Computers and Education: Artificial Intelligence 3 (2022) 100089

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

AI-Induced guidance: Preserving the optimal Zone of Proximal Development

Chris Ferguson, Egon L. van den Broek [*], Herre van Oostendorp

Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Artificial intelligence in education (AIEd)
Discovery learning
Textual specificity
Zone of proximal development (ZPD)
Personalization
Experience

ABSTRACT

Holding the promise of higher learning outcomes, discovery learning utilizes intrinsic motivation to provide an enjoyable self-directed learning experience. Unfortunately, this approach can also lead to a sub-optimal cognitive load, which hinders learning. To avoid this, players must be in the optimal Zone of Proximal Development (ZPD). A way of accomplishing this is to make use of Artificial Intelligence in a narrative-centered discovery game using adaptive guidance. Textual instructions were automatically adapted in real-time to ensure a personalized challenge for one group of learners, where a control group received static instructions. Compared to the control group, the learners with personalized instructions showed higher story and spatial learning, while having decreased cognitive load and a similar learning experience. So, instructions given to self-directed learners can be personalized in real-time, which not only reduces learners' cognitive load but also leads to enhanced learning outcomes without affecting the learning experience.

1. Introduction

Artificial Intelligence in Education (AIEd) has many shapes, such as Intelligent Tutoring Systems (ITS) (Conati, Barral, Putnam, & Rieger, 2021; Mousavinasab et al., 2021) and adaptive learning systems (Kabudi, Pappas, & Olsen, 2021; Tang, Chang, & Hwang, [in press]), and has provided new opportunities as well as offering the potential of revolutionizing education (Hwang, Xie, Wah, & Gašević, 2020; Ouyang & Jiao, 2021). In this setting, knowledge often needs to be made explicit, decision rules need to be transparent, variance among learners is significant and the amount of data is overwhelming (Graafland et al., 2014; Romero & Ventura, 2020; Zliobaite et al., 2012). Consequently, Artificial Intelligence (AI) current trend of subsymbolic machine learning (e.g., deep learning) are often not the best choice. Symbolic AI such as traditional, symbolic *intelligent agents* makes more sense.

In AIED, utility-based *intelligent agents* seem to be appropriate. In particular, utility-based agents take into account i) the environment, ii) the *educational goals*, and iii) a performance measure, which allows the agent to make decisions on how a goal can be achieved (Russell & Norvig, 2021). Such agents have a strong tradition in AIED, specifically in serious games (i.e., video games that serve an educational purpose (Caballero-Hernández, Palomo-Duarte, & Dodero, 2017; Doesburg, Heuvelink, & van den Broek, 2005; Geuze & van den Broek, 2011, chap. 8)), which have become more prominent in educational settings as complementary learning resources (Lacka, Wong, & Haddoud, 2021) and have shown that the knowledge transferred to learners can be retained for extended periods (Hu et al., 2021). One particular serious game genre, narrative-centered discovery games, could benefit greatly from a utility-based agent. This game genre combines two genres: narrative-centered games and discovery games.

Narrative-centered games provide a strong narrative through powerful, emotive storylines (Lester et al., 2014) and have been found to provide a motivating, engaging, and organized learning experience (Dickey, 2006; Lester et al., 2014; Marsh, 2010). In discovery games, learners freely explore the virtual environment within the game with minimal guidance whilst solving problems (i.e. completing in-game tasks by discovering useful items) (Toh & Kirschner, 2020), which takes a constructivist approach: learners actively construct representations of reality by linking new information to prior knowledge. Combining both game genres results in a self-directed learning experience that exploits intrinsic motivation, where a learner performs an activity for its own sake or enjoyment without reward (Liao, Chen, & Shih, 2019), which can lead to higher learning outcomes (Froiland & Worrell, 2016) (see Fig. 1).

However, using a self-directed learning approach means that learners may not discover important learning materials due to a lack of

* Corresponding author.
E-mail addresses: chris@chrisferguson.co.uk (C. Ferguson), vandenbroek@acm.org (E.L. van den Broek), h.vanoostendorp@uu.nl (H. van Oostendorp).

https://doi.org/10.1016/j.caeai.2022.100089
Received 14 July 2021; Received in revised form 7 July 2022; Accepted 8 July 2022
Available online 19 July 2022

knowledge of tasks and challenges that need to be solved (i.e. finding items related to educational content). This can lead to disorientation (i.e., poor navigational efficiency), where users lose their sense of location and direction (Head, Archer, & Yuan, 2000) because navigation is too much of a cognitive burden, exceeding the capacity of a learner's cognitive abilities (Chen & Ismail, 2008; Gwizdka & Spence, 2007). This is referred to as cognitive overload (Sweller, Ayres, & Kalyuga, 2011). Cognitive Load Theory (CLT) states that there are three different types of cognitive load: intrinsic, which is associated with the complexity of the task, extraneous, which is related to task presentation, and germane, which is produced by processing and constructing schemas to handle new information for learning new skills (Sweller et al., 2011). Therefore, the way a task is presented, in terms of instruction and navigation, has a large influence on the overall cognitive load so it is often extraneous cognitive load that is aimed to be reduced to allow more capacity for germane load (Sweller, 2016). Conversely, cognitive underload, excessively low cognitive load, has similar effects (Van der Sluis, Van den Broek, Glassey, Van Dijk, & de Jong, 2014) and recent research has found that extraneous load is necessary for learning success, especially in serious games (Skulmowski & Xu, 2022). Human-Computer Interaction (HCI) literature shows, for instance, that making the task easier often does not lead to better learning, or, conversely, making the task more difficult can even lead to better performance, at least when the player has sufficient relevant knowledge or skills (O'Hara & Payne, 1998; Svendsen, 1991; Trudel & Payne, 1995; van Nimwegen, 2008). Payne c.s. demonstrated that making the task more difficult with more implicit information (more extraneous load) can stimulate players to construct more reflective, plan-based strategies leading to better performance, compared to players who got an easy, explicit presentation of task-related information. Similar effects are also found in other areas than HCI. In our previous work, we found that reducing cognitive load by giving learners a 'guided tour' resulted in a lesser in-game experience, which may have also led to non-optimal learning (Ferguson, van den Broek, & van Oostendorp, 2020). For effective learning, there must be an appropriate level of cognitive load, the amount of working memory resources that are available for cognitive processing (Sweller et al., 2011). This is in line with research showing that full discovery learning is not as successful as expository learning but, rather, *guided discovery learning*, with appropriate guidance is more successful (de Jong & Lazonder, 2014).

We adopted the concept of the Zone of Proximal Development (ZPD) (Vygotsky, 1978) to describe the relationship between disorientation, cognitive load and learning, and the effects of appropriate guidance. The essence of ZPD is that optimal learning takes place when the space between what a learner can do without guidance and what a learner can do with guidance is at a minimum. ZPD puts forward that there are simple tasks that do little to challenge the learner (i.e., cognitive underload), leading to little learning. Conversely, there are more complex tasks that cannot be completed without assistance (i.e., cognitive overload), again compromising learning. The tasks in the center of the ZPD will provide an appropriate level of challenge without assistance being needed (i.e., an appropriate amount of cognitive load necessary for effective learning (Schnotz & Kürschner, 2007)), which can be referred to as the *"sweet spot"* where tasks are not too easy nor too difficult (Van der Sluis et al., 2014). Dependent on the proficiency of the learner on that moment during learning, a game can be made easier (by giving guidance) when needed or the opposite, making it more difficult (by providing extra challenges). In the virtual environment we use, navigation efficiency is crucial, and can form an important source of disorientation, hindering learning, by imposing extra cognitive load (Ferguson & van Oostendorp, 2020; Gwizdka & Spence, 2007). On basis of the navigation efficiency the intelligent agent will adapt the difficulty of the game in real-time.

Fig. 1. An example of a task being carried out in "The Chantry", a narrative-centered discovery game. This game tells the story of Dr. Edward Jenner, his invention of vaccination, and the smallpox virus it helped eradicate. The game involves exploring the house of the late Dr. Jenner using node-based movement, where accessible locations and items are predefined (Habgood, Moore, Wilson, & Alapont, 2018). When encountering a closed door, the learner is presented with a task, which represents a particular story topic and a list of descriptions of items that must be found. To open the doors and progress further, learners must discover and interact with the items referenced in the list, which contain further information on the story topic, in the form of an audio narrative.

Besides increased learning, being in the appropriate area of ZPD is also expected to lead to increased cognitive interest (Trif, 2015) (i.e., understanding topics and becoming more interested (Harp & Mayer, 1997), and higher engagement (Hamari et al., 2016) (i.e., heightened concentration, interest, involvement, and enjoyment (Kim, 2018)). This may also assist with maintaining a learner's sense of presence (i.e., feeling physically present in a virtual environment (Slater, 2002)), a key aspect related to the physically immersive nature of games, which can be broken when a learner feels like they require assistance (Steiner & Voruganti, 2004). For optimal learning, a personalized experience, which takes into account personal differences, should be provided to learners to ensure that they are within the correct area of ZPD. This is also referred to as adaptivity: adjustments of the game fitting to the proficiency of the learner that improve the learner's experience. (Lopes & Bidarra, 2011).

Adaptivity is part of the second paradigm of AIED: *AI-supported, learner-as-collaborator*, where the AI system serves as a supporting tool, with the learner collaborating to focus on the individual learner's learning process (Ouyang & Jiao, 2021). Using a learning theory, such as ZPD, as a base for an AI adaptive system is necessary for optimal learning as such learning theories are vital for explaining the learning process, learning outcomes, and evaluation methods for assessing the learner (Er et al., 2021; Radianti, Majchrzak, Fromm, & Wohlgenannt, 2020; Slussareff, Braad, Wilkinson, & Strååt, 2016). This would offer new opportunities for learning as well as new, possibly unobtrusive, techniques for learning assessment.

As previously mentioned, utility-based agents require two things: a goal to be achieved and a performance measure to make rational decisions on how this goal can be achieved. Based on the principles of ZPD, the goal of the agent is to ensure that a learner finds items and completes tasks in the optimal ZPD. To make rational decisions to achieve this goal, a quantitative measure of navigational efficiency will be used: the Lostness measure (Smith, 1996):

$$\sqrt{\left(\frac{N}{S}-1\right)^2+\left(\frac{R}{N}-1\right)^2}, \tag{1}$$

For an information-searching task (i.e., searching for and discovering an item), Lostness considers the minimum number of steps that are needed to be taken (R), the number of unique steps that a learner has taken (N), and the total number of steps that a learner has taken (S) (Smith, 1996). This will return a value between 0 and $\sqrt{2}$. 0 indicates that the task has been completed perfectly and the learner was not disoriented (high navigational efficiency) and $\sqrt{2}$ indicates that they were completely disoriented (low navigational efficiency) whilst completing the task. This measure is originally from the domain of hypertext; but, has shown its use in serious games as well (Ferguson, van den Broek, van Oostendorp, de Redelijkheid, & Giezeman, 2020). Moreover, this measure showed to be a highly predictive and valid indicator for learning (Ferguson & van Oostendorp, 2020).

The Lostness measure can be used to determine how and when to support a learner so that they complete a task in the optimal ZPD. In other words, the information obtained from the performance measure is used to determine when, and in what direction, to adapt to the learner. A way the agent can accomplish this adaptivity is through personalized automated real-time feedback, which is one of the merits of serious games, and computer-assisted learning in general. Personalized real-time feedback is an essential component of scaffolding for learning and has been shown to increase performance, aid learning outcomes, and improve knowledge construction (Cavalcanti et al., 2021; Shute, Ke, & Wang, 2017). This guidance can give learners additional information, which helps them access key knowledge for learning before they are *over*loaded (Kirschner, Sweller, & Clark, 2006). Conversely, this real-time feedback can also provide a greater challenge by providing less information to learners that are feeling *under*loaded.

More AI-enabled adaptive learning systems have emerged in recent years (Kabudi et al., 2021) and learning outcomes have been shown to be improved through such systems that utilize adaptive real-time feedback (Cavalcanti et al., 2021). These include Steiner and Voruganti (2004), who found that adding a trail of brightly colored stones was able to quickly and effectively guide learners to a target location, and Peirce, Conlan, and Wade (2008), who provided a companion, in the form of Galileo, to give learners hints throughout the game. These two examples, however, only assist struggling learners, which does not always lead to better learning (O'Hara & Payne, 1998; Svendsen, 1991; Trudel & Payne, 1995). One such AI-system that would meet the needs of both *over*loaded and *under*loaded learners was developed by Clark, Virk, Barnes, and Adams (2016). They manipulated the layout of tasks to provide appropriate personalized challenge depending on the learner's current in-game performance. A possible alternative to manipulating the layout of tasks would be to modify the educational instructions given to learners depending on their current performance, ensuring automatic presentation of appropriate personalized challenges.

In this paper, we will present a ZPD-based narrative-centered learning environment, which will automatically adapt and personalize educational instructions to ensure that learners are in the most optimum area of the ZPD with an appropriate amount of germane cognitive load. In the next Section, we will detail the theory behind using textual specificity to adapt educational instructions and how this is expected to aid learning. Section 3, the methods section, will describe the empirical study carried out to evaluate the proposed textual specificity-based adaptive system. This is followed by the results of this study in Section 4 before we end with Section 5, which provides a general discussion.

2. Automatic, personalized, adaptive specificity of educational instructions

There is scant literature available that mentions the use of information searching in video games (Beheshti, 2012), yet there is no literature available that utilizes textual specificity in this area. However, recent research by Albus, Vogt, and Seufert (2021) found that signaling, in the form of textual information, can increase learning performance and impact cognitive load. It is expected that using a similar system, which automatically adapts textual information, depending on the learner's in-game performance, could be even more beneficial and improve learning outcomes.

Words rapidly guide early visual processing (Boutonnet & Lupyan, 2015). Specific instructions or descriptions are more likely to lead to a successful search (Spärck Jones, 1988), whereas a decrease in specificity of instructions or descriptions makes this less likely (Wolfe, 2020). Following this observation, the utility-based agent adapts its instructions to either widens or narrows the learner's search space so that they remain in the center of the ZPD, resulting in higher levels of intrinsic motivation (Froiland & Worrell, 2016; Liao et al., 2019) with higher cognitive interest (Trif, 2015), engagement (Hamari et al., 2016), and presence (Kim, 2018). In this regard, if a learner performs poorly, the agent eases the process of finding educational content. In contrast, if the learners performs very well, it widens his/her search space by presenting less specific/generic instructions, making it more challenging to find educational content. If the learner is performing moderately, the agent will stick to the active difficulty level. In terms of CLT, the amount of germane cognitive load that a learner experiences will be impacted as this utility-based agent changes how the task is presented.

When carrying out an information-searching task, there are two different ways of increasing the chance of an item matching a request (i.e., an efficient search): adding terms to the search string or making use of more specific phrases, such as choosing the word 'tea' over 'beverage' (Spärck Jones, 1988). Contrarily, through the use of generic terms, one adds distractors, which can increase the task difficulty (Wolfe, 2020). Accordingly, it can be argued that describing an item more specifically (i.e. feature guidance) decreases the difficulty of an information-searching task and describing an item more generally increases difficulty (Wolfe, 2020). Guidance towards a location (i.e., scene guidance) that is more likely to contain the target item is also important in real-life search (Wolfe, 2020). Consequently, the description of both the item and the location is a key component of a successful search.

To provide an appropriate challenge for a wide range of learners, ensuring a personalized experience so that learners are within the optimal ZPD, an appropriate number of difficulty levels should be made available for the intelligent agent to provide to the learner. These will be defined along the two dimensions identified as being important for a successful search: the description of the item (feature guidance) and the description of the location of the item (scene guidance). Based on Fig. 1, Table 1 shows how learners are given more information to help them complete a task when they are not performing well and less information when performing very well.

For the utility-based agent to know when to modify the difficulty level (see Table 1), thresholds for the Lostness measure need to be predetermined to assess a learner's performance. In the Chantry (see Fig. 1), learners are given 11 tasks overall, with a median of 3 items. Before the start of a new task, a reliable impression of the learners' performance needs to be obtained. Therefore, the navigational efficiency in the previous 6 items was determined. To circumvent the cold start problem, we will make use of data collected from our previous research using this game. This dataset consists of data from 12 learners from the Virtual Reality (VR) condition in (Ferguson et al., 2020b) and 13 learners from an identical condition in (Ferguson et al., 2020a) was used, leading to 179 data points (mean: 0.427, SD: 0.160), following a normal distribution.

Two *k*-means cluster analyses were carried out on the data using both 3 clusters, (7 iterations, $F(2, 176) = 454.665$, $p < .001$) and 5 clusters (10 iterations, $F(4, 174) = 589.368$, $p < .001$). Based on the equality of clusters' data distribution, 3 clusters could be distinguished. The clusters' center-points were: 0.260 ($n = 63$), 0.454 ($n = 81$), and 0.665 ($n = 35$), which gave the following inter-cluster border values: 0.357 and 0.559. If the Lostness value of a learner drops below the lower threshold, the utility-based agent will increase the difficulty, as a player is deemed to be in the lower ZPD. Conversely, a player with a Lostness value above the upper threshold will be deemed to be in the upper ZPD so the utility-based agent will decrease the difficulty. A player with a Lostness value between these two thresholds will be considered in the optimal ZPD so the difficulty will not be modified.

Through adaptive real-time feedback, the utility-based agent uses textual specificity to provide a personalized experience for learners. Using the thresholds described, this should lead to the *"sweet spot"*, where each learner is given an appropriate level of challenge (Van der Sluis et al., 2014). To investigate whether such adaptive feedback is effective in terms of reducing cognitive load (Schnotz & Kürschner, 2007) and increasing learning (Shute et al., 2017), a study was executed on a VR narrative-centered discovery game.

Table 1
Five levels of specificity-based textual adaptation.

Difficulty Level	Information Made Available	Example
Very Easy	Specific location information Specific item description	A County Map on the Wall (Look Left)
Easy	Generic location information Specific item description	A County Map on the Wall
Normal	No location information Specific item description	A County Map
Hard	No location information Less Specific item description	Representation of the County
Very Hard	No location information Generic item description	The County

The effect that the adaptive system has on story and spatial knowledge will be investigated separately. Additionally, the learner's experience (i.e., presence, engagement, and cognitive interest) (Hamari et al., 2016) will be evaluated.

The following research question was examined:

Compared to a game providing static instructions, will a game featuring a utility-based agent, which provides adaptive instructions, result in less cognitive load and lead to better learning, as well as a better learning experience?

Based upon the above theoretical background, this initial question is decomposed into the following hypotheses:

Compared to a game providing static instructions, a game featuring a utility-based agent, which provides adaptive instructions, will lead to:

1. A lower amount of cognitive load;
2. Higher transfer of both story and spatial information;
3. Higher engagement and cognitive interest in the subject matter; and
4. A higher feeling of presence.

These hypotheses will be evaluated, in a randomized control study, using a commercially published serious game.

3. Methods

3.1. Material

As shown in Fig. 1, "the Chantry" (Steel Minions and the Jenner Trust, 2018) (https://jennermuseum.com/), a narrative-centered discovery game for PlayStation VR, will be used. The item descriptions given to players, which serve as educational instructions, will manipulated by the utility-based agent to provide different levels of textual specificity for these activities and the overall task.

Each learner wore a Sony PlayStation VR headset (model: CUH-ZVR1), connected to a PlayStation 4. They navigated using a standard PlayStation DualShock 4 controller (model: CUH-ZCT1) and wore noise-canceling over-ear headphones. Learners used their head movements to look at a movement/item node and used a single button press on the controller to jump to that node or pick up an item. After picking up an item, learners could move and rotate it by doing the same action holding on to the controller, as if they were holding the physical item.

3.2. Participants

A total of 40 learners, 22 males and 18 females, aged 21–55 (mean: 31.80, standard deviation (SD): 7.27), who were residents of Utrecht, the Netherlands were recruited to participate through email and online adverts. They had differing levels of VR experience, as indicated on a post-experiment 5-point Likert scale (from 1 = "Used very little" to 5 = "Use all the time", mean: 1.73, SD: 1.09). Informed consent was obtained along with information that would disqualify any of these learners from taking part, such as being susceptible to migraines or not having a professional comprehension of the English language. To avoid introducing biases, no participants were selected that had a background in health sciences or medicine. Moreover, as the in-game story is very specific, it is safe to assume participants had no prior knowledge of the topics within the game. Eligible learners were randomly assigned to either the adaptive or the non-adaptive group, both consisting of 11 males and 9 females.

3.3. Questionnaires

To evaluate learning, learners were provided with a short bespoke knowledge test, consisting of 24 true/false statements about the parts of the game to gauge how well that knowledge was transferred. 16 of these questions concerned facts related to the story and 8 involved spatial aspects related to the location of items/rooms in the game. This questionnaire was used in previous research Ferguson and van Oostendorp (2020), where we reported very high correlations between the two knowledge tests and the lostness measure, as well in two other empirical studies and it was sensitive enough to pick up effects of manipulations applied (see (Ferguson, van den Broek, van Oostendorp, de Redelijkheid, & Giezeman, 2020,a)).

Standard questionnaires consisting of 5-point Likert scales were used to measure engagement (Brockmyer et al., 2009), presence (Schubert, 2003), and cognitive interest (Schraw, Bruning, & Svoboda, 1995). For all these standard questionnaires acceptable levels of reliability and validity are reported in the literature.

To measure cognitive load, the NASA-TLX scale (Hart & Staveland, 1988) was used. This is a highly valid and reliable cognitive load questionnaire, consisting of 6 questions, which has been used in many studies on cognitive load (Lum, Greatbatch, Waldfogle, & Benedict, 2018; Xu, Liang, Zhang, & Baghaei, 2020). To avoid the questions becoming too complicated and time-consuming, along with results shown to be similar to the weighted version (Byers, Bittner, & Hill, 1989), the unweighted version of the questionnaire was used.

3.4. Procedure

Upon being seated, the learners were given instructions in both oral and written form. This involved safety information, such as not to attempt to physically grab anything and what to do in the event of motion sickness, along with their right to withdraw. This was followed by instructions on the game, including controls. The learners were then given 5 min to experience a simple controls tutorial before being instructed to play through the game for 30 min and learn about the story.

As each learner played the game, information on which tasks were started and completed, including found items and associated Lostness values, were logged. In the adaptive version, when a learner started a new task, their Lostness was compared to the thresholds and the difficulty (min: 0, max: 4) set, which remained constant for the duration of the task. For example, as shown in Table 1, a learner playing a game at normal difficulty would be given the description ‘a county map’. Increasing the difficulty would give the description ‘representation of the county’, giving less information to the learner in the form of a more generic item description. Conversely, decreasing the difficulty would give the description ‘a county map on the wall’, giving more information to the player in the form of a location. For the non-adaptive condition, the difficulty remained at the ‘normal’ level (2) for the whole game.

After 30 min, the game stopped, fading to black, and the learners were invited to complete the knowledge test and questionnaires. Once completed, game logs and questionnaire answers were exported and saved. Also, all learners were debriefed and informed about the nature of the study and our future plans.

4. Results

To determine the differences across conditions, adaptive and non-adaptive, a Multivariate Analysis of Variance (MANOVA) was performed after ensuring that the experiment dataset met the requirements of this analysis, such as following a normal distribution and the absence of multicollinearity. The presence/absence of adaptivity was the only independent between-subjects variable. Presence, engagement, cognitive interest, and the results on the knowledge test, separated into the story and spatial aspects, were examined as dependent variables, which all followed a normal distribution. The presence of adaptivity was found to have an overall significant effect ($F(6, 33) = 5.641, p < .001, \eta_p^2 = .506$). For univariate effects, it was found that cognitive load and knowledge transfer, of both story and spatial information, were affected by the presence of adaptivity at a significant level ($p < .050$, see Fig. 2). There were no significant effects found for presence, engagement, and

Table 2
Means, standard deviations (SDs), effect (*F*), significance (*p*) and ratio of variance (η_p^2) for the variables included in the MANOVA analysis.

	Non-Adaptive Mean (SD)	Adaptive Mean (SD)	F(6, 33)	p	η_p^2
Story Knowledge (0–1)	.625 (.111)	.706 (.084)	6.817	.013	.152
Spatial Knowledge (0–1)	.525 (.165)	.719 (.167)	13.636	.001	.264
Presence (1–5)	3.214 (.359)	3.136 (.375)	.458	.503	.012
Cognitive Interest (1–5)	3.760 (.692)	3.580 (.652)	.835	.367	.022
Engagement (1–5)	3.139 (.472)	2.934 (.334)	2.395	.130	.059
Cognitive Load (0–100)	54.208 (13.111)	40.750 (13.456)	10.264	.003	.213

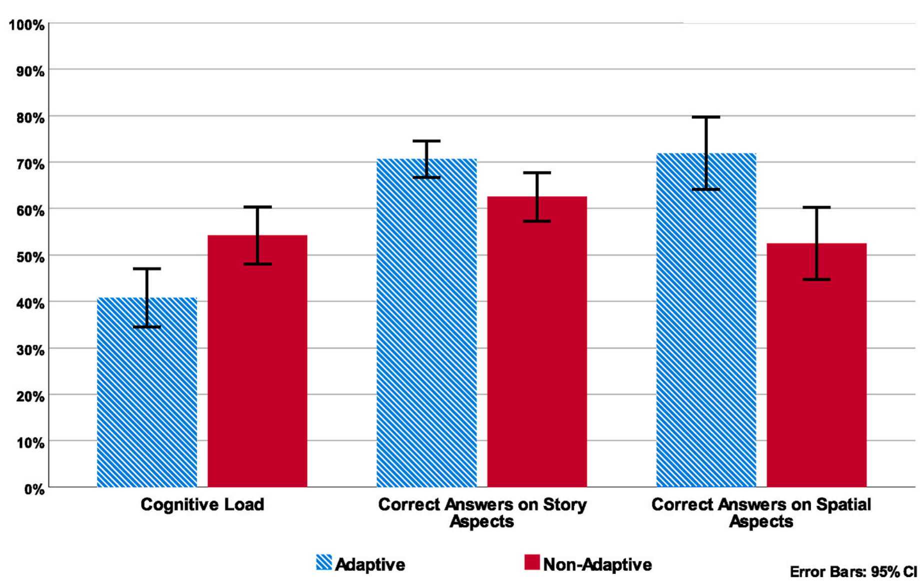

Fig. 2. Chart showing the means of cognitive load and knowledge transfer (story and spatial information) for learners in both the adaptive and non-adaptive condition.

cognitive interest. The full results are shown in Table 2.

As shown in Fig. 3, Lostness values over time showed a similar pattern in both conditions, which follows events that occur at these points. For example, some of the higher peaks correspond to the events that lead to more of the environment being unlocked, meaning that the learners needed to explore more to discover these new areas and the items within them. For the most part, the adaptive system was able to keep the average Lostness of the learners within the optimal range to avoid cognitive overload, this is also shown by the smaller variance in the results compared to when the adaptive system was not present. Overall, the mean Lostness, measured from the beginning of the condition until its end was .491 (SD: 0.139) for the adaptive and .549 (SD: 0.135) for the non-adaptive condition. No significant differences were expected between participants across conditions due to the adaptive system increasing the difficulty of tasks when navigational efficiency was high. This was confirmed through an independent-samples t-test, which found no significant difference in Lostness across conditions ($t(38) = 1.343$, $p = .187$).

Overall, 4 participants completed the adaptive version of the game (mean progress: 0.717, SD: 0.183) compared to 1 for the non-adaptive version of the game (mean progress: 0.628, SD: 0.241). However, an independent-samples t-test found no significant difference ($t(38) = 1.324$, $p = .173$).

For the adaptive condition, the average difficulty level for participants was 1.682 (SD: 1.030) with an average difficulty change of 0.087 (SD: 0.213) for each task.

5. Discussion

Artificial Intelligence in Education (AIEd) has many different forms and can have many different uses (Conati et al., 2021; Mousavinasab et al., 2021; Kabudi et al., 2021; Tang et al., [in press]), offering the potential to revolutionize education (Hwang et al., 2020; Ouyang & Jiao, 2021). We showed that AIED can help avoid eventual problems with cognitive overload and non-optimal learning during discovery learning. A utility-based agent was developed using the principles of the Zone of Proximal Development (ZPD) and Cognitive Load Theory (CLT) and applied to narrative-centered discovery learning with the aim of delivering an appropriate level of challenge to learners. The utility-based agent produced an adaptive system providing a personalized experience by giving learners automatic and personalized real-time feedback, presenting a task differently in the form of specific or generic descriptions for items to be found as part of a task so that learners are within the optimal ZPD and undergoing an appropriate amount of germane cognitive load. This feedback was in response to the learners' navigational efficiency obtained from the Lostness measure, which was used as a performance measure for the agent, as they were exploring the environment. The results showed that, compared to the non-adaptive version of the game, playing the adaptive version of the game led to: lower cognitive load (confirming hypothesis 1) and transfer retention of both story and spatial information (confirming hypothesis 2). Engagement, and cognitive interest, and presence were not affected (rejecting hypothesis 3 and hypothesis 4).

Cognitive load was significantly lower in learners playing the adaptive condition. Therefore, the support from the adaptive system (i.e., generic and specific search descriptions) led to these learners being in the appropriate area of the ZPD, avoiding disorientation and cognitive overload, which normally inhibits learning (Gwizdka & Spence, 2007).

It is interesting that this significantly lower average cognitive load occurred despite the additional challenge (i.e., by presenting generic descriptions) given to some learners, which is expected to lead to higher germane load and, consequently, overall cognitive load. On the one hand, it could simply be that being in the correct area of ZPD does, in fact, lead to higher learning with less effort, in the form of lower cognitive load. On the other hand, this would suggest that, overall, learners playing the adaptive version of the game received lower difficulty levels, which is backed up by the relatively low average difficulty of these learners. Furthermore, the high standard deviations showed that both the average difficulty and average change greatly varied for each participant, showing that participants varied in this aspect and each needed a different level of challenge. This shows the importance of personalization in games.

For knowledge retention, the adaptive condition led to better learning outcomes as shown by significant increases in knowledge retention for both story and spatial aspects in the form of higher knowledge test scores. It seems that the lower cognitive load, reported by learners in the adaptive condition, left cognitive capacity free for learning. Interestingly, this effect appeared to be stronger for spatial knowledge, compared to story knowledge. This could be due to the additional location information given by the adaptive system assisting with remembering the location of different objects.

Finally, it was shown that the experience aspects (i.e., engagement, cognitive interest, and presence) were not affected by using the adaptive system. It appears that the learners had the same experience in both conditions, although they were given more support in the adaptive condition. It was expected that this personalized level of challenge would lead to higher engagement and cognitive interest, in line with the findings of others (Hamari et al., 2016). As shown in Table 2, these values were already high in the non-adaptive condition so perhaps there was not much room for improvement. Moreover, regarding engagement in particular, the Game Engagement Questionnaire (Brockmyer et al., 2009), which was used to measure engagement, was developed with violent games in mind. Perhaps differences may have been found with an alternative questionnaire (e.g., Player Experience of Need Satisfaction (Ryan, Rigby, & Przybylski, 2006) or User Engagement Scale (Wiebe, Lamb, Hardy, & Sharek, 2014)). Presence was also not affected by the adaptation which was not as expected. Building a (spatial) mental model of a situation (e.g. of the Chantry) is expected to correlate to a sense of presence (Bailey & Witmer, 1994; Lee, Wong, & Fung, 2010; Slater, 2002). So we might expect that the support we provided would lead to construct an appropriate mental model (Wasserman & Banks, 2017) and, consequently, to a higher sense of presence (which appeared not to be the case). However, other recent research was also unable to find a significant relationship between presence and learning (Alsina-Jurnet & Gutiérrez-Maldonado, 2010; Coxon, Kelly, & Page, 2016; Ling, Nefs, Brinkman, Qu, & Heynderickx, 2013). In our own previous research using the same game and the same presence and learning measurements, we even found a (strong) negative correlation, where participants learned less when they felt more present in the game (Ferguson & van Oostendorp, 2020). These negative correlations between presence and learning are also found in studies by Schrader and Bastiaens (2012), Makransky et al. (2019), and Frederiksen et al. (2020). It is clear from these sets of conflicting outcomes that the role of presence in virtual environments needs more research.

Although the adaptive system was successful, the higher Lostness values, in certain areas, for learners in the adaptive condition (see Fig. 3) indicate potential issues with the adaptive system. The reason could be that the agent set the difficulty at the beginning of each task for each learner and this difficulty level did not change for the duration of the task. Therefore, in some cases, a learner may have been given a difficulty level that was too high, putting them in the upper area of ZPD, and this remained the case until the task was completed or they ran out of time. Unquestionably, the timing of feedback is an important issue of adaptive systems. Further research is needed to determine its optimal usage, particularly for adaptive personalized feedback.

To fully generalize these results, additional empirical studies should be executed. Firstly, further research should go into further detail into types of adaptivity to see whether or not the additional challenge, given by providing generic item descriptions, was actually beneficial or if the overall positive effect attributed to adaptivity is mainly due to the assistance given through specific item descriptions leading to a lower level of challenge. Interestingly, a recent study by Beege, Nebel,

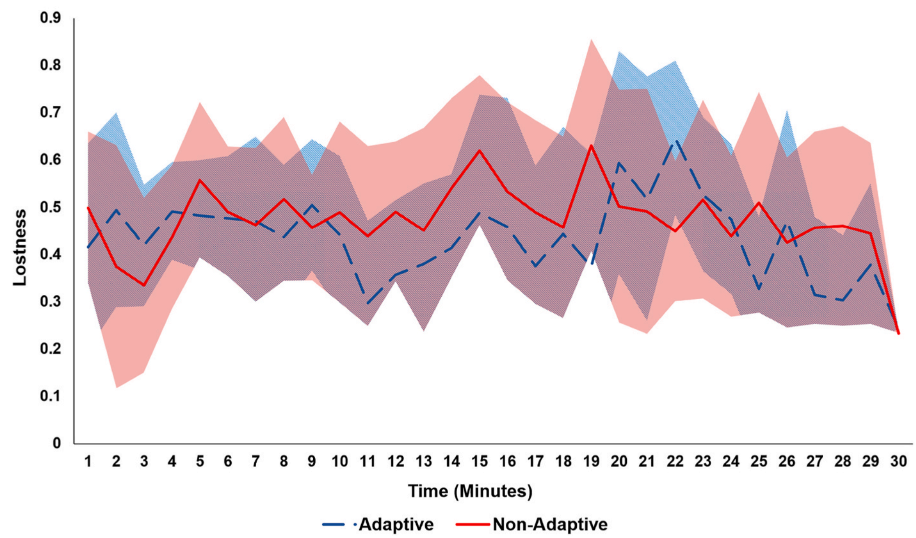

Fig. 3. Mean Lostness values from the previous six items found throughout both the adaptive and the non-adaptive condition. Shaded values indicate ± 1 SD.

Schneider, and Rey (2021) found that adding additional germane load did not influence learning outcomes, yet signaling, similar to the lower difficulty levels shown in this study, reduced cognitive load overall and led to higher learning outcomes. In addition, to fully evaluate the advantages and disadvantages of a textual specificity-based adaptive system, alternative forms of adaptivity could be investigated, such as guiding learners towards an item (Steiner & Voruganti, 2004) and more research on the validity of predictors for learning, similar to the Lostness measure, would also be beneficial to this area of research (Ferguson & van Oostendorp, 2020). This may include making use of temporal measures (i.e., task time) alongside the Lostness measure. Another area of interest could even be the effect of such an adaptive system outside of VR games, such as so-called 'scavenger hunts' inside museums or other tasks where information must be searched for in specific locations. Moreover, as utility-based agents are a type of symbolic AI, perhaps more complex solutions, such as sub-symbolic machine learning, could engender even better results.

In conclusion, inspired by the ZPD concept of Vygotsky (1978) and CLT of Sweller et al. (2011), we developed a utility-based agent, which created an adaptive system that provided a personalized experience through real-time feedback by adapting the level of challenge in narrative-centered discovery learning game based on a learner's navigation efficiency. This system modified the specificity of item descriptions to either increase or decrease the difficulty of a task so that each learner faced an appropriate level of challenge to their performance. Our results showed that such a system was highly successful and leads to lower levels of cognitive load, indicating that a learner is in the correct area of ZPD (i.e., the *"sweet spot"* (Van der Sluis et al., 2014)), resulting in increased learning outcomes for both story and spatial aspects. This increased learning was achieved with the same feelings of engagement, cognitive interest, and presence. The value of such personalized experiences is becoming more well-known and are even being introduced into real-world museums (Not & Petrelli, 2019). As more AI-enabled adaptive learning systems continue to emerge (Kabudi et al., 2021), with automated real-time feedback showing to be beneficial (Cavalcanti et al., 2021) follow-up research is encouraged in this area. Such research is crucial for the advancement of Artificial Intelligence in Education (AIEd) and the identification of other real-time learning analytics and monitoring variables, such as the Lostness measure of navigational efficiency.

CRediT authorship contribution statement

Chris Ferguson: Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing, Visualization, Project administration. **Egon L. van den Broek:** Methodology, Validation, Formal analysis, Writing, Supervision, Funding acquisition. **Herre van Oostendorp:** Conceptualization, Methodology, Validation, Formal analysis, Writing, Supervision, Project administration, Funding acquisition.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 732599, the REVEAL project. This study was carried out on the grounds of Utrecht University, Faculty of Science and approved by the faculty's Ethics Review Board.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education, 166*, Article 104154. https://doi.org/10.1016/j.compedu.2021.104154

Alsina-Jurnet, I., & Gutiérrez-Maldonado, J. (2010). Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments. *International Journal of Human-Computer Studies, 68*, 788–801.

Bailey, J. H., & Witmer, B. G. (1994). Learning and transfer of spatial knowledge in a virtual environment. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 38*, 1158–1162.

Beege, M., Nebel, S., Schneider, S., & Rey, G. D. (2021). The effect of signaling in dependence on the extraneous cognitive load in learning environments. *Cognitive Processing, 22*, 209–225. https://doi.org/10.1007/s10339-020-01002-5

Beheshti, J. (2012). Teens, virtual environments and information literacy. *Bulletin of the American Society for Information Science and Technology, 38*, 54–57. https://doi.org/10.1002/bult.2012.1720380313

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience, 35*, 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*, 624–634. https://doi.org/10.1016/j.jesp.2009.02.016

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw task load index (tlx) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics & safety* (pp. 481–485). London: Taylor & Francis.

Caballero-Hernández, J. A., Palomo-Duarte, M., & Dodero, J. M. (2017). Skill assessment in learning experiences based on serious games: A systematic mapping study. *Computers & Education, 113*, 42–60. https://doi.org/10.1016/j.compedu.2017.05.008

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., et al. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers & Education: Artificial Intelligence, 2*, Article 100027. https://doi.org/10.1016/j.caeai.2021.100027

Chen, C. J., & Ismail, W. M. F. W. (2008). Guiding exploration through three-dimensional virtual environments: A cognitive load reduction approach. *Journal of Interactive Learning Research, 19*, 579–596. URL: https://www.learntechlib.org/primary/p/24213/.

Clark, D. B., Virk, S. S., Barnes, J., & Adams, D. M. (2016). Self-explanation and digital games: Adaptively increasing abstraction. *Computers & Education, 103*, 28–43. https://doi.org/10.1016/j.compedu.2016.09.010

Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized xai: A case study in intelligent tutoring systems. *Artificial Intelligence, 298*, #–103503. https://doi.org/10.1016/j.artint.2021.103503

Coxon, M., Kelly, N., & Page, S. (2016). Individual differences in virtual reality: Are spatial presence and spatial ability linked? *Virtual Reality, 20*, 203–212.

Dickey, M. D. (2006). Game design narrative for learning: Appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. *Educational Technology Research & Development, 54*, 245–263. https://doi.org/10.1007/s11423-006-8806-y

Doesburg, W., Heuvelink, A., & van den Broek, E. L. (2005). Tacop: A cognitive agent for a naval training simulation environment. In M. Pechoucek, D. Steiner, & S. Thompson (Eds.), *Proceedings of the industry track of the fourth international joint conference on autonomous agents and multi-agent systems (AAMAS-05)* (pp. 34–41). New York, NY, USA: ACM. Utrecht, The Netherlands.

Er, E., Villa-Torrano, C., Dimitriadis, Y., Gasevic, D., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., et al. (2021). Theory-based learning analytics to explore student engagement patterns in a peer review activity. In *LAK21: 11th international learning analytics and knowledge conference* (pp. 196–206). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3448139.3448158.

Ferguson, C., & van Oostendorp, H. (2020). Lost in learning: Hypertext navigational efficiency measures are valid for predicting learning in virtual reality educational games. *Frontiers in Psychology, 11*, 3264. https://doi.org/10.3389/fpsyg.2020.578154

Ferguson, C., van den Broek, E. L., & van Oostendorp, H. (2020). On the role of interaction mode and story structure in virtual reality serious games. *Computers & Education, 143*, Article 103671. https://doi.org/10.1016/j.compedu.2019.103671

Ferguson, C., van den Broek, E. L., van Oostendorp, H., de Redelijkheid, S., & Giezeman, G. J. (2020). Virtual reality aids game navigation: Evidence from the hypertext lostness measure. *Cyberpsychology, Behavior, and Social Networking 0*. https://doi.org/10.1089/cyber.2019.0435. null.

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., et al. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: A randomized trial. *Surgical Endoscopy, 34*, 1244–1252. https://doi.org/10.1007/s00464-019-06887-8

Froiland, J. M., & Worrell, F. C. (2016). Intrinsic motivation, learning goals, engagement, and achievement in a diverse high school. *Psychology in the Schools, 53*, 321–336. https://doi.org/10.1002/pits.21901

Geuze, J., & van den Broek, E. L. (2011). E-learning through gaming: Unfolding children's negotiation skills. In S. Stankov, V. Glavinić, & M. Rosić (Eds.), *Intelligent tutoring systems in E-learning environments: Design, implementation and evaluation* (pp. 141–165). https://doi.org/10.4018/978-1-61692-008-1.ch008. Information Science Reference/IGI Global. Premier Reference Source. chapter 8, Hershey, PA, USA.

Graafland, M., Dankbaar, M., Mert, A., Lagro, J., De Wit-Zuurendonk, L., Schuit, S., et al. (2014). How to systematically assess serious games applied to health care. *JMIR Serious Games, 2*, e11. https://doi.org/10.2196/games.3825

Gwizdka, J., & Spence, I. (2007). Implicit measures of lostness and success in web navigation. *Interacting with Computers, 19*, 357–369. https://doi.org/10.1016/j.intcom.2007.01.001

Habgood, M. P. J., Moore, D., Wilson, D., & Alapont, S. (2018). Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique. In *2018 IEEE conference on virtual reality and 3D user interfaces* (pp. 371–378). VR). https://doi.org/10.1109/VR.2018.8446130.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170–179. https://doi.org/10.1016/j.chb.2015.07.045

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*, 92–102. https://doi.org/10.1037/0022-0663.89.1.92

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9. volume 52 of *Advances in Psychology*.

Head, M., Archer, N., & Yuan, Y. (2000). World wide web navigation aid. *International Journal of Human-Computer Studies, 53*, 301–330.

Hu, L., Zhang, L., Yin, R., Li, Z., Shen, J., Tan, H., et al. (2021). Neogames: A serious computer game that improves long-term knowledge retention of neonatal resuscitation in undergraduate medical students. *Frontiers in Pediatrics, 9*. https://doi.org/10.3389/fped.2021.645776

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers & Education: Artificial Intelligence, 1*, Article 100001. https://doi.org/10.1016/j.caeai.2020.100001

de Jong, T., & Lazonder, A. W. (2014). The guided discovery learning principle in multimedia learning. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (2 ed., pp. 371–390). Cambridge Handbooks in Psychology: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.019.

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). Ai-enabled adaptive learning systems: A systematic mapping of the literature. *Computers & Education: Artificial Intelligence, 2*, Article 100017. https://doi.org/10.1016/j.caeai.2021.100017

Kim, S. (2018). *Gamification in learning and education : Enjoy learning like gaming*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-47283-6

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86. https://doi.org/10.1207/s15326985ep4102_1

Lacka, E., Wong, T., & Haddoud, M. Y. (2021). Can digital technologies improve students' efficiency? Exploring the role of virtual learning environment and social media use in higher education. *Computers & Education, 163*, Article 104099. https://doi.org/10.1016/j.compedu.2020.104099

Lee, E. A. L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education, 55*, 1424–1442. https://doi.org/10.1016/j.compedu.2010.06.006

Lester, J. C., Spires, H. A., Nietfeld, J. L., Minogue, J., Mott, B. W., & Lobene, E. V. (2014). Designing game-based learning environments for elementary science education: A narrative-centered learning perspective. *Information Sciences, 264*, 4–18. https://doi.org/10.1016/j.ins.2013.09.005

Liao, C. W., Chen, C. H., & Shih, S. J. (2019). The interactivity of video and collaboration for learning achievement, intrinsic motivation, cognitive load, and behavior patterns in a digital game-based learning environment. *Computers & Education, 133*, 43–55. https://doi.org/10.1016/j.compedu.2019.01.013

Ling, Y., Nefs, H. T., Brinkman, W. P., Qu, C., & Heynderickx, I. (2013). The relationship between individual characteristics and experienced presence. *Computers in Human Behavior, 29*, 1519–1530.

Lopes, R., & Bidarra, A. (2011). Adaptivity challenges in games and simulations: A survey. *IEEE Transactions on Computational Intelligence and AI in Games, 3*, 85–99. https://doi.org/10.1109/TCIAIG.2011.2152841

Lum, H. C., Greatbatch, R., Waldfogle, G., & Benedict, J. (2018). How immersion, presence, emotion, & workload differ in virtual reality and traditional game mediums. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 62*, 1474–1478. https://doi.org/10.1177/1541931218621334

Makransky, G., Mayer, R. E., Veitch, N., Hood, M., Christensen, K. B., & Gadegaard, H. (2019). Equivalence of using a desktop virtual reality science simulation at home and in class. *PLoS One, 14*. https://doi.org/10.1371/journal.pone.0214944. e0214944–e0214944.

Marsh, T. (2010). Activity-based scenario design, development and assessment in serious games. *Gaming and Cognition, 213–226*. https://doi.org/10.4018/978-1-61520-717-6.ch010

Mousavinasab, E., Zarifsanaiey, N., Kalhori, S. R. N., Rakhshan, M., Keikha, L., & Saeedi, M. G. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29*, 142–163. https://doi.org/10.1080/10494820.2018.1558257

van Nimwegen, C. (2008). *The paradox of the guided user: Assistance can be counter-effective*. Ph.D. thesis. Universiteit Utrecht.

Not, E., & Petrelli, D. (2019). Empowering cultural heritage professionals with tools for authoring and deploying personalised visitor experiences. *User Modeling and User-Adapted Interaction, 29*, 67–120. https://doi.org/10.1007/s11257-019-09224-9

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology, 35*, 34–70. https://doi.org/10.1006/cogp.1997.0676

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers & Education: Artificial Intelligence, 2*, Article 100020. https://doi.org/10.1016/j.caeai.2021.100020

Peirce, N., Conlan, O., & Wade, V. (2008). Adaptive educational games: Providing non-invasive personalised learning experiences. In *2008 second IEEE international conference on digital game and intelligent toy enhanced learning* (pp. 28–35).

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, Article 103778. https://doi.org/10.1016/j.compedu.2019.103778

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery, 10*, e1355. https://doi.org/10.1002/widm.1355

Russell, S. J., & Norvig, P. (2021). Artificial intelligence: A modern approach. In *Pearson series in artificial intelligence* (4th ed.). Hoboken, NJ, USA: Pearson Education, Inc.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*, 344–360. https://doi.org/10.1007/s11031-006-9051-8

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*, 469–508. https://doi.org/10.1007/s10648-007-9053-4

Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior, 28*, 648–658.

Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior, 27*, 1–17. https://doi.org/10.1080/10862969509547866

Schubert, T. W. (2003). The sense of presence in virtual environments. *Zeitschrift für Medienpsychologie, 15*, 69–71. https://doi.org/10.1026//1617-6383.15.2.69

Shute, V., Ke, F., & Wang, L. (2017). Assessment and adaptation in games. In P. Wouters, & H. van Oostendorp (Eds.), *Instructional techniques to facilitate learning and motivation of serious games* (pp. 59–78). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_4.

Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review, 34*, 171–196. https://doi.org/10.1007/s10648-021-09624-7

Slater, M. (2002). Presence and the sixth sense. *Presence: Teleoperators and Virtual Environments, 11*, 435–439. https://doi.org/10.1162/105474602760204327

Slussareff, M., Braad, E., Wilkinson, P., & Strååt, B. (2016). Games for learning. In R. Dörner, S. Göbel, M. Kickmeier-Rust, M. Masuch, & K. Zweig (Eds.), *Entertainment computing and serious games: International GI-dagstuhl seminar 15283, dagstuhl castle, Germany, july 5-10, 2015, revised selected papers* (pp. 189–211). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46152-6_9.

Smith, P. A. (1996). Towards a practical measure of hypertext usability. *Interacting with Computers, 8*, 365–381. https://doi.org/10.1016/s0953-5438(97)83779-4

Spärck Jones, K. (1988). A look back and a look forward. In *Proceedings of the 11th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 13–29). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/62437.62438.

Steel Minions and the Jenner Trust. (2018). The Chantry. URL: https://store.playstation.com/en-gb/product/EP4325-CUSA12254%5F00-DRJENNERSMALLPOX.

Steiner, K. E., & Voruganti, L. (2004). A comparison of guidance cues in desktop virtual environments. *Virtual Reality, 7*, 140–147. https://doi.org/10.1007/s10055-004-0125-1

Svendsen, G. B. (1991). The influence of interface style on problem solving. *International Journal of Man-Machine Studies, 35*, 379–397. https://doi.org/10.1016/S0020-7373(05)80134-8

Sweller, J. (2016). Cognitive load theory, evolutionary educational psychology, and instructional design. In D. C. Geary, & D. B. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 291–306). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-29986-0_12.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory in perspective. In *Cognitive load theory* (pp. 237–242). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-8126-4_18.

Tang, K.Y., Chang, C.Y., Hwang, G.J., (in press). Trends in artificial intelligence–supported e-learning: A systematic review and co-citation network analysis (1998–2019). Interactive Learning Environments doi:10.1080/10494820.2021.1875001.

Toh, W., & Kirschner, D. (2020). Self-directed learning in video games, affordances and pedagogical implications for teaching and learning. *Computers & Education, 154*, Article 103912. https://doi.org/10.1016/j.compedu.2020.103912

Trif, L. (2015). Training models of social constructivism. teaching based on developing a scaffold. *Procedia - Social and Behavioral Sciences, 180*, 978–983. https://doi.org/10.1016/j.sbspro.2015.02.184

Trudel, C. I., & Payne, S. J. (1995). Reflection and goal management in exploratory learning. *International Journal of Human-Computer Studies, 42*, 307–339. https://doi.org/10.1006/ijhc.1995.1015

Van der Sluis, F., Van den Broek, E. L., Glassey, R. J., Van Dijk, E. M. A. G., & de Jong, F. M. G. (2014). When complexity becomes interesting. *Journal of the Association for Information Science and Technology, 65*, 1478–1500. https://doi.org/10.1002/asi.23095

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Wasserman, J. A., & Banks, J. (2017). Details and dynamics: Mental models of complex systems in game-based learning. *Simulation & Gaming, 48*, 603–624. https://doi.org/10.1177/1046878117715056

Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the user engagement scale. *Computers in Human Behavior, 32*, 123–132. https://doi.org/10.1016/j.chb.2013.12.001

Wolfe, J. M. (2020). Visual search: How do we find what we are looking for? *Annual Review of Vision Science, 6*, 539–562. https://doi.org/10.1146/annurev-vision-091718-015048

Xu, W., Liang, H. N., Zhang, Z., & Baghaei, N. (2020). Studying the effect of display type and viewing perspective on user experience in virtual reality exergames. *Games for Health Journal, 9*, 405–414. https://doi.org/10.1089/g4h.2019.0102

Zliobaite, I., Bifet, A., Gaber, M., Gabrys, B., Gama, J., Minku, L., et al. (2012). Next challenges for adaptive learning systems. *SIGKDD Explor. Newsl, 14*, 48–55. https://doi.org/10.1145/2408736.2408746