative condition we find that once we compare CollEasy and ColdHard, there is not a statistically significant difference between the means of n_completed, avg_time_completed or avg_actions_time on the puzzle levels.

As a summary, although the creation of cohorts followed a random protocol, we observe that in both cases, the cohorts that received the Hard condition had a significantly higher level of gaming ability based on the performance metrics in the tutorial levels. We think that these differences in gaming ability between groups heavily influenced gameplay such that when analyzing puzzle completion, the SoloHard cohort was able to solve more puzzles than SoloEasy, which is the opposite of the expected outcome, and that there was not an statistically significant difference in the case of CollEasy and ColdHard.

5.2. Effect of Solo vs. Collaborative Gameplay

In this subsection, we focus on the results regarding our research question in terms of effect of the collaboration condition. To answer this issue, we collapse metrics together for both SoloEasy and SoloHard for the Solo condition, and CollEasy and ColdHard for the Collaborative condition.

5.2.1. Interaction with Tutorial and Puzzle Levels

Analogously to the previous visualization in Figure 2, Figure 3 also compares the metrics active_time, n_started, n_completed, p_abandon, avg_time_completed and avg_actions_time, but now between the collapsed Solo and Collaborative conditions.

After collapsing, some patterns become more apparent. For the tutorial level metrics, the Solo cohorts invested a higher active_time of 19 minutes

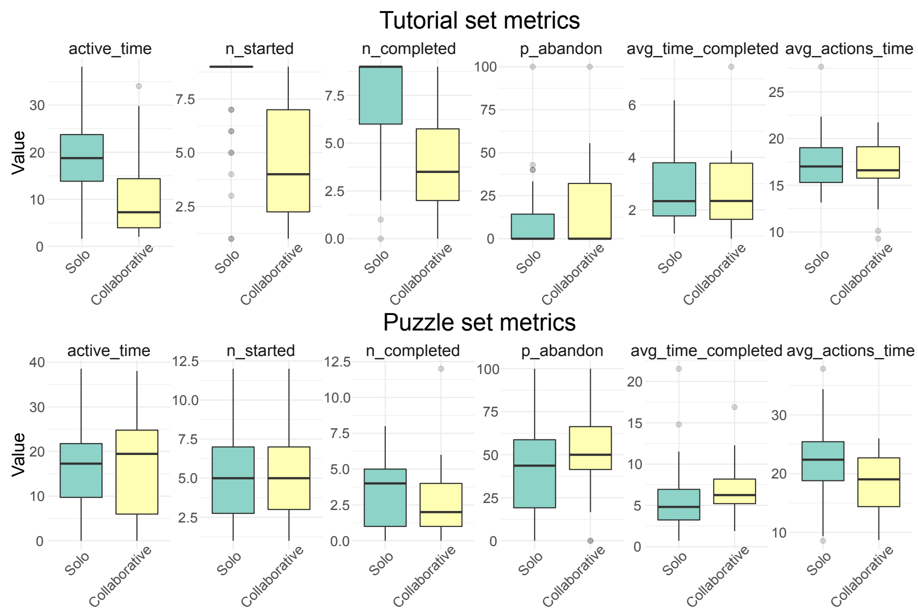


Figure 3: Boxplot with metrics regarding how students engaged with the game by play style condition (solo and collaborative play).

compared to 10 ($t = 3.5, p = 0.001, d = 0.97$), a higher average `n_started` of 7.9 compared to 4.8 ($t = 4.2, p \ll 0.01, d = 1.22$), and a higher average of `n_completed` of 7.3 compared to 4 ($t = 4.4, p \ll 0.01, d = 1.2$). However, there was not a statistical difference in game fluency for the `avg_time_completed` and `avg_actions_time` metrics.

For the puzzle level metrics, the `Solo` cohorts show slightly higher average of `n_completed` with 3.2 compared to 2.7 and `avg_actions_time` with 22.5 compared to 18.5, and lower `p_abandon` with 43% compared to 52% and `avg_time_completed` 5.8 mins per puzzle compared to 7.1, but none of these are statistically significant except for `avg_actions_time` ($t = 2.7, p = 0.008, d = 0.68$).

5.2.2. Level Pathways

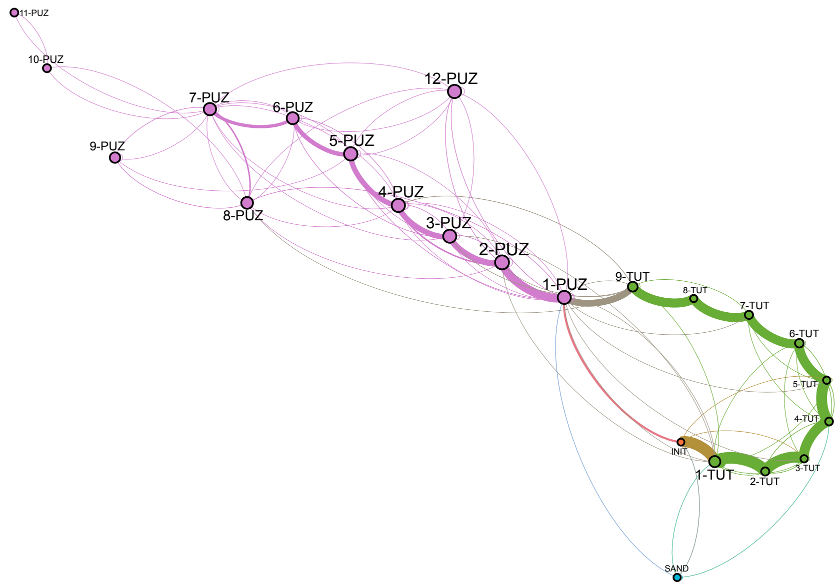
This section applies graph analysis to explore what were the actual level pathways for each play style condition. Figure 4 shows the graphs represent-ing the started levels transitions, with a) for the `Solo` cohort and b) for the `Collaborative` cohort. In both graphs, each node represents a level, with green for tutorials and purple for puzzles. The initial state is called *INIT* in orange, and the sandbox level is called *SAND* in blue. Then, the thickness of the line represents the frequency that students under that condition move between these two levels.

Therefore, the main interpretation of these graphs is quite straightforward. While students in the `Solo` condition moved through *Shadowspect* quite linearly and following the proposed order, starting from the tutorial levels and then moving and advancing through the puzzle levels, this is not the case for the `Collaborative` condition where this linear progress was not the main trend and students moved through the different levels of *Shadowspect* with a much higher entropy. We can formalize this degree of entropy by presenting two standard graph statistics such as the average degree, with is 4.5 for `Solo` and 5.5 for `Collaborative` that implies that `Collaborative` is more connected, or the standard deviation of the edge weights conforming each graph, which would be 1.74 for `Solo` and 1.09 for `Collaborative` which implies that the `Collaborative` graph is more dispersed and the entropy is higher.

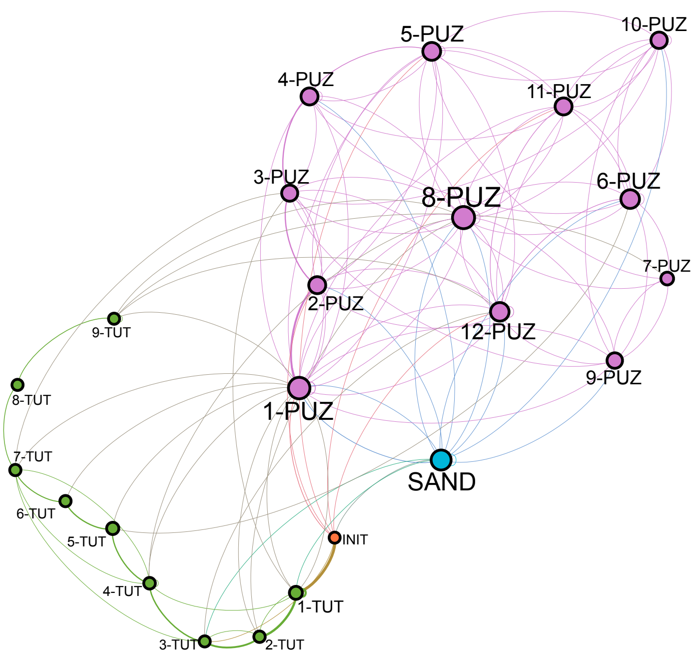
5.2.3. In-game Engagement Metrics

The last subsection in Figure 5 shows results by computing a series of global metrics, instead of dividing by tutorial and puzzle levels. First, `active_time` and `n_events` can provide insights regarding the level of engagement within the game; we can see that students in the `Solo` condition engaged with the game more, with an average `active_time` of 35 minutes compared to 29 ($t = 2.1, p = 0.03, d = 0.5$) and an average `n_events` of 890 events compared to 709 ($t = 2.3, p = 0.02, d = 0.52$). Then, we use `median_start_to_exit` as a measure of persistence in trying to solve a puzzle, and we find lower persistence in the `Collaborative` condition with an average `median_start_to_exit` of 3.2 minutes compared to 4.6 ($t = 1.7, p = 0.08, d = 0.44$).

The metrics `n_different_events` and `n_paint_events` represent if students



(a) Solo play



(b) Collaborative play

Figure 4: Directed graph of started levels by play style.

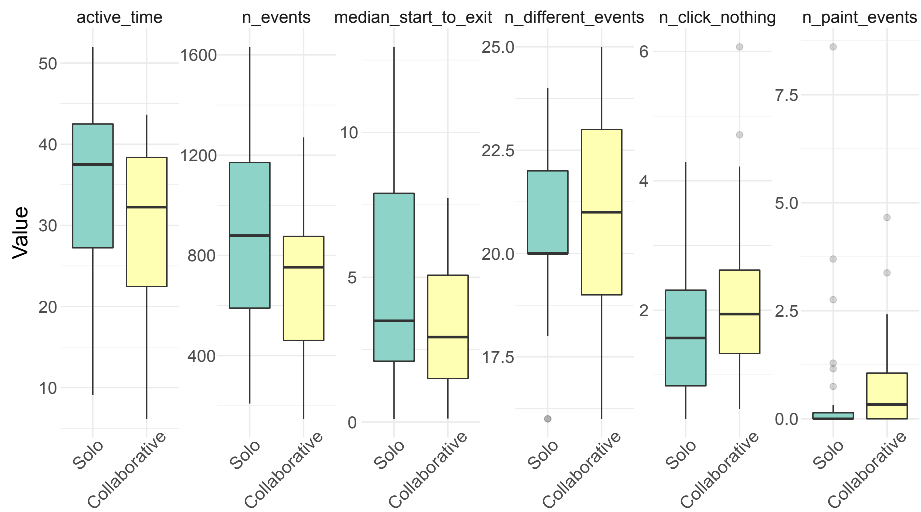


Figure 5: Boxplot with metrics regarding how students engaged with the game by play style condition (solo and collaborative play).

explored out of the box features that were not required to solve puzzles, and although we see that the `Collaborative` cohort has a slightly superior average of `n_different_events` with 21 different events compared to 20.4 and an average of `n_paint_events` with 0.81 compared to 0.43, these two differences in means are not statistically significant.

Finally, students in the `Collaborative` condition tended to do more random clicking with an average of `n_click_nothing` of 2.21 clicks compared to 1.7 $(t = 1.8, p = 0.07, d = 0.44)$.

6. Discussion

6.1. Differences between Gameplay Styles

Our initial hypothesis was that a `Collaborative` play mode could make students slightly less engaged with the game environment and tasks, due to more talking, switching controls, exchange of ideas and social interactions, which in turn could be helpful to actually make students more engaged with the overall activity and less prone to disengage due to problems solving puzzles, but at the same time could make students slower to solve simple tasks.

Based on the data-driven engagement analytics, we observed a number of significant differences in how students interacted with *Shadowspect* between the `Solo` and `Collaborative` conditions. Students were encouraged to first solve tutorial levels and then move on to the puzzle levels, however, *Shadowspect* does not enforce that linearity as all levels are unlocked from the beginning and students can always quit a level and move to another one. As a result, we found that students in the `Collaborative` condition were much less linear in how they advanced through the levels, whereas `Solo` students advanced much more linearly, as recommended in the scripted chat. This might have caused one of the most noteworthy statistical differences, which is that students in the `Collaborative` condition solved an average of four tutorial levels compared to the 7.3 of the `Solo` condition–and note this was not directly related to struggle with these levels, but because they did not even access the additional tutorial levels. This could be explained by the pair feeling overconfident or failing to establish roles to productively negotiate different ideas.

Other differences that we observed include that students playing `Solo` had statistically significant higher in-game engagement metrics in terms of more `active_time` and `n_events`, and that students under the `Collaborative` condition explored more due to a higher number of `n_different_events` and `n_paints`, though these were not statistically significant. This could be explained by social interactions and dynamics within the pair. For instance, turn-taking and negotiating commonly occur in collaborative play Koivula et al. (2017), and this might have led to more exploration among possible set of choices (e.g., "Should I click this?" or "Why don't you try this shape?"). This is also consistent with the finding that despite `Collaborative` students spent more time interacting with puzzle levels, they were able to solve fewer of them than `Solo` students. One key detail to add is that even though in the `Collaborative` condition students engaged less with the game, this does not necessarily mean that they were less engaged with the overall activity. More nuanced approaches would be required to monitor and match how the dyads interacted with each other (e.g., looking at their conversations) physically outside the game environment. We also observed a higher prevalence of unproductive clicks and slightly lower persistence for `Collaborative` condition. The former can be related to previous research which found that preschoolers engage in random clicking when they do not know when to complete a task Plowman & Stephen (2005).

There are some limitations to these findings, such as that we did not perform a systematic random pairing for the `Collaborative` condition, we did not collect other sources of data (like qualitative observations or video feed) and that we did not perform a pre- and post-test design to assess the effect on learning. However, we did not plan this experiment to have a confirmatory resolution on the effect on learning, instead we were aiming to analyze differences in how learners engaged with the game and to evaluate those differences based on in-game trace data, which is a novel approach that scales up to large groups. The results from this experiment support clear differences regarding in-game engagement metrics between the two groups, raising the idea that the effect of sharing a computer with interactive environments in the classroom is something that deserves more experimental research. Future studies following this line of work can address the effect on learning by performing experiments with more sessions and pre- and post-test designs, as well as capturing additional qualitative observations about the collaboration.

6.2. Implementing Gameplay Sessions in the Classroom

Game-based learning is multifaceted and multiple factors are at play Plass et al. (2015). Previous studies suggested that collaborative gameplay facilitated better in-game performance than solo play Shih et al. (2010); Plass et al. (2013). However, in our results we found lower level completion in `Collaborative` compared to `Solo` gameplay, and a more exploratory behavior of *Shadowspect* features in the case of the `Collaborative` style. One of the potential factors influencing these results could be that the activity was completely ungraded. Previous work has found that students interacted in a different way when working alone or paired if the activity was graded or not Arterberry et al. (2007). Another possibility to consider is that the play time might have been too limited to really achieve mastery in the game, especially if they were collaborating. Previous work has found that when comparing solo vs collaborative game play, some of the strategies selected by the students collaborating were less efficient and error-prone, and that perhaps more time was necessary to achieve a higher joint game fluency than students playing alone Plass et al. (2013). This could also be influenced by the actual collaboration implementation that we did, i.e., having students sit together with a single laptop and switch controls; previous work has explored different ways to implement collaboration with games, for example, by using games that support multi-mouse control or augmented reality markers Echeverría et al. (2012). These alternative options and others should be considered by teachers implementing game collaboration in the classroom.

Another factor to consider is the sociocultural influence of having students to play alone or together. As usual in multicultural K12 settings, there are big differences in how each learner engages socially with each other learner, with some of them being very extroverted while playing *Shadowspect*, and others silently playing the game. For the individual condition, students were sitting at individual tables and most of them were focused on their computers, some of them quietly focused on the task and others talking with friends from a distance, bragging about the puzzles they completed or challenging their friends. The students who worked in pairs were sitting in clusters of around four desks facing each other. Most students were willing to work with whomever they were sitting next to, and they shared one computer between them. They looked at the screen together and either discussed the levels frequently while taking turns controlling the mouse, or watched each other explore the controls and solve problems, occasionally chiming in with comments and suggestions. In both cases, some students who did not like the activity or struggled, quickly disengaged with the game and barely did anything during the session, sometimes performing an off-task activity at their desks or talking to other peers. However, pairing up students was helpful to avoid this issue, as if one of the students was less fluent in the geometry game, the support of the other student would make the first student less likely to quit.

Previous work found that some students prefer individual playing methods instead of collaborative ones Hamalainen (2008), or that students might feel higher levels of stress during competitive gameplay Romero et al. (2012). One first thought could be to try to accommodate for different students' preferences when implementing these experiences. However, another point of view is that these different game play styles can be seen as a feature, since each game style can reinforce different skills. For example, solo play could reinforce the self-efficacy and independence of students, collaborative gameplay the communication and teamwork skills, competitive gameplay the capacity to work under pressure and focus; all of these are important future-ready skills. Therefore, developers should consider implementing learning games that support different game play styles, so that teachers can rotate between different game styles to get the best of each mode.

Although we did not conduct systematic observations as part of this study, we can report anecdotally that there was plenty of student interaction and a range of affect elicited from the gameplay. The overall perception is that students enjoyed the activity and it helped them become more interested in computer science courses (based on feedback from the high school staff). Students expressed in the feedback survey how they felt about the session in terms of "Happy", "OK" or "Sad". Our perceived enjoyment is aligned with the feedback survey where 83% of students (N = 147) reported feeling "Happy" or "OK" during the session. While we could not separate the feedback by condition, it would have been interesting to see if the collaborative style had any effect on enjoyment, as previous work has found that collaborative and competitive play were more enjoyable than solo play Plass et al. (2013).

7. Conclusions

We have performed a cohort-based factorial experiment focusing on the effect of solo vs. collaborative gameplay using the digital geometry game *Shadowspect* with high school first year students. The main findings include that students playing `Solo` had a higher in-game engagement and solved more levels, while `Collaborative` students were less linear in their gameplay patterns (i.e., skipping more tutorial levels and demonstrating more exploratory behaviors with *Shadowspect* features). Our results do not necessarily imply one gameplay mode or the other is good or bad, but we want to raise awareness that we see significantly different outcomes in how students engaged with the game and the social interaction that emerges by having students working alone or in dyads. Educators who want to incorporate games in classrooms can take these findings into account to implement learning games more effectively in their classrooms. For example, when deciding to have students work individually or collaboratively, if the goal is having students complete a given task most efficiently, then the educator might want students to work individually. If the digital environment is more open-ended and has great affordances for exploration, then the educator might consider collaborative play.

Overall, the feedback from students was positive. The students reported that this was a nice exposure to alternative playful approaches for learning of academic topics such as math that are usually regarded as boring. They also reported that this experience made them more interested in technology and computer science for their upcoming selection of elective classes. This experiment using in-game metrics demonstrates the potential of learning analytics to provide insights about how students approached the game beyond their performances. However, it also exemplifies its limitations; for example, we cannot know if lower levels of activity within the game can be translated to more disengaged students, or if it meant that students were engaging outside the game environment, which could also be quite positive for the learning process depending on the type of activity. Additional data collection method in the future can connect in-game metrics with external activities by either using multimodal approaches (e.g., image and voice processing) or by systematically coding what students are doing. We utilized the common tutorial levels as a pre-test measurement, however we found that learners within the `Collaborative` condition tended to skip some of these tutorials because following the pre-established sequence of puzzles in the game was not mandatory. This highlights the complexity of anticipating human behavior, specially when working with groups of kids with diverse backgrounds and interests in a classroom. Future experiments might want to consider administering external pre- and post-tests, or perhaps making some game levels mandatory to provide a suitable and less-obtrusive in-game equivalent to such tests.

We believe that these findings are interesting as the first experiment that investigates different play patterns using analytics and calls for additional experiments around solo and collaborative approaches in digital games and other digital tasks in the classroom, especially in K12 settings where digital games can play a key role to maintain students motivated and engaged in critical ages. We plan to deepen our understanding of the effects of game play style with more experiments on the role that collaboration can have in measures like persistence or creativity, and perhaps in other contexts, such as higher education.

References

Adachi, P. J., & Willoughby, T. (2016). Does playing sports video games predict increased involvement in real-life sports over several years among older adolescents and emerging adults? *Journal of youth and adolescence*, *45*, 391–401.

Admiraal, W., Huizenga, J., Akkerman, S., & Ten Dam, G. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior*, *27*, 1185–1194.

An, Y.-J., & Cao, L. (2017). The effects of game design experience on teachers attitudes and perceptions regarding the use of digital games in the classroom. *TechTrends*, *61*, 162–170.

Arellano, D. G., Tokarchuk, L., & Gunes, H. (2016). Measuring affective, physiological and behavioural differences in solo, competitive and collaborative games. In *International Conference on Intelligent Technologies for Interactive Entertainment* (pp. 184–193). Springer.

Arterberry, M. E., Cain, K. M., & Chopko, S. A. (2007). Collaborative problem solving in five-year-old children: Evidence of social facilitation and social loafing. *Educational Psychology*, *27*, 577–596.

Baranowski, T., Blumberg, F., Buday, R., DeSmet, A., Fiellin, L. E., Green, C. S., Kato, P. M., Lu, A. S., Maloney, A. E., Mellecker, R., Morrill, B. A., Peng, W., Shegog, R., Simons, M., Staiano, A. E., Thompson, D., & Young, K. (2016). Games for Health for Children-Current Status and Needed Research. *Games for health journal*, *5*, 1–12. doi:10.1089/g4h.2015.0026.

Baumöl, U., & Bockshecker, A. (2017). Evolutionary change of higher education driven by digitalization. In *2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1–5). IEEE.

Berland, M., & Lee, V. R. (2011). Collaborative strategic board games as a site for distributed computational thinking. *International Journal of Game-Based Learning (IJGBL)*, *1*, 65–81.

Bressler, D. M., Oltman, J., & Vallera, F. L. (2018). Inside, outside, and off-site: Social constructivism in mobile games. In *Handbook of Research on Mobile Technology, Constructivism, and Meaningful Learning* (pp. 1–22). IGI Global.

Buckingham, D., & Willett, R. (2013). *Digital generations: Children, young people, and the new media*. Routledge.

Chapman, P. M. (1997). *Models of engagement: Intrinsically motivated interaction with multimedia learning software*. Ph.D. thesis University of Waterloo.

Council, N. R. et al. (2011). *Learning science through computer games and simulations*. National Academies Press.

De Freitas, S. (2006). Learning in immersive worlds: A review of game-based learning. *JISC*, .

De Grove, F., Bourgonjon, J., & Van Looy, J. (2012). Digital games in the classroom? a contextual approach to teachers adoption intention of digital games in formal education. *Computers in Human behavior*, *28*, 2023–2033.

Divjak, B., & Tomić, D. (2011). The impact of game-based learning on the achievement of learning goals and motivation for learning mathematics-literature review. *Journal of Information and Organizational Sciences*, *35*, 15–30.

van Drie, J., van Boxtel, C., Jaspers, J., & Kanselaar, G. (2005). Effects of representational guidance on domain specific reasoning in cscl. *Computers in Human behavior*, *21*, 575–602.

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

Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game learning analytics: learning analytics for serious games. *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, (pp. 1–29).

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, *1*, 20–20.

Hamalainen, R. (2008). Designing and evaluating collaboration in a virtual game environment for vocational learning. *Computers & Education*, *50*, 98–109.

Homer, B. D., Plass, J. L., Raffaele, C., Ober, T. M., & Ali, A. (2018). Improving high school students' executive functions through digital game play. *Computers & Education*, *117*, 50–58.

Kaye, L. K., & Bryce, J. (2014). Go with the flow: The experience and affective outcomes of solo versus social gameplay. *Journal of Gaming & Virtual Worlds*, *6*, 49–60.

Ke, F., & Grabowski, B. (2007). Gameplaying for maths learning: cooperative or not? *British Journal of Educational Technology*, *38*, 249–259.

Kim, Y. J., & Shute, V. J. (2015). Opportunities and challenges in assessing and supporting creativity in video games. In *Video games and creativity* (pp. 99–117). Elsevier.

Koivula, M., Huttunen, K., Mustola, M., Lipponen, S., & Laakso, M.-L. (2017). The emotion detectives game: Supporting the social-emotional competence of young children. In *Serious games and edutainment applications* (pp. 29–53). Springer.

Li, Q. (2013). Digital game building as assessment: A study of secondary students' experience. In *Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL conference*. volume 40.

Marfisi-Schottman, I., & George, S. (2014). Supporting teachers to design and use mobile collaborative learning games. In *International Conference on Mobile Learning* (pp. 3–10).

Martey, R. M., Kenski, K., Folkestad, J., Feldman, L., Gordis, E., Shaw, A., Stromer-Galley, J., Clegg, B., Zhang, H., Kaufman, N. et al. (2014). Measuring game engagement: multiple methods and construct complexity. *Simulation & Gaming*, *45*, 528–547.

Mee, R. (2009). *A comprehensive guide to factorial two-level experimentation*. Springer Science & Business Media.

Ofcom (2019). *Children and parents: media use and attitudes report 2018*. Technical Report.

Papastergiou, M. (2008). Online computer games as collaborative learning environments: Prospects and challenges for tertiary education. *Journal of educational technology systems*, *37*, 19–38.

Pastore, R. S., & Falvo, D. A. (2010). Video games in the classroom: Pre-and in-service teachers' perceptions of games in the k-12 classroom. *International Journal of Instructional Technology and Distance Learning*, *7*, 49–57.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*, 258–283.

Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of educational psychology*, *105*, 1050.

Plowman, L., & Stephen, C. (2005). Children, play, and computers in pre-school education. *British journal of educational technology*, *36*, 145–157.

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE)*, *1*, 21–21.

Romero, M., Usart, M., Ott, M., Earp, J., & de Freitas, S. (2012). Learning through playing for or against each other? promoting collaborative learning in digital game based learning. In *European Conference on Information Systems* (p. Paper 93).

Ruiperez-Valiente, J. A., Gaydos, M., Rosenheck, L., Kim, Y. J., & Klopfer, E. (2020). Patterns of engagement in an educational massive multiplayer online game: A multidimensional view. *IEEE Transactions on Learning Technologies*, .

Shaffer, D. W. (2006). *How computer games help children learn*. Macmillan.

Shih, J.-L., Shih, B.-J., Shih, C.-C., Su, H.-Y., & Chuang, C.-W. (2010). The influence of collaboration styles to childrens cognitive performance in digital problem-solving game "william adventure": A comparative case study. *Computers & Education*, *55*, 982–993.

Squire, K. (2005). Changing the game: What happens when video games enter the classroom? *Innovate: Journal of online education*, *1*.

Starkey, P. L. (2013). *The effects of digital games on middle school students' mathematical achievement*. Ph.D. thesis Lehigh University.

UNESCO Bangkok (2016). *School and Teaching Practices for Twenty-first Century Challenges: Lessons from the Asia-Pacific Region - Regional Synthesis Report*. Technical Report. Https://unesdoc.unesco.org/ark:/48223/pf0000244022.

Voulgari, I., & Komis, V. (2008). Massively multi-user online games: The emergence of effective collaborative activities for learning. In *Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning* (pp. 132–134). IEEE.

Watson, W., & Yang, S. (2016). Games in schools: Teachers perceptions of barriers to game-based learning. *Journal of Interactive Learning Research*, *27*, 153–170.

Wendel, V., Göbel, S., & Steinmetz, R. (2012). Game mastering in collaborative multiplayer serious games. In *E-learning and games for training, education, health and sports* (pp. 23–34). Springer.

Wendel, V., Hertin, F., Göbel, S., & Steinmetz, R. (2010). Collaborative learning by means of multiplayer serious games. In *International Conference on Web-Based Learning* (pp. 289–298). Springer.