

Evaluating the Impact of Automated Hints in a 3D Educational Escape Game: A Comparative Study of Accessibility and Computer Science Versions

El Montaser Osmane
osmane.el-montaser@lip6.fr
Sorbonne Université, CNRS, LIP6
Paris, France

Lallé Sébastien
sebastien.lalle@lip6.fr
Sorbonne Université, CNRS, LIP6
Paris, France

Muratet Mathieu
mathieu.muratet@lip6.fr
Sorbonne Université, CNRS, LIP6
Paris, France
INSEI
Suresnes, France

ABSTRACT

This paper focuses on analyzing a 3D virtual educational escape game, which helps players solve puzzles by offering hints automatically. The study aimed to determine whether these automated hints improved player retention and progression, and if these effects are influenced by the players' gaming skills. We present our results on two instances of an escape game, which have the same game structure but different learning content, the first focusing on accessibility and inclusive education (ACCESS version) and the second on computer science (CS version). Although the structure of the two versions is identical, we found that the results differ between them. The automatic hint system was very effective in the ACCESS version, improving both retention and progression. In contrast, it is less effective in the CS version, where the complexity of the puzzles might have hindered progression.

CCS CONCEPTS

• **Applied computing** → **Interactive learning environments; Computer games**; • **Human-centered computing** → *Field studies*; • **Software and its engineering** → *Petri nets*.

KEYWORDS

Educational escape game, Automatic hints, Naturalistic setting, Learning analytics

ACM Reference Format:

El Montaser Osmane, Lallé Sébastien, and Muratet Mathieu. 2025. Evaluating the Impact of Automated Hints in a 3D Educational Escape Game: A Comparative Study of Accessibility and Computer Science Versions. In *30th International Conference on Intelligent User Interfaces (IUI '25)*, March 24–27, 2025, Cagliari, Italy. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3708359.3712132>

1 INTRODUCTION

Adaptive support in Technology-Enhanced Learning (TEL) is important to help learners progress in problem solving and improve their learning process [4, 32, 35]. In educational games, designing

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only. Request permissions from owner/author(s).

IUI '25, March 24–27, 2025, Cagliari, Italy

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1306-4/25/03

<https://doi.org/10.1145/3708359.3712132>

adaptive support remains a challenge because of their exploratory and open-ended nature, making it difficult to understand when a learner needs help, and how to frame and deliver this help [6, 7]. This is especially true in virtual educational escape games, where learners can freely interact with numerous game objects and explore the game environment at will.

Escape games are problem-based, offer limited playing time, and require active and cooperative participants to look for clues to solve a series of puzzles [23]. Several puzzles are typically available at the same time, and the players must associate the right clues to the right puzzle before the time limit. Educational escape games add a pedagogical dimension to these games by integrating learning material and requiring learners to use their knowledge to solve the puzzles [30, 36]. A key role in such escape games is that of the game master who supervises the players' progress and provides them with hints upon request [1, 17, 24, 36, 39], albeit without sharing the solutions [28, 30]. However, this key role is not always filled in education [31], either when learners play on their own without any supervision, or when teachers want to integrate the game in their courses but lack the training and expertise to be the game master themselves. Without the support of a game master, learners may progress too slowly or even get stuck [1, 17, 24, 28, 39], which is especially concerning for novice players who have little experience and skills with such games.

In this paper, our goal is to evaluate the effectiveness of the existing help system of an educational escape game named E-LearningScape, when learners play the game without the supervision of a game master. This help system provides automatic hints to the players when their progress and behaviours in the game indicate that they are having difficulties. This research goal leads to two research questions:

- (RQ1) Do automatic hints increase players' retention and progression in E-LearningScape?
- (RQ2) Does this increase depend on the player's experience with virtual escape games?

To address these research questions, we leverage the data of in-the-wild players who interacted with two versions of E-LearningScape, about accessibility and computer science. We then formally compare in both game versions the retention and progression of learners who played with and without the hint system, while accounting for their gaming experience.

The originality of this work lies in experimenting with a hint system in two similar educational escape games in the wild. Our

results show the importance of replicating such experiment across several games, as it allowed to examine in more details what results could not be replicated across versions and why. These findings are especially interesting because, to the best of our knowledge, there has been no study that formally evaluated a hint system with end-users in virtual escape games so far.

The plan of this paper is as follows. We review related work in Section 2. Then, we present E-LearningScape in Section 3, and the experimental setup in Section 4. Next, Section 5 details the results and we analyze them in Section 6 before concluding.

2 RELATED WORK

There has been extensive research on escape games [11], especially in educational context [12, 23, 34, 37]. Most of the games described in these systematic reviews are physical escape rooms where digital technologies are mainly used to monitor the safety and progression of learners from an adjacent room, to structure the gameplay (scanning a QR code for instance) or to support the narrative and enhance immersion [37]. Fotaris and Mastoras [12] identify in their study that only 10.6% of educational escape games (7) are virtual.

Thus, studies on virtual educational escape games remains fairly unexplored, and none of them have formally evaluated automatic hints in this context, as we do in this paper. For instance, López-Pernas *et al.* [21] studied how learners' behaviours are predictive of knowledge acquisition in an educational escape game. Gordillo *et al.* [14] studied the effectiveness of a remotely conducted educational escape room designed to teach software modelling in a software engineering course, demonstrating that such activities can significantly enhance student learning and engagement, even in an online setting. In both [14, 21], students could request a pre-defined hints at will, meaning that the hints were not automated nor tailored to their behaviours in the game. The value of the hints, as compared to not receiving hints, was also not evaluated in their work, as it was not the purpose of their research.

The advantage of virtual escape games over physical rooms is that they allow for tracking the players' actions in real time to provide automatic feedback as needed. We know of two papers [15] [26] that addressed this challenge of how to deliver hints automatically using players' data in educational escape games, though neither evaluated their effectiveness. Specifically, Hare and Tang [15] proposed a hierarchical reinforcement learning approach to build a hint system based on Petri nets. However, their methodology solely relied on simulated data to train their model, and no evaluation of the hints with actual learners was provided.

In previous work [26], we modelled the resolution process of all puzzles in E-LearningScape with Petri nets, so as to provide hints on what to do next to progress through the game. The Petri nets model how far a learner is in solving each puzzle by monitoring all of their in-game actions, as well as generate so-called *pedagogical labels* to tag relevant learners' actions (e.g., erroneous action, action done too late). We integrated in E-LearningScape the Learner Activity Analyzer (Laalys) [27], an algorithm to leverage Petri nets in real time in a game. However, in [26], we focused on designing the Petri nets and the hints, but did not actually evaluate their effectiveness with learners. In this paper, we extend this previous work by formally evaluating the value of the automatic hint system,

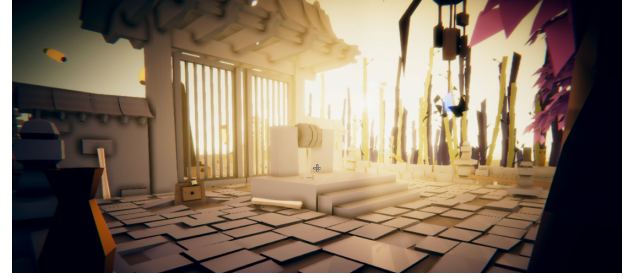


Figure 1: Screenshot of “E-LearningScape” second room.



Figure 2: Screenshot of “E-LearningScape” rope puzzle.

as compared to not receiving hints. To the best of our knowledge, this is the first formal evaluation of the value of automatic hints in virtual escape games (educational and non-educational alike).

Beyond escape games, research has also explored providing adaptive guidance in video games to optimize engagement and the gaming experience, including via tutorial, tooltips and hints, e.g., [8, 9, 16, 18, 25, 38]. The findings of such research with end-users are still mixed, as some studies found that adaptive hints may not be effective, depending on the players' skills [8, 25, 38] and the complexity of the game [2]. This highlights the need to further evaluate adaptive hints in video games and replicate results across games, while accounting for individual and game-specific factors, which we contribute to exploring with the two versions of E-LearningScape.

3 PRESENTATION OF E-LEARNINGSCAPE

E-LearningScape (see Figures 1 2) is an open source¹ digital adaptation of a physical escape game². In the digital adaptation, the players play the role of a sandman immersed in the dream of Camille. Their challenge will be to help Camille structure her thoughts in her dream by solving puzzles. The player interacts with the game and moves inside the 3D virtual environment with a first-person perspective to discover interactive game objects and clues. Clues are named “**Dream fragments**” (in reference to the game pitch: help Camille structure her thoughts in her dream) and allow players to solve the puzzles.

¹E-LearningScape source code: <https://github.com/Mocahteam/E-LearningScape>, accessed September 19, 2024

²<https://projects.learningplanetinstitute.org/projects/learningscape/>, accessed August 06, 2024

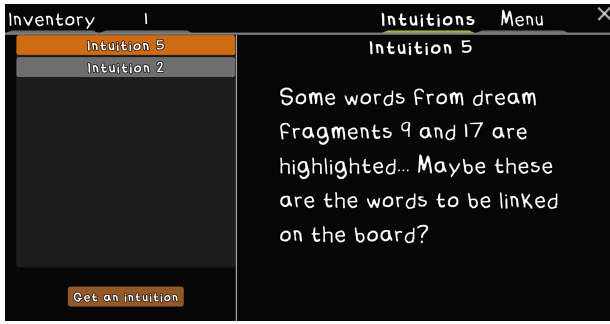


Figure 3: Example of hint in “E-LearningScape”.

E-LearningScape is made up of 17 puzzles that require the players to combine more than 40 game objects found in the game. Without being exhaustive, here are some examples of puzzles: finding a rope to connect words on a pin panel; deciphering a message with a cylinder mirror; revealing hidden symbols with a black light; etc. As in classic escape rooms, several puzzles can be performed in parallel and the players have to discover and associate relevant Dream Fragments to solve them.

The content (but not game mechanics) of E-LearningScape is fully customizable to target different learning topics. In this paper, we use the two game versions based on E-LearningScape from which we get user data. The first one is about accessibility and inclusive school and the second one is about introducing what is computer science. We name these two versions ACCESS and CS in the rest of the paper. There exists a third version about pedagogy, but without user data yet.

These two versions have the same architecture that can’t be changed: same number of rooms; same position of game objects (puzzles, dream fragments, interactive objects); and same number of Dream Fragments and puzzles per room. The first “room” is an introduction and acts as a tutorial for learning how to navigate in the game and how to pass one simple puzzle. This introduction is followed by three sequential rooms: room 1 contains five puzzles; room 2 contains seven puzzles; and room 3 contains the last four puzzles. Even if the game architecture is fixed as said above, all of the text and pictures in the game, including the Dream Fragments linked to puzzles, can be customized in each game version. For instance, the question for the rope puzzle (room 1) in the ACCESS version (see Figure 2) is “What does the 2001 International Classification of Functioning offer?” and the players have to link “Functions”, “Capacity” and “Participation”, while the CS version asks “Link the three key ideas of computer science with the rope” and players have to link “Information”, “Processing” and “Science”. Obviously, the content of the Dream Fragments associated to this puzzle was changed accordingly.

In addition, Dream Fragments can also be links to resources outside the game such as online videos and online interactive applications. This functionality is used in the CS version to enable player discovery of third party applications such as Compute-it³ and BlocklyMaze⁴.

We now provide a brief description of the **hint system** in E-LearningScape, sufficient for the purpose of this paper, as our research goal is to focus on the user study and data analysis to evaluate the hints. The hint system was designed to provide gradual support through incremental levels of details, in line with best practices for adaptive support in learning environments [5, 20]. Consulting the hints also remains fully optional, so as to minimize intrusiveness and preserve a sense of autonomy, as per [25, 29, 38]. As reported in Section 2, the hint system was detailed in [26] and is based on Petri nets that model the resolution process of the 17 game puzzles. Depending on the pedagogical labels provided by Laalys, the number of puzzles solved and the remaining time, the hint system selects the next hint to be provided to the player for an unsolved puzzle. The content of the hints was phrased by the game designers and teachers, and associated to their relevant Petri net. The default hint database includes 121 predefined hints (named “intuition” in the game) that are meant to be independent of the game version (CS and ACCESS use the same default hint database). The purpose of these hints is to help players find the Dream Fragments related to unsolved puzzle, without referring to game content or the learning material. Three levels of hints can be defined: level 1 provides general information (“I should take a better look at the ground” for instance), level 2 provides more specific indications (“I have a feeling that one of the floor tiles is different” for instance), and level 3 provides very explicit instructions (“There is a different floor tile on the left as you enter the building” for instance). As shown by these examples, the wording of the hints does not refer to the CS or ACCESS subject and is focused to help players find useful objects and Dream Fragments to solve the puzzles. The default database contains 83 level-1 hints, 36 level-2 hints and 2 level-3 hints. Figure 3 shows an example of hint level 2 in the game. This database was designed by researchers and teachers involved in this work. In addition, for each game version, this default hint database is augmented with 13 level-3 version-specific hints (9 for ACCESS, 4 for CS), which provide help specifically about the learning content. The learners can display the hints at any time by clicking the “Intuitions” button, which opens a dedicated panel (shown in Figure 3).

4 EXPERIMENTAL SETUP

The data we analyze in this paper were generated from a field user study we designed and conducted in naturalistic settings. To get participants, we advertised the game versions on two main networks via dedicated mailing lists and word-of-mouth. The ACCESS version was advertised on the network of teachers invested on the field of accessibility, disabilities and inclusive school. The CS version was advertised to computer science beginners (grades 11, 12 and first university year). Interested players were free to play the game remotely, without any supervision from a researcher. When launching the game, the hint system was randomly enabled/disabled, to enable comparisons of learners playing with the hints, versus without. At the end of the game, players are invited to fill in a qualitative questionnaire⁵ to collect demographics and feedback about their gaming experience. The questionnaire is composed of three parts.

³Compute-it: <https://compute-it.toxicode.fr/>, accessed August 06, 2024

⁴BlocklyMaze: <https://blockly.games/maze>, accessed August 06, 2024

⁵French qualitative questionnaire used at the end of a game session: <https://sondages.inshea.fr/index.php/153235?lang=fr>, accessed January 16, 2025.

Version	Gender			Ages				
	Female	Male	Undef.	<18	18-30	31-40	41-55	>55
ACCESS	79%	13%	8%	0%	46%	31%	21%	2%
CS	30%	63%	7%	34%	58%	0%	4%	4%

Table 1: Players' gender and age.

The first part contains five closed-ended questions on gender, age and game experience and habits. The second part consists of six 5-point Likert scale questions on the players' needs, based on Siang and Radha's work [33]. The third and last part includes nine open-ended questions to gather feedback on the game and puzzles. The hints

All of the data collected in the game and the questionnaire were kept fully anonymous. When downloading the game, learners were informed about the data collection process, which was assessed as compliant with the EU's General Data Protection Regulation. A total of 1,908 game sessions were collected from January 2022 to August 2024, of which we retained 658 for the analysis (see next subsections on data cleaning). Table 1 provides the gender and age of the players per game version, as reported in the post-game questionnaire.

4.1 Data gathering

All in-game interactions were tracked in xAPI⁶ format and stored on a Learning Record Store (LRS). Each trace in xAPI format is called a statement and consists of an Actor/Verb/Object triplet (Who? Does what? On what?). Each new game session generates a unique, anonymous session identifier. This session ID defines the statement actor. Verbs and objects are used to characterize the nature of the interactions. The vocabulary of E-LearningScape is made of 21 verbs for 25 types of objects. To evaluate the hint system in this paper, we needed to focus on five verbs:

- Moved: indicates that the player moves to a target position.
- Received: indicates that the player receives a new hint.
- Read: indicates that the player reads an hint.
- Answered: indicates that the player solved a puzzle.
- Completed: indicates that the player completes a room.

4.2 Data cleaning

We gather 1,163,828 rows of data. Each row corresponds to one action (statement) performed in the game by a player. We convert these rows into distinct game sessions by grouping them by their anonymous session ID, thus getting 1,908 sessions. Because the game versions were played into the wild without supervision from researchers as stated above, a lot of these sessions are not meaningful and need to be removed. We showcase in Figure 4 the data cleaning process we devised to do so. First, for the purpose of this paper, we consider only remote sessions (1,686 sessions), to exclude in-classroom sessions where we were playing the role of the game master. Next, we remove sessions where the players cleared a room without performing the mandatory actions of looking for Dream Fragments. This could happen when players use teleportation and/or already know the game. Teleportation is accessible

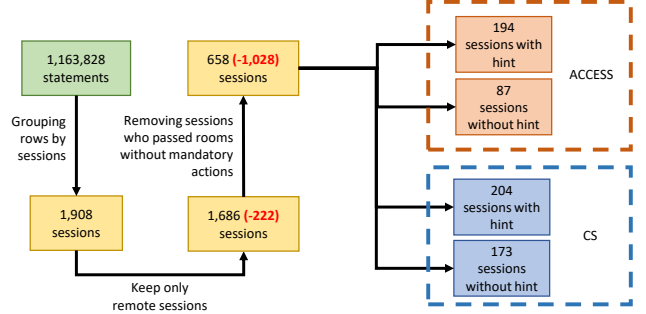


Figure 4: Data cleaning pipeline of E-LearningScape's data.

with a cheat code that enables to spawn the avatar in any room of the game, which is mostly used by the game designers to test a specific puzzle. These sessions with teleportation are thus not actual learning session and are removed from the dataset. We also identify “speedrunners”, i.e., players who try to complete the game as quickly as possible. These players tend to play many times by solving puzzles without consulting any Dream Fragments. Thus, we remove sessions where the learners did not look for those Dream Fragments that we deem mandatory to solve puzzles. This filter reduces the amount of sessions to 658.

Table 2 shows the number of sessions we kept after data cleaning. The “With hints” group contains sessions where the hint system was randomly enabled and players accessed the hints panel at least twice and actually read at least one hint. It should be remembered that the data analyzed in this paper came from game sessions conducted in the wild in an naturalistic context, so players were not aware of the presence of the hint system. The first consultation of the hint panel is therefore an opportunity for players to discover the hint system. If the players read a hint and consult the hint panel again later, we consider that in this session the player intentionally used the hint system. The “Without hints” group contains sessions where the hint system was randomly disabled, as well as sessions where it was enabled but the player did not read any hints or did not access the hint panel more than once. As a result, we obtained for the data analysis presented in the next section 194 sessions with hints and 87 sessions without hints for the ACCESS version, and 204 sessions with hints and 173 sessions without hints for the CS version (see Table 2). A power analysis⁷ confirmed that these sample sizes are sufficient to detect even small effects in the statistical analysis reported Section 5.

⁶<https://adlnet.gov/projects/xapi/>, accessed August 06, 2024

⁷Using G*Power [10] with $\alpha = .05$, $power = .9$, and $f = .2$.

Version	With hints	Without hints
ACCESS	194	87
CS	204	173

Table 2: Number of sessions with and without hints by game version after cleaning the data.

4.3 Metrics

We measure **retention** by the total time played in the game. The **progression** is measured by the number of puzzles solved out of the 17 puzzles that have to be done to reach the end of the game. We use these two metrics to gauge learners’ engagement and performances following previous work on educational escapes rooms, e.g., [13, 14, 21]. The rationale is that if the hints are truly effective, players who read them should remain engaged longer and have less difficulties overcoming the puzzles, thus playing longer to solve more puzzles, as compared to not reading the hints. We also hypothesize that players with lower gaming skills should benefit more from the hints, as they may have even more difficulties with solving puzzles and may drop-out faster, as mentioned earlier.

5 RESULTS

We first present how we inferred learners’ gaming skills in our work in Section 5.1 and the extent of hint usage 5.2. Next, we report on the results of our statistical analysis to answer our research questions for retention (Section 5.3) and progression (Section 5.4). Notice that our code and data are available in open source at <https://gitlab.lip6.fr/elmontaser/e-learning-scape-analysis>.

5.1 Gaming experience classification

Our second research question aims at exploring whether the value of the automatic hints dependant on the player’s experience and skills with virtual escape games like E-LearningScape. As we presented in Section 3, E-LearningScape is based on a 3D virtual universe with a first-person perspective. This kind of navigation is common in video games but not all learners have experience with it. Furthermore, learners may have never played escape games and thus could have a hard time progressing through the game puzzles. Thus, we aim to capture as a factor in our analysis the players’ underlying skills in navigating complex 3D virtual environments and in using game mechanics typically found in virtual escape games. We denote these skills as “gaming skills” for the rest of the paper.

To identify these gaming skills, we opted to analyze the time spent in the introduction by the players on both versions. Indeed, the game introduction is exactly the same between ACCESS and CS, and acts as a tutorial for learning how to move, interact with objects, use the inventory, and look for a first Dream Fragment to solve one simple puzzle. These are common actions in E-LearningScape and in virtual escape games in general. Specifically, we perform a cluster analysis on the time spent in the introduction. We focus on simple clustering algorithms which we deem sufficient to perform clustering on solely one features, namely k-means and Gaussian mixture models (GMM), both using a number of clusters ranging from 2

Version	Cluster 1 (High skill)	Cluster 2 (Low skill)
ACCESS	135 (48%)	146 (52%)
CS	273 (72%)	104 (28%)

Table 3: Number and percentage of sessions in each version by gaming skills group (GMM clustering).

to 10^8 . We also use a simple median split as a baseline, and use the Silhouette score to compare them, a popular metric to choose the appropriate number of clusters and assess their quality [19]. GMM with two clusters ended up yielding the best Silhouette score (0.54), so we opt to focus on the GMM clusters only. They indicate that players in Cluster 1 cleared the introduction in 1.7 minute on average, against 5.6 minutes on average for Cluster 2. As a sanity check to formally compare the clusters, we run a two-way Anova with the completion time of the introduction as the dependent measure, clusters and hint condition as the factors. The Anova confirms that the difference in completion time across clusters is significant ($p < .0001$). Based on the assumption made earlier, this suggests that players in Cluster 1 have better prior gaming skills with E-LearningScape than those in Cluster 2. For the rest of the paper, we label Cluster 1 as the “**High skill**” group, and Cluster 2 as the “**Low skill**” group. Table 3 shows how many players fall into each cluster per game version. The Anova also showed no significant main effect of hint condition nor a significant interaction with the clusters, indicating that the gaming skills of the participants are independent of the hint conditions.

5.2 Hint usage

This section focuses on how many times players in the *With hints* group read a hint, granted that each hint can be read multiple times. The results shown in Table 4 indicate that hints are widely used, as players read feedback over 20 times on average in both game versions (see top two rows of Table 4). Compared to the mean and median, the standard deviation is high. This suggests that the use of the hint system varies substantially between players. The value 274 in the *max* column for High skill group is due to one player who encountered lots of difficulties to solve one puzzle in the first room. Without this player, the max value would be 112. We also looked at the number of hints read in the introduction (see bottom two rows of Table 4) to examine whether it is consistent with the clusters built in the previous section. The results are indeed in line with the clusters: players in the High skill group made very little use of the hints in the introduction, indicating again that their prior gaming skills is sufficient to clear the very simple puzzle of the introduction. In contrast, players in the low skill group already needed to read feedback several times on average to clear the introduction.

5.3 Player retention

In this section, we analyze player retention by examining the total time each player spent in the game, regardless of whether they completed it. Tables 5 and 6 show descriptive statistics of total time for the ACCESS and CS versions, depending on the learners’ gaming

⁸Larger values for the number of clusters ended up hindering the performance of the clustering algorithms, as measured by the Silhouette score.

Version	High skill					Low skill				
	μ	median	σ	min	max	μ	median	σ	min	max
ACCESS (full game)	22.7	19	18.3	2	90	29	29	20	2	112
CS (full game)	23.9	19	28.8	2	297 (112)	22.6	21	13.6	2	55
ACCESS (introduction)	0.43	0	0.6	0	2	2.75	2	2.77	0	22
CS (introduction)	0.49	0	0.88	0	5	2.93	2	3	0	14

Table 4: Descriptive statistics on hint usage for the entire game with a focus on the introduction.

Group	High skill		Low skill	
	μ	σ	μ	σ
With hints	45.7	31.7	54.8	33.5
Without hints	23.1	54.8	42.7	33.1

Table 5: Mean and Standard Deviation of Total Time in minutes - ACCESS version.

Group	High skill		Low skill	
	μ	σ	μ	σ
With hints	45.4	25.3	40.3	20.8
Without hints	27.1	19.7	48.4	21.8

Table 6: Mean and Standard Deviation of Total Time in minutes - CS version.

skills. In addition, Figure 5 provides kernel density estimates of the total time distributions per game versions and skill groups, to get a better idea of their shape. These tables and figures suggest that the “With hints” group plays for longer than the “Without hints” group in both game versions, except for the Low skill learners in the CS version.

We formally compare the groups by running a statistical analysis. Due to the total time and the residuals not being normally distributed, as shown by Shapiro-Wilk tests, we opt to use the Aligned Rank Transform (ART) ANOVA [40], a non-parametric alternative for two-way ANOVA. Specifically, we run two ART-ANOVAs, one per game version, with total time as the dependent variable, and hint groups (two levels: with/without hints) and gaming skill groups (two levels: High/Low) as the factors. We report in this paper significance levels at 0.05, and effect size as high for $\eta p^2 > 0.14$, medium for $\eta p^2 > 0.06$, small otherwise [22]. To account for family-wise errors, we applied the Benjamini and Hochberg correction [3] and confirmed that all significant p -values remained significant.

ACCESS version: The ART-ANOVA reveals a significant main effect of hint groups ($F(1, 277) = 39.80, p < .0001, \eta p^2 = 0.13$) and gaming skills ($F(1, 277) = 24.15, p < .0001, \eta p^2 = 0.08$). These effects show that the *With hints* group played significantly longer than their counterparts, and that the same goes for Low skills players. The interaction effect between hints and gaming skills is close to significance ($F(1, 277) = 3.15, p = .077, \eta p^2 = 0.01$), indicating a potential interaction between the two factors, albeit not at the conventional significance level.

CS version: There is a significant main effect of hints ($F(1, 373) = 20.23, p < .0001, \eta p^2 = .05$) and gaming skills ($F(1, 373) = 12.01, p =$

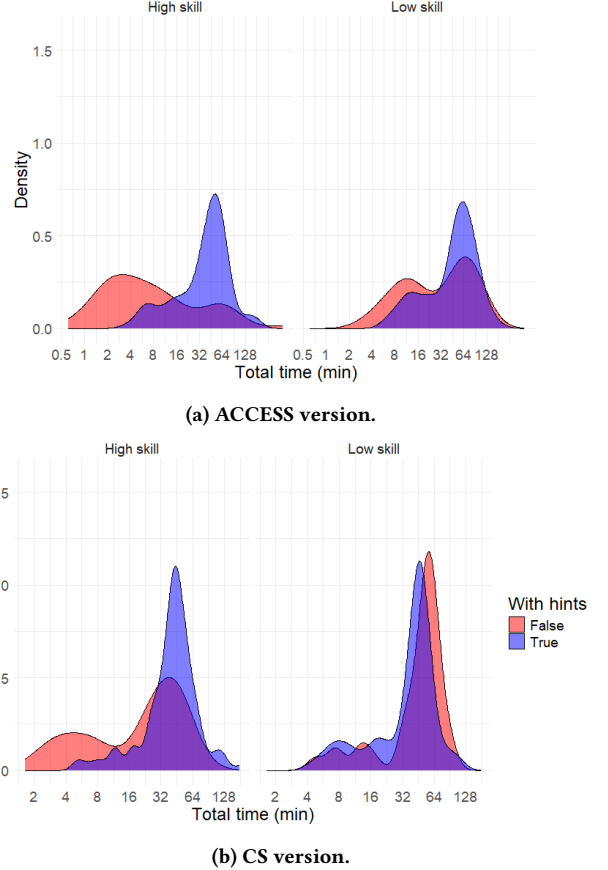


Figure 5: Density of log-transformed total time in each game version depending on gaming skill and hint group.

$.0006, \eta p^2 = .03$), showing again that the *With hints* group and the *Low skill* group play longer than their counterparts. The interaction effect between hints and gaming skills is also significant ($F(1, 373) = 27.67, p < .0001, \eta p^2 = .07$), suggesting that the effect of hints depends on the gaming skills. Pairwise comparisons indicate that the interaction is due to High skill players without hints playing significantly less (27.1 mins) than all other groups.

5.4 Player progression

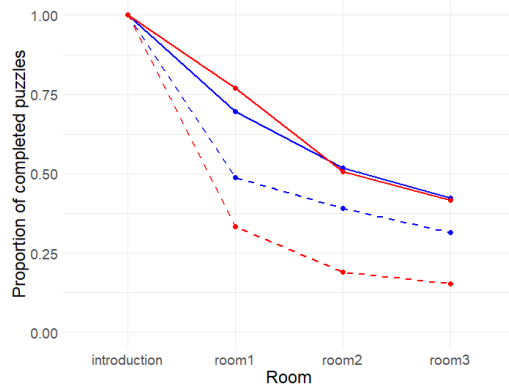
We now analyze learners’ progression as measured by the number of puzzles solved. Tables 7 and 8 show means and standard deviations of number of puzzles solved for the ACCESS and CS versions.

Group	High skill		Low skill	
	μ	σ	μ	σ
With hints	10.1	6.5	9.8	6.9
Without hints	4.6	5.8	7.4	7.3

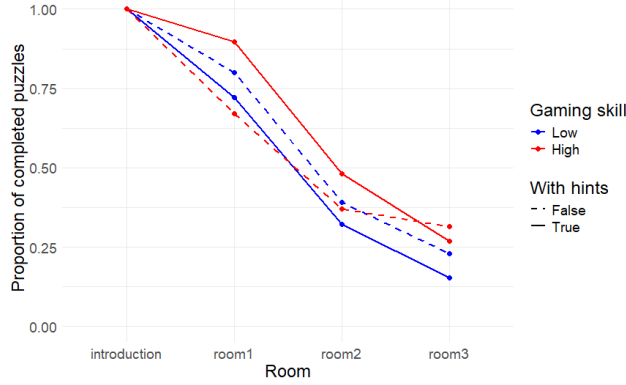
Table 7: Mean and Standard Deviation of Number of puzzles solved - ACCESS version.

Group	High skill		Low skill	
	μ	σ	μ	σ
With hints	9.9	4.9	7.5	5.0
Without hints	8.2	6.0	8.7	5.4

Table 8: Mean and Standard Deviation of Number of puzzles solved - CS version.



(a) ACCESS version.



(b) CS version.

Figure 6: Comparison of puzzles solved in each game version depending on gaming skills and hints group.

Figure 6 also illustrates the proportion of puzzles solved per room, so as to see how the hint effectiveness varies over time. These tables and figures show similar trends as what we found in the previous section for retention, that learners solved more puzzles in all rooms in the *With hints* group, as compared to the *Without hints* group, except for the *High skill* group in the CS version.

We formally verify whether these trends are significant by running a similar statistical analysis as done above for retention, using again ART-ANOVA with the number of puzzle solved as the dependent variable, and the same factors (hint groups, gaming skill groups). We used ART-ANOVA again due to violation of the normality assumption. As done above, we verified the significance of all p -values after applying the Benjamini-Hochberg correction.

ACCESS version: The ART-ANOVA reveals a significant main effect of hint groups ($F(1, 277) = 22.74, p < .0001, \eta p^2 = .08$), but not of skill groups ($F(1, 277) = 2.38, p = .12, \eta p^2 = .01$). As for retention above, these effects show that the *With hints* group outperforms its *Without hints* counterpart in terms of number of puzzles solved. The interaction effect between hints and gaming skills is slightly above the 0.05 cutoff for significance, $F(1, 277) = 3.63, p = .058, \eta p^2 = .01$, indicating a potential interaction between the two factors, albeit not at the conventional significance level.

CS version: There is no significant main effect on progression of hint groups ($F(1, 373) = 2.31, p = .13, \eta p^2 = 0.01$), nor of gaming skills ($F(1, 373) = 2.91, p = .09, \eta p^2 = .01$). The interaction effect between hints and gaming skills is again close to the significance cutoff ($F(1, 373) = 3.43, p = .06, \eta p^2 = .01$), and could be worth studying further in future work.

6 DISCUSSION

In this section, we discuss our results pertaining to our two research questions introduced earlier:

- (RQ1) Do automatic hints increase players' retention and progression in E-LearningScape?
 (RQ2) Does this increase depend on the player's experience with virtual escape games?

The results presented in Sections 5.3 and 5.4 for RQ1 show that the hints were very effective in the ACCESS version, as they enabled learners to play significantly longer and solve significantly more puzzles, as compared to not reading the hints. Results are more contrasted in the CS version, as we found that the hints significantly increased retention, but not progression. This means that the increased play time enabled by the hints did not lead to the learners solving more puzzles. Thus, while we found that the hint system can be very beneficial to the learners (i.e., in the ACCESS version), its effectiveness depends on the game version, meaning that the answer to RQ1 is only partially affirmative.

Our results for RQ2 show that gaming skills with virtual escape games can significantly interact with the hint effectiveness for both retention and progression. Regarding retention, the hints were effective for players with high gaming skills in both game versions, as their total play time was about two times higher with the hints than without (45.7 min vs. 23.1 min in ACCESS; 45.4 min vs. 27.1 min in CS, see Tables 5- 6). Results are similar for progression, with high skill players solving two times more puzzles with the hints than without in the ACCESS version (10.1 vs. 4.6, see Table 7). This increase in solved puzzles was present, albeit less pronounced in the CS version (9.9 vs. 8.2, see Table 8). Overall, these results show that the high skill players were able to take advantage of the hints to progress farther in the game. Low skill players also benefited from the hints, but only in the ACCESS version where the hints led to increased retention and progression (see Tables 5- 7). In itself, this is

an interesting result as these players with lower gaming skill might typically need more help in such games. In the CS version, there was no difference in retention nor progression with and without hints for these low skill players.

To further understand the mixed results we got in the CS version, especially for Low skill players, we analyzed the qualitative questionnaire provided to the learners after playing the game (see section 4). This questionnaire was meant to get players' feedback on the game and its puzzles via open-ended questions about the gaming experience. Here, we focused only on the one question that inquired about what the players disliked, to get insights into possible discrepancies among the ACCESS and CS versions. We extracted the players' criticisms for each of the 17 puzzles (also called Enigma below), as reported next. Players of the CS version stated that in Room 2, they encountered significant challenges with two puzzles (Enigma08 and Enigma11) involving block-based programming. These puzzles were perceived as too complex and tedious, thus reducing their engagement.

"I didn't like the panda [Enigma11] too much, too complicated."

"I didn't like the one where you had to guide a cosmonaut [Enigma08] as it was really very complex."

"Little panda [Enigma11] and astronaut [Enigma08]: difficult if not focused."

Players also complained about three puzzles (Enigma08, Enigma11 and Enigma12) that entailed interacting with learning material on the web (BlocklyMaze for Enigma08 and Enigma11, and Compute-it for Enigma12), i.e., outside of the game, because switching between a web browser and the game multiple times was deemed too tedious.

"Switching between game and browser are complicated."

"Enigma on the mazes <https://blockly.games/> [Enigma08 and Enigma11] were initially a good idea for me, but are poorly implemented in my opinion."

"The panda enigma [Enigma11] wasn't very clear, so you might want to add a hint about that and make it clear that it's two different sites [Enigma08 and Enigma11]."

"The Blockly reference [Enigma08 and Enigma11] is nice, but not being able to switch easily from the web page to the game window is a pain [...] same for Compute-it [Enigma12]."

This criticism on Enigma08, Enigma11 and Enigma12 are aligned with the data. Figure 7 presents the percentage of puzzles resolved for each game version for Room 2. We observe a drop of about 10% in puzzle completion rate for Enigma12, Enigma08 and Enigma11 in the CS version as opposed to the ACCESS one. Thus, these findings suggest that the players' issues in Room 2 might have been not so much related to the hints themselves, but rather to the challenge of solving difficult puzzles requiring external resources outside of the game where hints cannot be delivered. This could also explain why low skill players had more difficulties in the CS version, as learning how to use external resources, while handling multiple windows using the hint panel might have overwhelmed them. This suggests to review these puzzles, for instance by integrating the external resources in the game and designing version-specific hints to help players solve them.

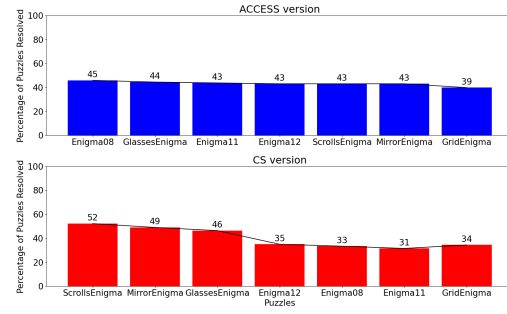


Figure 7: Histograms of puzzles' completion rates for each version in Room 2 sorted in descending order except for the GridEnigma which is necessary the last one in room 2.

The demographics of the CS and ACCESS players were also different, with CS players being younger and predominantly male, while the ACCESS' players were older and predominantly women (see Table 1). This difference in demographics could also have influenced the effectiveness of the hints, albeit we are unsure how, and ascertaining this could be the focus of future work.

7 CONCLUSION AND FUTURE WORK

In this work, we investigated the value of a automatic hint system in two version of E-LearningScape, a virtual educational escape games. These version (ACCESS and CS) featured the same puzzles, Dream Fragments and game mechanics, and just differ in their actual learning content, thus allowing us to examine whether our results can be replicated across several game versions. To evaluate the hint system, we designed a user study into the wild where learners could play either versions of the game on their own without the supervision of a researcher, to maintain high ecological validity. Learners were randomly assigned into two groups, one with the automatic hints enabled, and one without. We leverage the game data to formally compare these groups across game versions, while controlling for the gaming skills of the learners, via a statistical analysis.

Our finding revealed that the hint system based on resolution process modelling of puzzles with Petri net has proven to be highly effective in our context, particularly in the ACCESS version, as it significantly increased both player retention and progression. We also found that the hint system is especially useful to learners with high gaming skills in both game versions. Players with lower gaming skills benefited from the hint system too, but only in the ACCESS version, whereas in the CS version the hints had no impact on them.

The hint system does have its limitations, as it was overall not as effective in the CS version. One possible reason we highlighted in the discussion is that the more complex puzzles, as perceived by players in the CS version, had to be completed on the web outside of the game, where no hints can be delivered. This led to a significant drop in completion rates for these puzzles. Based on these findings, we advise to limit or avoid external puzzles altogether to ensure that players are supported throughout the entire game, as it was very much the case for the ACCESS version.

The originality of this work lies in experimenting with a hint system in two similar educational escape games in the wild. Hence, we did not investigate the effectiveness of the hints only in one context, but we replicated the experiment across two instances of the games. Our results show the importance of replicating such experiment, as it allowed to examine in more details what results could not be replicated across versions and why. These findings are especially interesting because, to the best of our knowledge, there has been no study that formally evaluated with end-users a hint system in virtual escape games so far. Therefore, in this paper, we contributed to deepen our understandings of how hints could benefit learners in this context, as well as discussed how to improve their deployment in actual games. In addition, for the sake of open science, we provided the code and data of our analysis to enable replication, as well as the hint mechanism in open source so that game designers could reuse it.

In future work, we first plan to reify the CS version as stated in the discussion section, to gauge whether getting rid of external resources would improve the effectiveness of the hint system. We also plan to study whether there are other factors, such as players' individual differences, that could interact with the value of the hints, similarly to what we found with gaming skills. This could lead to personalize the hints to the learners, e.g., by providing hints that would be more engaging to learners with low gaming skills who did not benefit as much from the hints.

REFERENCES

- [1] Fraguas-Sánchez Ana Isabel, Serrano Dolores R., and Elena González-Burgos. 2022. Gamification Tools in Higher Education: Creation and Implementation of an Escape Room Methodology in the Pharmacy Classroom. *Education Sciences* 12, 11 (2022), 833. <https://doi.org/10.3390/educsci12110833>
- [2] Erik Andersen, Eleanor O'Rourke, Yun-En Liu, Rich Snider, Jeff Lowdermilk, David Truong, Seth Cooper, and Zoran Popovic. 2012. The impact of tutorials on games of varying complexity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 59–68.
- [3] Yoav Benjamini and Yoel Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.
- [4] Andrew Thomas Bimba, Norisma Idris, Ahmed Al-Hunaiyyan, Rohana Binti Mahmud, and Nor Liyana Bt Mohd Shuib. 2017. Adaptive feedback in computer-based learning environments: a review. *Adaptive Behavior* 25, 5 (2017), 217–234. <https://doi.org/10.1177/1059712317727590>
- [5] Andrew Thomas Bimba, Norisma Idris, Ahmed Al-Hunaiyyan, Rohana Binti Mahmud, and Nor Liyana Bt Mohd Shuib. 2017. Adaptive feedback in computer-based learning environments: a review. *Adaptive Behavior* 25, 5 (2017), 217–234.
- [6] Anderson Pinheiro Cavalcanti, Arthur Barbosa, Ruan Carvalho, Fred Freitas, Yi-Shan Tsai, Dragan Gašević, and Rafael Ferreira Mello. 2021. Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence* 2 (2021), 100027.
- [7] Cristina Conati and Sébastien Lallé. 2023. Student modeling in open-ended learning environments. *Handbook of Artificial Intelligence in Education* (2023), 170–183.
- [8] Cristina Conati and Micheline Manske. 2009. Evaluating adaptive feedback in an educational computer game. In *Intelligent Virtual Agents: 9th International Conference, IVA 2009 Amsterdam, The Netherlands, September 14–16, 2009 Proceedings* 9. Springer, 146–158.
- [9] Tobias Drey, Fabian Fischbach, Pascal Jansen, Julian Frommel, Michael Rietzler, and Enrico Rukzio. 2021. To be or not to be stuck, or is it a continuum? a systematic literature review on the concept of being stuck in games. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–35.
- [10] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.
- [11] Clara Fernandez-Vara and Ira Fay. 2021. A Special Issue on Escape Rooms. Well Played.
- [12] Panagiotis Fotaris and Theodoros Mastoras. 2019. Escape Rooms for Learning: A Systematic Review. *Proceedings of the 12th European Conference on Game Based Learning* (2019). <https://api.semanticscholar.org/CorpusID:208072357>
- [13] Leon Freudenthaler and Michael Strommer. 2022. Towards the Implementation of a Digital Escape Room in a Higher Education (Java) Programming Course. In *International Conference on Interactive Collaborative Learning*. Springer, 959–970.
- [14] Aldo Gordillo, Daniel López-Fernández, Sonsoles López-Pernas, and Juan Quemada. 2020. Evaluating an Educational Escape Room Conducted Remotely for Teaching Software Engineering. *IEEE Access* 8 (2020), 225032–225051. <https://doi.org/10.1109/ACCESS.2020.3044380>
- [15] Ryan Hare and Ying Tang. 2022. Petri Nets and Hierarchical Reinforcement Learning for Personalized Student Assistance in Serious Games. In *Proceedings of the International Conference on Cyber-Physical Social Intelligence, ICCSI 2022 (Proceedings of the International Conference on Cyber-Physical Social Intelligence, ICCSI 2022)*, Xuemin Chen, Jun Wang, Jiacun Wang, and Ying Tang (Eds.). Institute of Electrical and Electronics Engineers Inc., United States, 733–738. <https://doi.org/10.1109/ICCSIS5536.2022.9970680> Publisher Copyright: © 2022 IEEE.; 2022 International Conference on Cyber-Physical Social Intelligence, ICCSI 2022 ; Conference date: 18-11-2022 Through 21-11-2022.
- [16] Danial Hooshyar, Liina Malva, Yeongwook Yang, Margus Pedaste, Minhong Wang, and Heuiseok Lim. 2021. An adaptive educational computer game: Effects on students' knowledge and learning attitude in computational thinking. *Computers in Human Behavior* 114 (2021), 106575.
- [17] Kedar Jambhekar, Rachel P Pahls, and Linda A Deloney. 2020. Benefits of an escape room as a novel educational activity for radiology residents. *Academic Radiology* 27, 2 (2020), 276–283.
- [18] Dominic Kao, Alejandra J Magana, and Christos Mousas. 2021. Evaluating tutorial-based instructions for controllers in virtual reality games. *Proceedings of the ACM on human-computer interaction* 5, CHI PLAY (2021), 1–28.
- [19] Nuntawut Kaoungku, Keerachart Suksut, Ratiporn Chanklan, Kittisak Kerdprasop, and Nittaya Kerdprasop. 2018. The silhouette width criterion for clustering and association mining to select image features. *International journal of machine learning and computing* 8, 1 (2018), 69–73.
- [20] Kenneth R Koedinger and Vincent Aleven. 2007. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review* 19 (2007), 239–264.
- [21] Sonsoles López-Pernas, Mohammed Saqr, Aldo Gordillo, and Enrique Barra. 2023. A learning analytics perspective on educational escape rooms. *Interactive Learning Environments* 31, 10 (2023), 6509–6525. <https://doi.org/10.1080/10494820.2022.2041045>
- [22] Jessica Middlemis Maher, Jonathan C Markey, and Diane Ebert-May. 2013. The other half of the story: effect size analysis in quantitative research. *CBE—Life Sciences Education* 12, 3 (2013), 345–351.
- [23] Helena Manojlovic. 2022. Escape room as a teaching method. *Opus et Educatio* 9 (08 2022). <https://doi.org/10.3311/ope.504>
- [24] Ana Manzano-León, José M Rodríguez-Ferrer, José Manuel Aguilar-Parra, Ana María Martínez Martínez, Antonio Luque de la Rosa, Dario Salguero García, and Juan Miguel Fernández Campoy. 2021. Escape rooms as a learning strategy for special education master's degree students. *International Journal of Environmental Research and Public Health* 18, 14 (2021), 7304.
- [25] Javier Melero, Davinia Hernández-Leo, and Josep Blat. 2012. Considerations for the design of mini-games integrating hints for puzzle solving ICT-related concepts. In *2012 IEEE 12th International Conference on Advanced Learning Technologies*. IEEE, 154–158.
- [26] Mathieu Muratet, Thibault Carron, and Amel Yessad. 2022. How to assist designers to model learning games with Petri nets?. In *Proceedings of the 17th International Conference on the Foundations of Digital Games (Athens, Greece) (FDG '22)*. Association for Computing Machinery, New York, NY, USA, Article 15, 10 pages. <https://doi.org/10.1145/3555858.3555937>
- [27] Mathieu Muratet, Amel Yessad, and Thibault Carron. 2016. Understanding Learners' Behaviors in Serious Games. In *Advances in Web-Based Learning – ICWL 2016*, Dickson K.W. Chiu, Ivana Marenzi, Umberto Nanni, Marc Spaniol, and Marco Temperini (Eds.). Springer International Publishing, Cham, 195–205.
- [28] Scott Nicholson. 2015. Peeking behind the locked door: A survey of escape room facilities. (2015).
- [29] Neil Peirce, Owen Conlan, and Vincent Wade. 2008. Adaptive educational games: Providing non-invasive personalised learning experiences. In *2008 second IEEE international conference on digital game and intelligent toy enhanced learning*. IEEE, 28–35.
- [30] Eric Sanchez and Maud Plumettaz-Sieber. 2019. Teaching and Learning with Escape Games from Debriefing to Institutionalization of Knowledge. In *Games and Learning Alliance*, Manuel Gentile, Mario Allegra, and Heinrich Söbke (Eds.). Springer International Publishing, Cham, 242–253.
- [31] Ana Moura Santos, Sofia Sá, Luis Felipe C Costa, and Luísa Coheur. 2021. Setting up educational escape games: Lessons learned in a higher education setting. In *2021 4th International Conference of the Portuguese Society for Engineering Education (CISPEE)*. IEEE, 1–8.
- [32] Valerie J. Shute. 2008. Focus on Formative Feedback. *Review of Educational Research* 78, 1 (2008), 153–189. <https://doi.org/10.3102/0034654307313795>
- [33] A.C. Siang and Radha Krishna Rao. 2003. Theories of learning: a computer game perspective. In *Fifth International Symposium on Multimedia Software Engineering*,

2003. *Proceedings*. 239–245. <https://doi.org/10.1109/MMSE.2003.1254447>
- [34] Lene Hayden Taraldsen, Frode Olav Haara, Mari Skjerdal Lysne, Pernille Reitan Jensen, and Eirik S Jenssen. 2022. A review on use of escape rooms in education—touching the void. *Education Inquiry* 13, 2 (2022), 169–184. <https://doi.org/10.1080/20004508.2020.1860284>
- [35] Kurt VanLehn, Collin Lynch, Kay Schulze, Joel A. Shapiro, Robert Shelby, Linwood Taylor, Don Treacy, Anders Weinstein, and Mary Wintersgill. 2005. The andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education* 15, 3 (2005), 147–204.
- [36] Alice Veldkamp, Johanna Rebecca Niese, Martijn Heuvelmans, Marie-Christine PJ Knippels, and Wouter R van Joolingen. 2022. You escaped! How did you learn during gameplay? *British Journal of Educational Technology* 53, 5 (2022), 1430–1458.
- [37] Alice Veldkamp, Liesbeth van de Grint, Marie-Christine P.J. Knippels, and Wouter R. van Joolingen. 2020. Escape education: A systematic review on escape rooms in education. *Educational Research Review* 31 (2020), 100364. <https://doi.org/10.1016/j.edurev.2020.100364>
- [38] Helen Wauck and Wai-Tat Fu. 2017. A data-driven, multidimensional approach to hint design in video games. In *Proceedings of the 22nd international conference on intelligent user interfaces*. 137–147.
- [39] Markus Wiemker, Errol Elumir, and Adam Clare. 2015. Escape room games. *Game based learning* 55 (2015), 55–75.
- [40] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.