Digital Learning Games in Artificial Intelligence in Education (AIED): A Review

Bruce M. McLaren and Huy Nguyen
Human-Computer Interaction Institute, Carnegie Mellon University
Pittsburgh PA 15213

Keywords: Digital learning games, artificial intelligence, learning technology, natural language processing

1. Introduction

Digital and computer games have captured the attention and imagination of people around the world. Lobel and colleagues (2017) report that Dutch children (7 to 12 years old) play digital games between 4.9 and 5.8 hours per week, while a slightly older age range of children in New York City (10 to 15 years old) has been reported as playing even more: between 30 and 42 hours a week (Homer et al., 2012). Digital game play has also been significantly on the rise: According to a 2019 survey by the NPD Group (NPD, 2019), 73 percent of Americans aged 2 and older play digital games, a 6% increase from the prior year. There are reports of more than 2.6 billion people world-wide being video game players, with an expected rise to over 3 billion people by 2023 (Gilbert, 2021).

This general interest in digital games has also transferred to schools and educational use of computer-based games. Digital learning games, also called educational games, began to be designed, developed and sporadically appear in classrooms by the late 1990s (Zeng et al., 2020). A flourishing in digital learning games has occurred since 2000, boosted by concurrent advancements in supporting technology, such as computer graphics and faster processors. President Obama helped to push this interest forward by launching the National STEM Game Design Competition in 2010. Today, according to Juraschka (2019), 74% of teachers use digital game-based learning to enhance their lessons and the majority of teachers who use games in the classroom believe games have been helpful in improving their students' learning.

In addition to the increasing interest in and playing of digital learning games, there is also increasing empirical evidence of the effectiveness of games in helping people learn. Academics became interested in the motivational potential of digital learning games some time ago (Castell & Jenson, 2003; Gros, 2007), and an increase in empirical research has followed over the past 15 years or so. Meta-analyses over the past five years have uncovered and discussed the educational benefits of digital learning games (Clark et al., 2016; Crocco et al., 2016; Ke, 2016; Mayer, 2019; Tokac et al., 2019; Wouters & van Oostendorp, 2017). Even video games without a specific aim to provide educational benefits have been shown to increase students' skills, for instance, perceptual attention (Bediou et al., 2018). The Clark et al. (2016) meta-analysis identified that careful attention to game design – the way in which a game's interface and mechanics supports interaction with the student player – can result in positive learning results. Furthermore, there is strong evidence that designing digital games based on cognitive theories of learning and empirical learning science results can lead to educational benefits. For instance, Parong and colleagues have shown that executive function skills can be trained through a digital learning game designed expressly for that purpose (Parong et al., 2017; 2020). McLaren and colleagues found that designing a game explicitly targeted at mathematics misconceptions and employing self-explanation prompts – a technique that has been shown to lead to learning benefits in a variety of contexts (Chi et al., 1994; 1989; Mayer & Johnson, 2010; Wylie & Chi, 2014) – can lead to learning benefits (McLaren et al., 2017).

While it is clear that game design and attention to what we know about how humans learn has been beneficial to the design and success of digital learning games, what about the use of artificial intelligence (AI) in the design and development of digital learning games? It seems natural to insert AI into digital learning games to make them more realistic, more challenging, and more adaptive to students' skill level and style of play. Yet, a recent meta-analysis of digital learning games notes that over the past 20 years AI has rarely been cited as a component of learning games (Schöbel et al., 2021). Thus, a natural question that arises is: Has the field of Artificial Intelligence in Education (AIED) actually made an impact on learning with digital games? In this chapter we explore this question by discussing the way that AI has been used in digital learning games until now, as well as how it might provide even more benefit to learning with digital learning games in the future.

2. Foundations of Learning From Games

Between the mid 1970s and 1990 Czikszentmihalyi developed and described the theory of *flow*, a state of optimal experience, where a person is so engaged in the activity at hand that self-consciousness disappears, a sense of time is lost, and the person engages in complex, goal-directed activity (often with immediate feedback) not for external rewards, but simply for the exhilaration of doing (Czikszentmihalyi, 1990; 1975). For over 20 years Czikszentmihalyi had been studying "states of optimal experience" – times when people, while undertaking an engaging activity, report feelings of concentration and deep enjoyment. Flow induces focused concentration and total absorption in an activity. Everyone experiences flow from time to time and will recognize its characteristics: one feels strong, alert, in effortless control, loses all self-consciousness, and is at the peak of their abilities. Often, the sense of time seems to disappear, and a person in flow experiences a feeling of transcendence.

Digital games often immerse children – and people more generally – in flow and have long been posited to help in the learning process (Gee, 2003). Some very early researchers in the area of cognitive science identified constructs that are often part of games and appear to promote flow. For instance, Malone (1981) identified fantasy, curiosity, and challenge as key to intrinsic motivation and learning with games. Fantasy can serve to connect the player with content that they might otherwise reject as conflicting with their identity (Kaufman & Flanagan, 2015). Curiosity can be promoted when learners have the sense that their knowledge needs to be revised, for example, if it is incomplete or inconsistent. Challenge depends upon activities that involve uncertain outcomes, hidden information or randomness. While this theoretical work pre-dated most of present-day digital learning game research, all of the constructs explored by Malone are clearly relevant and important to learning with games.

While flow and intrinsic motivation appear to be key to learning with digital games, states of human affect, such as determination, confusion, frustration, and boredom, also play an important role (Loderer et al., 2019). For example, even though frustration is a "negative" emotion, it could indicate that a student is highly engaged while playing a challenging learning game (Gee, 2003). Determination and curiosity have also been found to be strongly present during game play with learning games (Spann et al., 2019). A line of AIED research that has investigated affect in gameplay has thus emerged, most typically involving data mining of log files of student use of learning games (Baker et al., 2007; 2010; Shute et al., 2015) or using sensors, video, and/or eye trackers to detect affect (Bosch et al., 2016; Conati & Gutica, 2016; Shute et al., 2013).

Given this foundational theoretical (and some empirical) work, a number of researchers and proselytizers have strongly pushed digital learning games as a panacea to the many shortcomings of education in today's world. In the early days, many claims were made about the benefits of learning with digital learning games versus more traditional approaches (Gee, 2007; Prensky, 2006; Squire & Jenkins, 2003). Not long after the claims were made, however, others emphasized the lack of evidence for positive learning outcomes with digital learning games (Honey & Hilton, 2011; Mayer, 2014; Tobias & Fletcher, 2011). Yet, as pointed out above, in recent years evidence has started to accumulate that digital learning games *can* be beneficial to learning (Clark et al., 2016; Mayer, 2019), including substantial evidence from within the AIED community (Arroyo et al., 2013; Easterday et al., 2017; Lee et al., 2011; McLaren et al., 2017; Sawyer et al., 2017; Shute et al., 2015). How the AIED community has explored the space of learning from digital games is discussed in the next section.

3. Digital Learning Games Research in AIED

For the purposes of this review chapter it is important to define, first, what a digital learning game is and, second, when a digital learning game is an "AIED" learning game. Given that many digital learning games are arguably not "AIED," two separate definitions are necessary.

Note that many before us have made attempts to define what a "game" is (Rollings & Morris, 2000; Salen & Zimmerman, 2003) and, in turn, what a "digital learning game" is (Mayer, 2014; Prensky, 2004). While it is impossible to precisely define these terms, and any definition is subject to dispute, the following components have typically been part of prior definitions of a digital learning game: (1) an interactive program running on a computer or other electronic device; (2) "game play" in the form of an artificial environment in which fun, challenge, and/or fantasy are involved; (3) instructional content or an instructional objective is an integral part of the game play; (4) entertainment goals are part of the game play (e.g., competition, having fun); and, finally, (5) a set of pre-defined rules guide game play. Given this background, our working definition of a *digital learning game* is:

> *A <u>digital learning game</u> is an interactive, computer-based system in which users (i.e., players) engage in artificial activities involving fun, challenge, and/or fantasy, instructional and entertainment goals are part of the system, and pre-defined rules guide game play.*

In turn, our definition of an *AIED digital learning game* is:

> *An <u>AIED digital learning game</u> is a digital learning game that (1) employs AI within its operation and interaction with players and/or (2) has been developed and/or extended using AI techniques (e.g., educational data mining, learning analytics, or machine learning)*.

Note that this definition implies that we include games that may not have been published within the annual AIED conference proceedings or the International Journal of AI in Education. That is, we focus more on how games operate, the extent to which AI is part of their game mechanism or post-use analyses, rather than whether they have been published within the AIED community. That said, we also focus on games that have been subject of some study of their efficacy and for which articles have been published. Note further that we include in our review AI learning technology that is "gamified," i.e., technology that was not (necessarily) developed originally as a game, but that includes gaming elements such as badges, points, leaderboards, and interactive playful agents (Landers & Landers, 2014; Landers et al., 2017). These elements may alter the game's mechanics, the player's interactions, or aspects of player immersion and emotion, to improve the learner's engagement and experience (Deterding et al., 2011). Such technology falls, we believe, within the above definition of an AIED digital learning game. Key examples of gamified AIED learning technology are *MathSpring* (Arroyo et al., 2013; 2014), *iStart-2* (Jackson & McNamara, 2011; Jacovina et al., 2016), *Gamified Lynnette* (Long & Aleven, 2014; 2018), and *Gamified SQL-Tutor* (Tahir et al., 2020).

In what follows, we explicitly call out the AI aspect of the games we cite and discuss. In the interest of inclusiveness, recency and broad coverage, we are also somewhat liberal in including games that are nascent, without much (or any) empirical evidence of their instructional effectiveness, especially more recently developed and tested games that may be of interest to the AIED audience (e.g., *Navigo* - Benton et al., 2021; *TurtleTalk* - Jung et al., 2019). On the other hand, we are clear about the games that have been the subject of extensive empirical study (e.g., *MathSpring* - Arroyo et al., 2013; 2014; *Crystal Island* - Lester et al., 2013; *Decimal Point* - McLaren et al., 2017; *Physics Playground* - Shute et al., 2015; 2021).

There are many ways that AIED-based digital learning games can be categorized, for instance, by whether they are pure games or "gamified", by their instructional topic, by the AI techniques used, by the degree of learning impact they've had, etc. To provide a structure to the review of this chapter, we present AIED games according to the four major ways in which AI has been used in the context of the learning games:

1. *AI-based Adaptation*: Digital learning games that employ AI to perform adaptation in real-time during play. That is, games that provide customized support (e.g., hints and error messages, problems appropriate to a student's current level of understanding, difficulty adjustment) to help students solve problems and learn (Martin et al., 2021).
2. *AI-based Decision Support*: Digital learning games that feature AI-powered interactive dashboards or recommendations; that is, these are games that don't make decisions for students, such as presenting the next problem or step to take, but instead present options, based on students' on-going performance.
3. *AI Character Interaction*: Games that employ an AI-driven non-player character (or characters) (NPC) to support the learning process with the games. These types of games rely on a "companion" to support students as they learn.
4. *Use of Learning Analytics (LA) and/or Educational Data Mining (EDM) for Game Analysis and Improvement:* These are AIED games that don't explicitly use AI for game play or game mechanics, but instead employ AI to do post-game analysis. This analysis is typically done for one of two reasons: to better understand how students interact with the game or to iteratively improve the game.

We also present the games according to their general instructional domain (i.e., math, computer science, natural science, humanities). Table 1 summarizes the prominent AIED digital learning games that are reviewed in this chapter. We identified these games by using three prominent databases: Google Scholar, Springer and Elsevier. In each database, we entered the search query as a combination of three terms: (1) the learning domain, corresponding to a row in Table 1 (e.g., "math," "computer science"), the type of AI, corresponding to a column in Table 1 (e.g., "adaptation," "interaction"), and (3) the keyword "game." While Table 1 provides a substantial number of digital learning games (more than 30), it is likely missing some games that could have been included, i.e., it is meant to be a close-to (but perhaps not quite) comprehensive literature review[1]. In addition, as we focus our search dimensions on the learning domain and type of AI, we note that the reviewed games vary broadly in other characteristics, such as maturity, sample size and research findings.

Table 1: Summary of AIED digital learning games

	AI-based Adaptation	***AI-based Decision Support***	***AI Character Interaction***	***Use of Learning Analytics (LA) and/or Educational Data Mining (EDM) for***

[1] The authors apologize for any AIED digital learning games that may have been overlooked in this review.

				Game Analysis and Improvement
Math	*Maths Garden* (Klinkenberg et al., 2011) *MathSpring* * (Arroyo et al., 2014; 2013) *Prime Climb* * (Conati et al., 2013)	*Decimal Point* * (Harpstead et al., 2019; Hou et al., 2020; 2021; McLaren et al, 2017) *Gamified Lynnette* (Long & Aleven, 2014; 2018)	*MathSpring* * (Arroyo et al., 2013; Arroyo et al., 2014) *Squares Family* (Pareto, 2009; 2014; Sjödén et al. 2017)	*Battleship Numberline* (Lomas et al., 2013; 2012; 2011) *Decimal Point* * (Nguyen et al., 2019; 2020) *Heroes of Math Island* (Conati & Gutica, 2016) *Prime Climb* * (Conati & Zhou, 2002) *Refraction* (O'Rourke et al., 2016; 2015; O'Rourke, Ballweber et al., 2014; O'Rourke, Haimovitz et al., 2014) *Reasoning Mind* (Ocumpaugh et al., 2013) *ST Math* (Peddycord-Liu et al., 2017) *Zombie Division* (Baker et al., 2007; Habgood & Ainsworth, 2011)
Computer Science	*AutoThinking* * (Hooshyar et al., 2021) *Gamified SQL-Tutor* (Tahir et al., 2020) *Minerva* (Lindberg et al., 2017; 2018)	*TALENT* * (Maragos, 2013)	*AutoThinking* * (Hooshyar et al., 2021) * *ELIA* (Kaczmarek & Petroviča, 2018) *TALENT* * (Maragos, 2013); *TurtleTalk* (Jung et al., 2019)	*Zoombinis* (Rowe et al, 2020; 2021)
Natural Science	*ELEKTRA* (Peirce et al., 2008)	*Physics Playground* *	*Betty's Brain* * (Biswas et al.,	*Beanstalk* (Aleven et al., 2013;

	Physics Playground * (Shute et al., 2021)	(Shute et al., 2019; Shute, 2011)	2016) *Crystal Island* * (Lester et al., 2013)	Harpstead & Aleven, 2015) *Betty's Brain* * (Kinnebrew et al., 2017; Munshi et al., 2018; Segedy et al., 2015) *Crystal Island* * (Sabourin et al., 2013; Sawyer et al., 2017) *Physics Playground* * (Shute et al., 2015)
Humanities	*iStart-2* (Jackson & McNamara, 2011; Jacovina et al., 2016) *Navigo* (Benton et al., 2021) *Policy World* (Easterday et al., 2017)	*Keep Attention* (Hocine, 2019; Hocine et al., 2019) *Tactical Language and Culture Training System* * (Johnson, 2010)	*ECHOES* (Bernardini et al. 2014) *Tactical Language and Culture Training System* * (Johnson, 2010)	*Downtown: A Subway Adventure* (Cano et al., 2018; 2016) *TC3Sim* (Henderson et al., 2020a; 2020b)

*** - A digital learning game that is in more than one of the AI categories**

3.1. AI-based Adaptation

Perhaps the most common way that AI is and has been used in AIED digital learning games is by adapting games to individual students and their learning progress. Usually, this means adapting the difficulty of a game, problems within the game, or hints to optimize the game's learning potential (Martin et al., 2021). This approach is derivative of what has long been done with intelligent tutoring systems (VanLehn, 2006; 2011), using approaches such as Bayesian Knowledge Tracing (Corbett & Anderson, 1995) and item response theory (Elo, 1978; Embretson & Reise, 2000).

Perhaps it is unsurprising that adapting instruction for individual students is a key research area of AIED digital learning games, given how much attention this type of computer-based instruction has been given in AIED research since its earliest days (Self, 2016) and until recently (see the chapter by Aleven, Mavrikis et al. in this volume). As John Self, one of the founders of the AIED research community, reported in his summary of the history of the field:

> *"AIED systems were, almost by definition, the only ones that carried out a significant, real-time analysis of the interaction with learners, in order to adapt that interaction. Other systems claimed to be adaptive, but they were really only reacting in pre-specified ways to different inputs. AIED systems responded in ways that had not been prespecified or even envisaged. And that, of course, is the essential difference between AI programs and general computer programs."* (Self, 2016, p. 9)

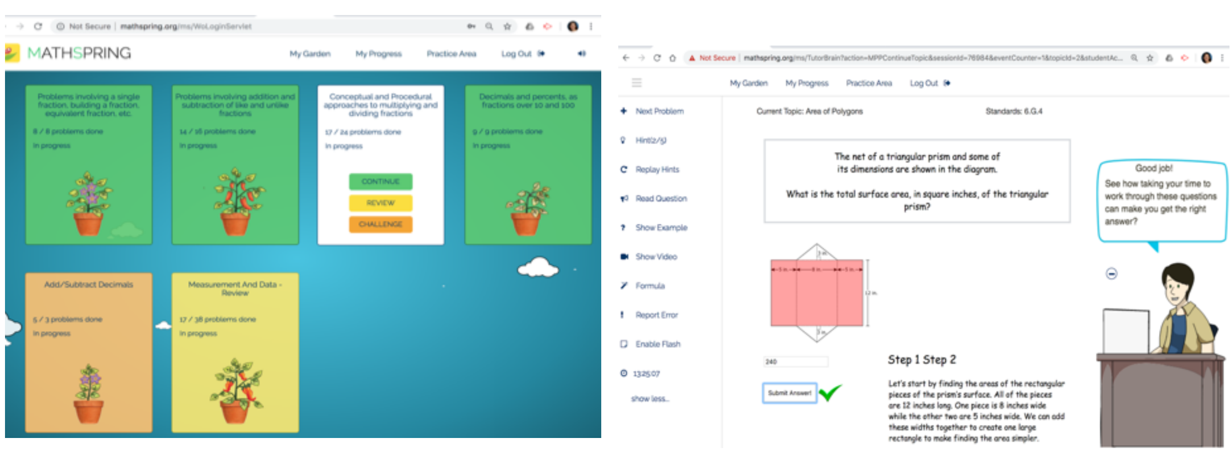
In fact, what Self says is essentially the difference between many of the games described in this chapter and, for instance, the games that have come from non-AIED research (e.g., Mayer, 2019). On the other hand, it is important that the emphasis is not only on the AI in games, but, like the many games that Mayer (2019) discusses, that AIED digital learning games are proven in well-designed, randomized controlled experiments.

In what follows, we first describe in detail two representative “AI-based adaptation” games from Table 1 – one in mathematics (*MathSpring*) and one in humanities (*Policy World*). We also discuss how AI has specifically been used for adaptation in those games, as well as the empirical results that studies with these games have uncovered. Finally, we present and discuss in a more succinct fashion other “AI-based adaptation” games and summarize what we have learned about AI-based adaption games.

MathSpring (Arroyo et al., 2013; 2014) is a single player, online game for 4th through 7th grade students to practice math problem solving (Figure 1(a)). The game, which was developed as a gamified extension to the tutoring system *Wayang Outpost*, adapts problem difficulty to student performance and offers hints, feedback, worked-out examples, and tutorial videos. The game also provides socio-emotional support from “learning companion” characters (hence this game is also found in the “AI Character Interaction” category). The game supports strategic and problem-solving abilities based on the theory of cognitive apprenticeship (Collins et al., 1989). The software models solutions via worked-out examples with the use of sound and animation and provides practice opportunities on math word problems.

MathSpring uses an AI-based adaptation algorithm to maintain students within their zone of proximal development (Vygotsky, 1978); in particular, the algorithm adapts problem difficulty. The *MathSpring* approach, referred to as “effort-based tutoring” (EBT), adapts problem selection depending on the effort exerted by a student on a practice activity based on three dimensions of student behavior: attempts to solve a problem, help requested, and time to answer. In addition, student affect is automatically predicted while students play the game. Initially, this was achieved through information from physiological sensors and student behavior within the game. Later, machine-learned detectors were created to predict student emotions. The game also uses AI to drive a learning companion, an animated digital character that speaks to the student (Figure 1(b)), deemphasizing the importance of immediate success and instead encouraging effort.

MathSpring has been used in middle and high schools in the U.S. as part of regular math classes since 2004, in some instances just before students take the Massachusetts statewide-standardized test exams. In a variety of studies involving hundreds of students, the game has led to improved performance in mathematics and on the state standardized tests. It has also led to improved engagement and affective outcomes for groups of students as a whole, as well as for certain subgroups, e.g., female students and low achieving students.



(a) (b)

Figure 1(a) and 1(b): In *MathSpring* students use math to grow plants representing progress and effort in the game (1(a)). Plants might bloom and give peppers, or wither if students show disengagement. Figure 1(b) shows an example of an AI-driven learning companion in *MathSpring*, used to encourage student's effort. (Figures provided by Ivon Arroyo, reproduced by permission)

Policy World (Easterday et al., 2017) is a digital learning game targeted at university students in which the player assumes the role of a policy analyst (Figure 2(a)) who must defend the public against an unscrupulous corporate lobbyist (Figure 2(b)) by persuading a Senator (Figure 2(c)) to adopt evidence-based policies that protect the public interest. The game has two modes of operation: game only and game+tutor, the second of which includes a back-end intelligent tutoring system that provides step-level feedback and immediate error correction. The narrative of Policy World emphasizes empowerment: a young policy analyst (i.e., the student playing the game) is recognized as having potential by the head of a policy think-tank. The student is guided by two mentor characters: another young but more senior analyst; and a sharp-tongued virtual tutor that teaches the student to analyze policies (Figure 2(d)). At the end of the game the player is tested through simulated senate hearings. The player must debate two policies with the corporate lobbyist to save the think-tank's reputation and defend the public against the corrupt agenda.

Policy World provides error flagging, situational feedback, and penalties for errors. The game+tutor version, which is the one that uses AI techniques for adaptation, is the same in all regards except that, as mentioned above, it also includes step-level feedback and immediate error correction (as in most intelligent tutoring systems - see VanLehn, 2006; 2011).

A total of 105 university students were recruited for a study with *Policy World*. Based on the results of a series of ANOVAs, the game+tutor version resulted in more learning of policy analysis skills and self-reported competence, compared to the game-only version. A path analysis supported the claim that the greater assistance provided by the game+tutor helped students learn analysis better, which increased their feelings of competence, which in turn increased their interest in the game.

(a)

(b)

(c) (d)

Figures 2(a), 2(b), 2(c), and 2(d)*: Policy World* screenshots of the agents involved in the policy debate: (a) the player; (b) the lobbyist; (c) the senator; and (d) the tutor (Figures from Easterday et al., 2017, reproduced by permission from the first author)

As shown in Table 1, other AI-based adaptive games include *Maths Garden*, *Prime Climb, AutoThinking, Minerva,* the *Gamified SQL Tutor, ELEKTRA, iSTART-2, and Navigo*. Mathematics is a key focus area of games in this category, with *Maths Garden* (Klinkenberg et al., 2011) as a prime example. *Maths Garden* is a single player game used to help K-8 students learn mathematics operations, such as whole number addition, subtraction, multiplication and division. Students click on different flower beds to try out the different mathematical operations and gain points (coins) by answering prompted questions correctly and lose coins by answering incorrectly. The game reward comes in two forms: first, the flowers in the various flower beds grow as the student makes progress; second, the coins earned from the math tasks can be used to purchase virtual prizes*. Maths Garden* is distinguished as an AIED digital learning game by adapting its problem content within each garden, using an item response model based on the Elo (1978) rating system. The game has been empirically tested with over 3,500 participants and results indicate that the Elo scoring model is highly correlated with an independent math test and students become highly engaged with the game. Another AI adaptive digital learning game that is targeted at math learning is *Prime Climb* (Conati et al., 2013). *Prime Climb* is a collaborative, two-player learning game in which 5th and 6th grade students practice number factorization by climbing a series of “number mountains,” composed of numbered hexagons. Players move to numbers in the hexagons that do not share common factors with their partner’s number. The two players rely on and cooperate with one another to reach the top of the mountain. Each player can make one or more moves before turning the control to the other player. The game is adaptive through its hints, which are provided to assist the students in climbing the mountain. Using a probabilistic student model, the game predicts when a student doesn’t have the factorization skill required for a particular move. The game gives hints at incremental levels of detail. Conati et al’s most important finding was that students with a positive attitude towards help tend to pay attention to hint content after correct moves, while students with a negative attitude towards help tend to pay attention to hints after incorrect moves, and students with a neutral attitude towards help show limited attention to hints.

A second key focus of AIED digital learning games in this category has been computational thinking and computer programming. For instance, *AutoThinking* (Hooshyar et al., 2021) is a recent adaptive, single player digital learning game designed to promote elementary-age students’ skills and conceptual knowledge in computational thinking (CT). In this game, the player takes the role of a mouse in a maze seeking cheese and evading two cats. Players write “programs” using icons representing program steps to create solutions to evade the cats. One of the two cats is “intelligent,” the other is random. After the player completes a solution, the “intelligent cat” adapts using student log data and a Bayesian Network algorithm that decides which algorithm it should use next to pursue the player and, if necessary, what kind of feedback or hints to provide to the player. In a comparison to a more conventional, computer-based approach to learning CT, *AutoThinking* was found to be especially helpful to students with lower prior knowledge. Hooshyar and colleagues also found that *AutoThinking* improved students’ attitude toward CT more than the conventional approach. Due to the AI-based cat, *AutoThinking* also falls into the “AI Character Interaction” category, which we introduce later in the chapter. A second CT example is *Minerva* (Lindberg et al., 2018; 2016) a single player game designed to teach programming to elementary school students, covering five concepts: input, output, math, loop and condition. In the game, players control a robot to navigate puzzles and repair a damaged ship, while avoiding aliens and other obstacles. The puzzles require players to learn the different programming concepts to achieve their task. The adaptation of Minerva is to students’ “learning styles,” a concept that is controversial, since students having different learning styles has been largely debunked in various learning science literature (Pashler et al., 2008). Despite the perhaps misguided focus of this research, it is another example of "AI-based Adaptation" and a programming focused game. As a final example of games focused on computational thinking and programming, the *Gamified SQL Tutor* (Tahir et al., 2020) includes game elements beyond the well-known, adaptive intelligent tutoring system, *SQL Tutor*, which teaches students the Standard Query Language (SQL). The *SQL Tutor* has been shown in many studies to lead to learning benefits (Mitrovic, 2012). The *Gamified SQL Tutor* was an attempt to gain additional learning benefits through adding "badges" to the tutor related to goals, assessment, and challenges. For instance, if students complete three problems in one session or five problems in one day, they receive goal badges. In a study with 77 undergraduate students, Tahir and colleagues found that while *Gamified SQL Tutor* didn't lead to better learning outcomes than *SQL Tutor*, time on task was found to be a significant mediator between badges and achievement in the gamified condition. This suggests that badges *can* motivate students to spend more time learning with the *SQL Tutor*.

AI-based adaptive games have also been developed to support reading (*iSTART-2* - Jackson & McNamara, 2011; Jacovina et al., 2016; *Navigo* – Benton et al., 2021) and science learning (*ELEKTRA* - Peirce et al., 2008). The highlights of these games include that *iSTART-2* (previously called *iStart-ME* - Jackson & McNamara, 2011) uses AI natural language techniques to assess student self-explanations and adapt game play and feedback accordingly; *Navigo* relies on AI-based rules to ensure the learner is at an appropriate reading level within the game, to be sure the language the student encounters is diverse; and that the student generally progresses towards reading fluency. In *ELEKTRA,* which is targeted at 13-to-15-year-old students, AI rules are also used to assess a student's holistic game experience and to execute adaptation that is both pedagogically helpful and non-invasive.

So what have we learned thus far from the research with AI-based adaptation of digital learning games? First, as mentioned above, it is clear that research with intelligent tutoring systems (VanLehn, 2006; 2011) has created a blueprint and paved the way for how adaptation has been implemented, at least thus far, in many digital learning games. The focus on adapting problems and feedback to the level of understanding or skill exhibited by a student is, unsurprisingly, core to most of the games in this category. Some of the most successful work in the "AI-based Adaptation" category of learning games, such as with *MathSpring* and *iStart-2,* essentially started as research with intelligent tutoring systems that later shifted into research with gamified intelligent tutors. Second, it is encouraging to see AI game adaptation be successfully applied in a wide variety of domains, including mathematics (e.g., *MathSpring*, *Maths Garden*), science (e.g., *ELEKTRA*), language learning (e.g., *iStart-2, Navigo*), computer science (e.g., *AutoThinking, Minerva*) and policy analysis (e.g., *Policy World*). This suggests that adapting instructional content is not constrained to any particular domain or game type. Finally, while adaptive digital learning games appear to be largely focused on elementary to middle school age students – for instance, with the games *MathSpring*, *Minerva, Prime Climb,* and *AutoThinking* – the games within this category show that adaptive games can also be effective with older students, for instance, as shown with *Policy World* and *Gamified SQL Tutor*.

3.2. AI-based Decision Support

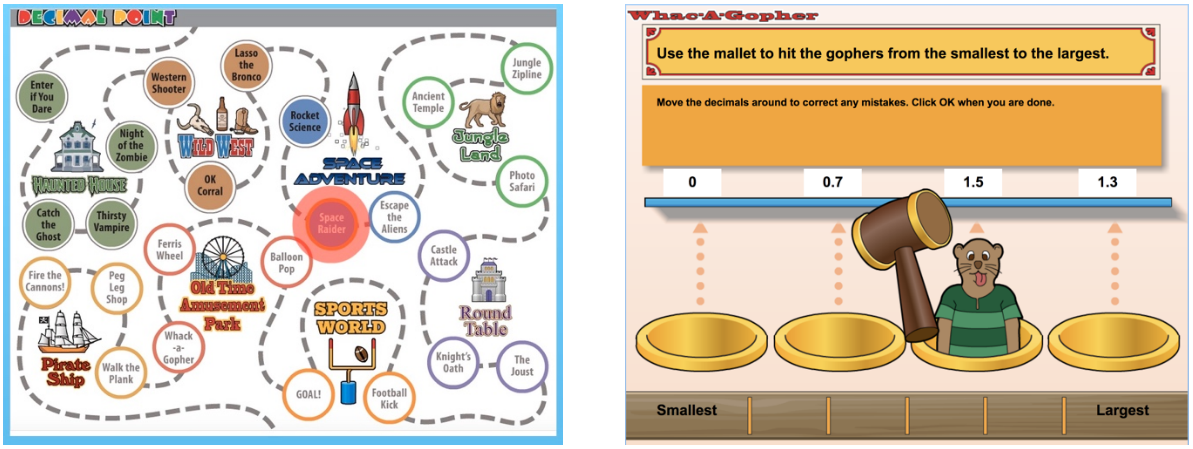
AI-based decision support is another relatively common approach found in AIED digital learning games. As opposed to "AI-based Adaptation," these games allow the student to make their own choices of problems to solve and game paths to follow, but with support from an AI recommender system. Often these types of games employ a dashboard or open learner model (OLM - Bull & Kay, 2008; Bull, 2020) to provide the choices to students, as well as data supporting their options. The games in this category are often intended to support and explore self-regulated learning (SRL - Zimmerman & Schunk, 2008), in particular, prompting students to carefully consider their learning trajectories and move thoughtfully forward in the game and in their learning.

As in the previous section, we first describe in detail two representative “AI-based Decision Support” games – one in the domain of mathematics (*Decimal Point*) and one in science (*Physics Playground*). We also discuss how AI has been used for decision support in these games, as well as empirical results from studies with the games. We then summarize and discuss other “AI-based Decision Support” games and, finally, discuss what we have learned thus far from research with “AI-based Decision Support” games.

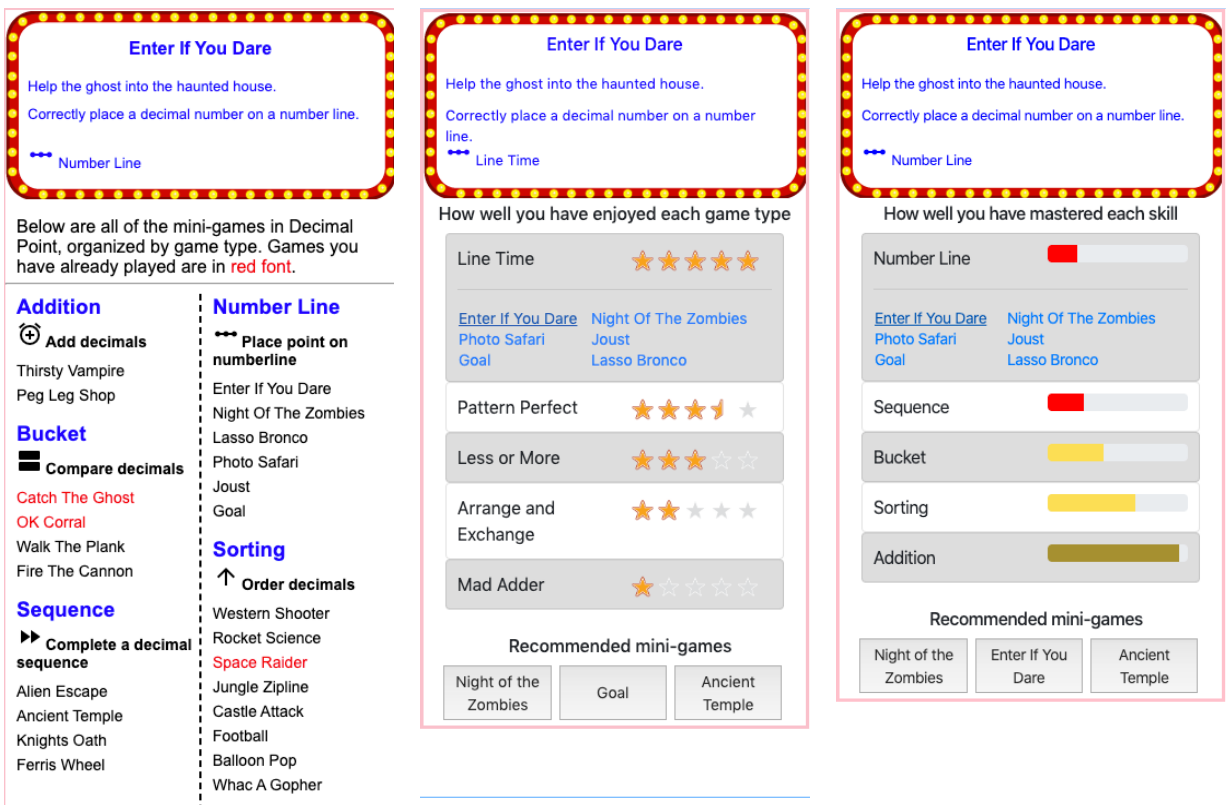
Decimal Point (Harpstead et al., 2019; Hou et al., 2020; 2021; McLaren et al., 2017; Nguyen et al., 2018) is a web-based, single-player digital learning game that helps 5th through 7th grade students reinforce their knowledge about decimal numbers and decimal operations. The game features an amusement park metaphor with eight theme areas and 24 mini-games (Figure 3(a)) that target common decimal misconceptions. The mini-games support playful problem-solving activities (e.g., entering a haunted house, shooting objects in the wild west), each of which connects with one of five decimal exercise types: number line, addition, sequence, bucket, and sorting. As an example, in the mini-game *Whack-A-Gopher* (Figure 3(b)), students need to correctly “whack” the gophers, who pop up and retreat at random times, in the order of their number labels. Students receive immediate feedback on their answers and play until they get the correct answer. After each mini-game, students are prompted to answer a self-explanation question to reinforce their understanding (Chi et al., 1994; 1989; Wylie & Chi, 2014). There have been a variety of studies of *Decimal Point* over the years, exploring issues such as student agency (Nguyen et al., 2018), indirect control (Harpstead et al., 2019), and adaptive recommendations (Hou et al., 2020; 2021).

One study with *Decimal Point* included an AI-based recommender system that recommends the next mini-game to select from the game map to maximize either learning or enjoyment (Hou et al., 2020; 2021). In the learning-oriented version of *Decimal Point*, the student’s mastery of each decimal skill is assessed in real time with Bayesian Knowledge Tracing (BKT - Corbett & Anderson, 1995). Students see their mastery visualized through an open learner model (Bull, 2020) and are recommended to play more mini-games in the two least mastered skills (Figure 3(c)). In the enjoyment-oriented version of the game, the student’s rating of each mini-game round is collected after they finish it, and students are recommended to play more mini-games of the type they like the most (Figure 3(d)). *Decimal Point* data has also been used in EDM research that uncovers students’ learning difficulties (Nguyen et al., 2020), as well as the relationship between in-game learning and posttest / delayed posttest performance (Nguyen et al., 2021). Thus, the *Decimal Point* game is also found in the “Use of Learning Analytics *...”* category of AIED digital learning games.

A media comparison study (Mayer, 2019) with the original version of the *Decimal Point* game showed a strong learning benefit over an intelligent tutoring system with the same content (McLaren et al., 2017). For the version that included AI-based recommender discussed above, 196 5th and 6th grade students participated in a study (Hou et al., 2020; 2021). Three game conditions were examined: a learning-focused condition featuring the BKT-based dashboard (Figure 3(c)), an enjoyment-focused condition featuring the playful dashboard (Figure 3(d)), and a control condition with a neutral dashboard (i.e., a dashboard that does not present either skill mastery or enjoyment scores to the student) (Figure 3(e)). Results from the study indicated that the students in the enjoyment-focused group learned more efficiently than the control group, and that females had higher learning gains than males across all conditions. Post hoc analyses also revealed that the learning-focused group re-practiced the same mini-games, while the enjoyment-focused group explored a wider variety of mini-games. These findings suggest that adaptively emphasizing learning or enjoyment can result in distinctive gameplay behaviors from students, and that *Decimal Point* can help bridge the gender gap in math education.



(a) (b)



(c) (d) (e)

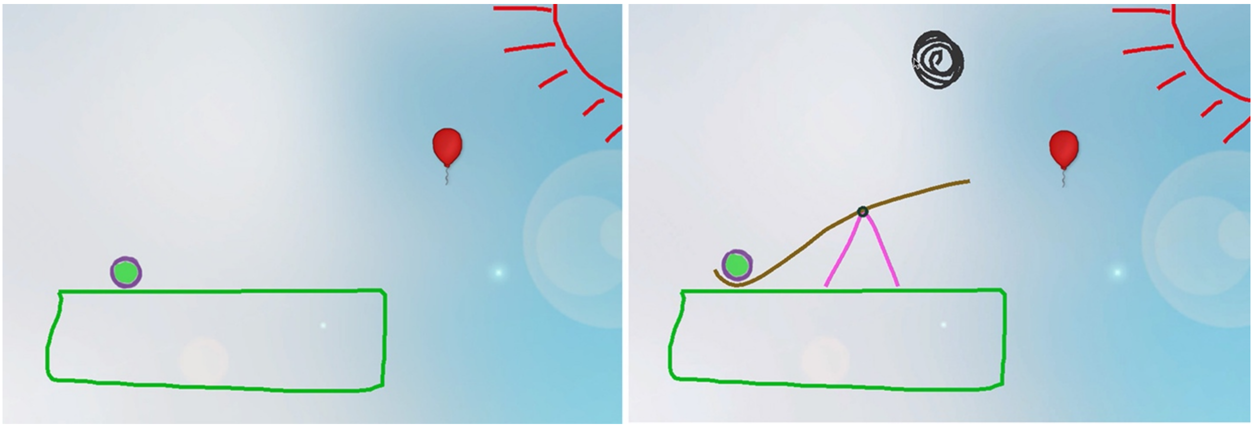
***Figures 3(a), 3(b), 3(c), 3(d), and 3(e)**:* The main game map on the top left allows to students to see the 24 mini-games of Decimal Point (a), an example mini-game on the top right in which the student "whacks" moles in the order of smallest to largest decimal number (b), while the recommender dashboard for the control version of the game (c), the enjoyment-oriented version of the game (d) and the learning-oriented version of the game (e) are shown left to right at the bottom. (Figures from Hou et al., 2021, reproduced by permission from the authors)

Physics Playground (PP - Shute et al., 2019; 2021; Shute, 2011) is a digital learning game designed to enhance the physics understanding of middle to high school students. *PP* was originally designed to be a single player game but has more recently been converted and also used in a collaborative learning mode

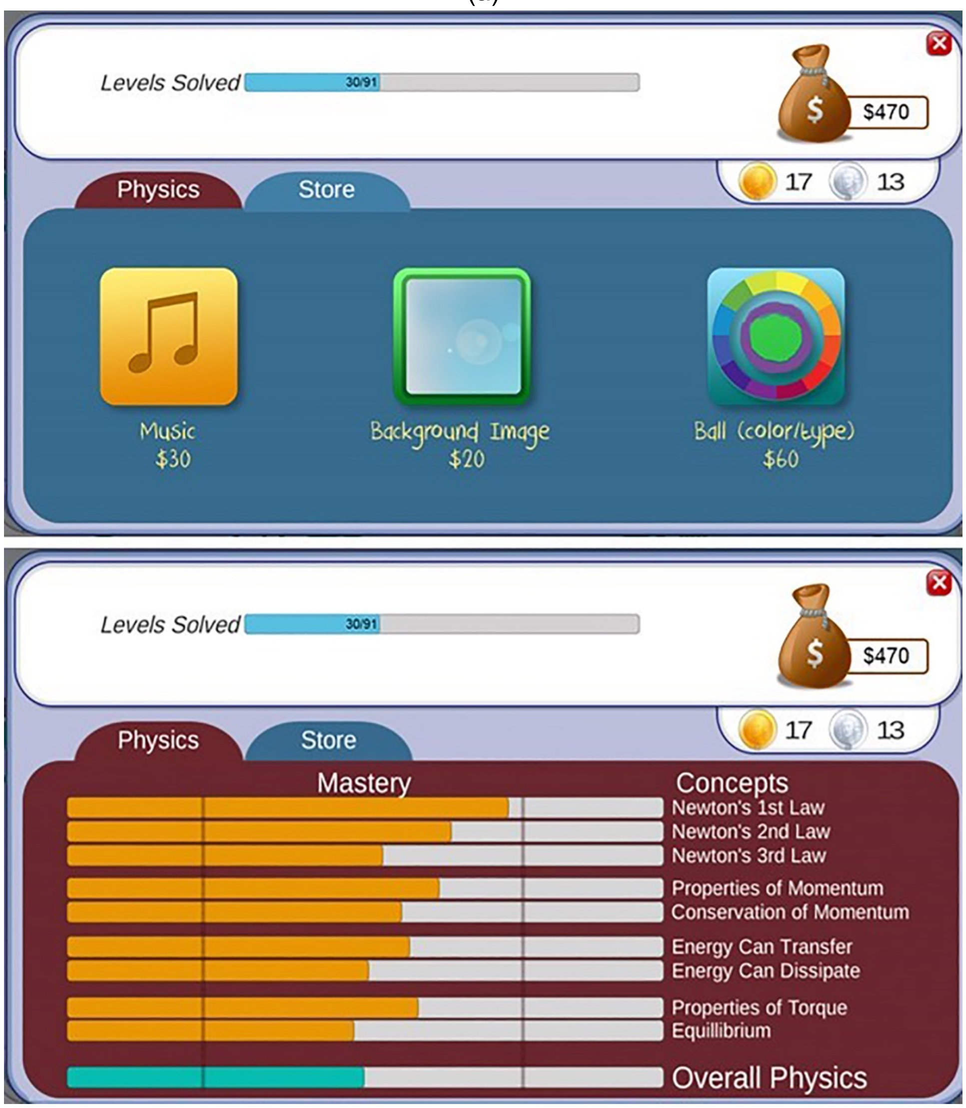
(Sun et al., 2020; 2022). The goal of students using *PP* is simple—hit a red balloon using a green ball and various provided tools. In its first version, *PP* included only one type of game level: sketching. Using a mouse or stylus, players could draw objects on the screen, create simple "machines" (i.e., ramp, lever, pendulum, or springboard), and target the red balloon with the green ball. The second and most recent version of the game incorporates a new task type, manipulation levels, where drawing is disabled with players instead move the ball to the balloon using new tools, including (1) sliders related to mass, gravity, and air resistance, (2) the ability to make the ball bounce by clicking the bounciness checkbox, and (3) new sources of external force (e.g., a puffer, and static and dynamic blowers). *PP* includes *stealth assessment*, in which real-time assessment is embedded in the learning environment, invisible to the learner (Shute, 2011). The idea behind stealth assessment is to assess student skills, and how students are dynamically advancing in those skills, without incurring the usual anxiety and time commitment that comes from standard testing.

To engage and enhance student performance, Shute and colleagues (2019) recently added AI-based learning supports and, in particular, an incentive and recommendation system called *My Backpack*. When clicked, *My Backpack* provides information about the player's progress and physics understanding, as well as a space to customize game play (e.g., the type of ball they use). More specifically, the "Physics" tab shows the estimated competency level for each targeted physics concept -- essentially an open learner model, based on real-time stealth assessment. Because stealth assessment is used to personalize content and adapt feedback (Shute et al., 2021), the *Physics Playground* is also in the "AI-based Adaptation" category discussed previously. In addition, Shute and colleagues (2015) have performed post-gameplay analyses of student log data to investigate student affect and predictors of learning outcomes while game playing; thus, *PP* is also in the "Use of Learning-Analytics …" category of AIED digital learning games.

Shute and colleagues have conducted many studies with *PP* since its initial creation as "Newton's Playground" (Shute & Ventura, 2013). For example, Shute et al. (2015) reports a study with 137 8th and 9th grade students who played *PP* for 2 1/2 hours. The students had a significant increase in scores from pretest to posttest, thus indicating that *PP* does indeed lead to physics learning. In post-gameplay analyses, they also found evidence that (1) both the pretest and the in-game measure of student performance significantly predicted learning outcomes and (2) a detector of frustration, a detector of engaged concentration, and the pretest predicted the in-game measure of performance. Finally, they found evidence for pathways from engaged concentration and frustration to learning, via the in-game progress measure. More recently, Shute and colleagues (2021) conducted a study to explore the benefits of the *PP* learning supports, such as *My Backpack,* a glossary, and animations of physics concepts. The study included 263 ninth-to-eleventh graders who played *PP* for 4 hours. In this study, students were randomly assigned to one of four conditions – an adaptive version of *PP*, in which a Bayesian algorithm was used to present game levels to students according to their physics competencies; a linear version, in which students followed a predetermined sequence of game levels; a free-choice version, in which the students were presented the linear sequence of levels, but could skip levels, and a no-treatment control. Surprisingly, there were no significant differences in learning between the four conditions. However, Shute and colleagues found that stealth assessment was an accurate estimate of students' physics understanding (i.e., stealth assessment was highly correlated with external physics scores), and physics animations were the most effective of eight supports in predicting both learning and in-game performance. Like the *Decimal Point* study discussed previously, this study is an excellent example of how digital learning games are becoming excellent platforms for pursuing various research questions, such as whether self-regulated learning can be productive in a game context and what the effects of various types of support are in game-based learning.



(a)



(b)

Figures 4(a) and 4(b): The figure at the top (4(a)) is an example of a student solving a *Physics Playground* problem by drawing a lever with a weight on one side of the lever to hit the balloon. At the bottom (4(b)) is the My Backpack view, which provides estimates of student progress, physics skill and concept understanding. (Figures from Shute et al., 2021, reproduced by permission from the first author)

There are fewer learning games in the “AI-based Decision Support” category than in any of the other learning games categories, but the few games that include decision support, typically in the form of an open learner model, include *Gamified Lynnette, TALENT, Keep Attention,* and the *Tactical Language and Cultural Training System (TLCTS)*. *Gamified Lynnette* (Long & Aleven, 2014) is an example of a gamified intelligent tutoring system, based on the linear equation tutor *Lynnette* (Waalkens et al., 2013). To extend *Lynette* with game features – creating what we (but not the authors) call *Gamified Lynnette* – Long and Aleven added two key features: (1) the possibility to re-practice problems, under student control, and (2) rewards given to students based on their performance on individual problems. The two features are connected (at least loosely) as students are encouraged to re-practice by earning rewards. The core AI technique of *Gamified Lynnette* is its use of BKT (Corbett & Anderson, 1995) to assess students’ skills and to support problem selection, either by the system (implementing mastery learning) or student (through an OLM dashboard, thus, its inclusion in the “AI-based Decision Support” category). *Gamified Lynnette*’s dashboard shows the student their rewards (i.e., badges) and allows students to select the BKT-assessed problem to tackle next. In their first experiment with *Gamified Lynnette,* Long and Aleven compared four versions of *Gamified Lynnette*, with and without re-practice enabled and with and without rewards, as well as to standard *Lynnette* (i.e., with problems all system selected and no dashboard) and *DragonBox* (http://www.dragonboxapp.com), a highly acclaimed and popular commercial digital learning game that covers the same mathematics content as *Gamified Lynnette*. A total of 267 7th and 8th grade students were randomly assigned to the six conditions, with 190 of those students finishing all activities and thus being subject to analyses. Long and Aleven did not find a significant difference between the different versions of *Gamified Lynnette* and the *Lynnette* control with respect to enjoyment or learning. However, *Gamified Lynnette* students who had the freedom to re-practice problems, but were not given rewards, performed significantly better on the post-tests than their counterparts who received rewards. This suggests that adding game features – in this case rewards – does not always enhance learning. Of particular note is that each of the *Gamified Lynnette* conditions led to more learning than the *DragonBox* condition, indicating that just a small dose of game features added to more traditional learning technology may be enough to lead to more learning versus a game. Long and Aleven (2018) later re-analyzed the 190 students of the Long and Aleven (2014) study, comparing the combined results of the five versions of *Lynnette* (i.e., the four gamified and the original, non-gamified versions of *Lynnette*) with *DragonBox.* They found that while students solved more problems and enjoyed playing *DragonBox* more, they learned more using *Lynnette* and *Gamified Lynnette*. This is an important additional finding to that of Long and Aleven (2014), as it shows that full-scale digital learning games, such as *DragonBox*, while often more engaging, may not always lead to better learning outcomes than more traditional and/or gamified technology. This finding further emphasizes the importance of conducting empirical studies of learning games (something that had not been done previously with *DragonBox*) before setting them loose with students in classrooms.

Another representative game in this category is *TALENT* (Maragos, 2013), a multiplayer adventure game that helps high school students learn introductory programming. To play and learn with *TALENT* students explore a game world, solving programming tasks as they proceed. They can chat with one another to discuss their game progress or exchange programming advice. *TALENT* has an open learner model based on information about the student’s in-game activities, such as navigation, tool usage, and learning progress. Further, it proposes the next mission to undertake, based on the student’s current level of game-assessed knowledge. The game also features a pedagogical agent whose role is to provide real-time hints (and thus *TALENT* is also in the “AI Character Interaction” category). A study of *TALENT* was conducted with 65 high school students, over a period of 8 weeks. An experimental condition in which students played *TALENT* was compared to a group of students that attended traditional lectures. While both groups improved from pretest to posttest, the experimental group had significantly higher posttest scores. However, given the stark comparison of traditional classroom instruction with the *TALENT* game, including collaboration and an OLM, it is not clear which feature, or combination of features, led to the better learning outcomes of the game-playing group.

Two other games in the "AI-based Decision Support" category are *Keep Attention* and *TLCTS. Keep Attention* (Hocine, 2019; Hocine et al., 2019) is a learning game for children with attention deficit disorders and trains attention skills (Hocine et al., 2019), as well as self-regulation skills (Hocine, 2019). The base game operates by first using basic tasks to assess students' attention skills, followed by personalizing their gameplay training – using the assessed skills as a guide – in a variety of game contexts (e.g., a zoo, card playing, outer space). For instance, in the zoo context players are prompted to rescue animals, while avoiding insects and monsters. The basic assessment is composed of selective attention tasks that prompt players, under time pressure, to select given targets in different contexts, while avoiding obstacles (distractors). In one version of the game, an open learner model (OLM) was developed to explore self-regulated learning by allowing users to reflect upon their actions during assessment and make decisions about their subsequent gameplay. The OLM uses a visual presentation of the player's performance, analyzed using AI techniques. To gamify the player's interaction with the OLM, *Keep Attention* prompts the player to predict their performance in the assessment exercises and levels based on an interactive questionnaire that provides feedback. Through the OLM, players can decide whether to accept or not the system's decisions on personalized gameplay. Unlike most of the other learning games reviewed in this chapter, no large-scale empirical studies have been conducted with *Keep Attention*. However, in a pilot study of the open learner model with 16 middle school students (Hocine, 2019) it was found that without the OLM, subjects found it difficult to assess their attention level, and consequently to choose the appropriate level of difficulty in the game context. Conversely, it appeared that what students learned about their attention from the OLM helped them during subsequent gameplay. While the *TLCTS* game is more representative of the "AI Character Interaction" category, given its extensive use of interactive NPCs to support language and culture learning, it also uses an AI recommender system to give learners advice on which lessons and game episodes to tackle next. *TLCTS* is described in more detail in the next section,

As with other forms of educational technology, open-learner models and recommender systems in digital learning games are designed to support and enhance students' self-regulated learning skills, ultimately leading, it is hoped, to better learning outcomes. OLMs seem especially appropriate for digital game contexts, given that computer-based games often allow players to make their own decisions about the next game level, challenge, or pathway to pursue during play. The difference in digital learning games is that player choices are not only about engagement and enjoyment but also about optimizing learning outcomes. So far, the results have been inconclusive regarding the benefits of OLMs in digital learning games. For instance, the learning-focused OLM in *Decimal Point* did not lead to clear learning benefits over the enjoyment-focused OLM; in fact, the enjoyment-focused OLM led to more efficient learning than for the students in the control group. Also, the *Physics Playground* OLM (i.e., *My Backpack*) may have helped students but did not lead conclusively to better learning. There is at least a hint that the dashboard of *Gamified Lynnette* may have been responsible for better learning outcomes, given the better learning outcomes of the *Gamified Lynnette* with a dashboard but without rewards over *Gamified Lynnette* with rewards. Even so, further study, focused on testing the dashboard, would be necessary to confirm the benefits of a dashboard. In short, the jury is still out on the impact and benefits of AI-supported OLMs and recommender systems in the context of digital learning games. Testing OLMs and recommender systems with digital learning games is perhaps more challenging and complicated than with other learning technologies, given the trade-offs and tension between engagement and learning inherent in digital learning games.

3.3. AI Character Interaction

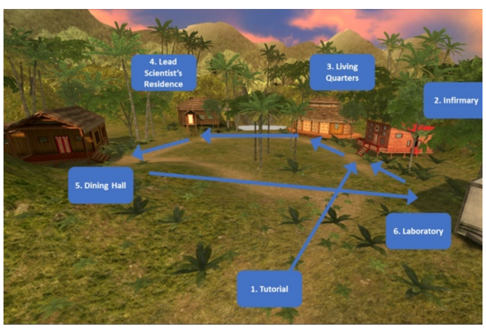
AI-based character interaction involves the use of Non-Player Characters, commonly called NPCs, engaging and interacting with students as they play a digital learning game. In the context of digital learning systems and games, these characters are also sometimes called “pedagogical agents” (Johnson & Lester, 2018): virtual characters that are part of learning scenarios and engage in rich interactions with students. Some games that have already been discussed in this chapter also fall into this category, including *MathSpring* (Arroyo et al., 2013; 2014), which uses NPCs to provide socio-emotional support to students, and the *Tactical Language and Culture Training System* (Johnson, 2010), which was briefly discussed in the previous section, and which uses NPCs to help adult trainees understand foreign language and culture. “AI Character Interaction” games often employ AI in the form of Natural Language Processing (NLP), as well as machine learning to develop the language capabilities and affect of such characters.

In this section we describe in detail three highly representative “AI Character Interaction” games, across the disparate domains of microbiology (*Crystal Island*), language and cultural instruction (the *Tactical Language and Culture Training System – TLCTS*), and programming (*TurtleTalk*). We describe the NPCs that have been implemented in those games and how AI has been used to bring the NPCs to life as interactive pedagogical agents with human learners. In some cases, in particular with *Crystal Island*, we also discuss other ways that AI has been employed in the game. We also summarize the results attained from experimenting with these three games, ranging from extensive and substantial, in the case of *Crystal Island,* to very preliminary, in the case of *TurtleTalk*. We also briefly discuss other games that fall within the category of “AI Character Interaction” games and conclude by discussing what has been learned thus far from “AI Character Interaction” game research.

Crystal Island (Lester et al., 2013) is a single player, narrative learning game designed for the domain of microbiology, typically targeted at late middle school students (but even university students have been a target population). The game features a science mystery situated on a remote tropical island (Figure 5(a)). Within the narrative of *Crystal Island,* the student plays the role of a medical field agent attempting to discover the identity of an infectious disease plaguing the island inhabitants. To solve the mystery, the student collects data and symptoms, forms hypotheses, and tests those hypotheses by interviewing a number of NPCs on the island, including a camp nurse, the camp cook, and virus and bacterial scientists. The student’s learning is scaffolded by interaction with the NPCs (Figure 5(b)) and a diagnosis worksheet (Figure 5(c)). When the student makes an incorrect diagnosis, the camp nurse identifies the error and provides feedback. The student successfully completes the game when they correctly identify the illness and specify an appropriate treatment.

A variety of AI techniques have been employed and tested with *Crystal Island*, including narrative-centered tutorial planning (Lee et al., 2011; 2012), student knowledge modeling (Rowe & Lester, 2010), and student goal recognition and affect recognition models (Sabourin et al., 2011). Most of these AI directions have been aimed at making the camp nurse smarter and more adaptable to and supportive of students playing the game. For instance, the narrative-centered tutorial planning uses a dynamic decision network (Dean & Kanazawa, 1989), to update the beliefs of the NPCs, and to select actions that maximize expected tutorial utility. For student modeling and affect recognition, Lester and colleagues developed a dynamic Bayesian network that connects students goals (e.g., mastery or performance) with emotions (e.g., boredom, confusion) and actions (e.g., notes taken, tests run) (Rowe & Lester, 2010). The game’s data has also recently been analyzed with deep learning techniques to build student models from log data and reflection texts (Geden et al., 2021), to recognize student goals from game events and eye tracking (Min et al., 2017), and to enable data-driven and interactive narrative personalization (Wang et al., 2017). These data mining studies place *Crystal Island* also in the “Use of Learning-Analytics *...”* game category still to be discussed.

Crystal Island is one of the earliest developed and most prolific AIED learning games; it has been the subject of many empirical studies, exploring a variety of issues – like *Decimal Point* and *Physics Playground* it has proven to be an excellent research platform. Here, we cite and summarize just four studies. An early observational study conducted by Rowe et al. (2011), in which the goal was to investigate the synergistic impact of engagement and learning, 150 students played *Crystal Island.* Students achieved significant learning gains from pre to posttest, providing evidence that the game is effective in engaging learners and supporting learning. In an experiment comparing three narrative-centered tutorial planning models, 150 eighth-grade students used *Crystal Island*, completing a pre- and posttest (Lee et al., 2012). While it was found that all students learned, only students in the "full guidance" condition (compared to intermediate and minimal guidance conditions) achieved significant learning gains. A study that did not yield positive learning outcomes for *Crystal Island* was conducted by Adams and colleagues (2012) in which students using *Crystal Island* were compared to students viewing a slideshow with the same microbiology materials. The *Crystal Island* group performed much worse than the slideshow group on a subsequent test of infectious disease with an effect size of -0.31. Finally, in Sawyer et al. (2017), Lester's lab explored the concept of *agency*, that is, student (versus system) control within a game. In this study, 105 college-age students were randomly assigned to one of three agency conditions: high agency, which allowed students to navigate to island locations in any order; low agency, which restricted students to a prescribed order of navigation, and no agency, in which students simply watched a video of an expert playing *Crystal Island*. They found that students in the low agency condition attempted more incorrect submissions but also attained significantly higher learning gains. Their results suggest that limiting agency in their game (and possibly in other games, as well) can improve learning performance but at the potential cost of undesirable student behaviors, such as a propensity for guessing.



(a) (b)

(c)

Figures 5(a), 5(b), and 5(c): The locations on Crystal Island where the student can visit to solve the infectious disease mystery (5(a)). The student interacts with AI-infused NPCs, such as the camp nurse and patients (5(b)). A diagnosis worksheet scaffolds students learning; they can record their findings in the worksheet (5(c)). (Figures from Taub et al., 2020 and Lester et al., 2013, reproduced by permission from the authors)

The *Tactical Language and Culture Training System* (*TLCTS* - Johnson, 2010) is a virtual learning environment that helps learners acquire basic communication skills and cultural understanding of foreign countries. This is done through interactive and practical lessons. *TLCTS* modules have been deployed to teach Arabic, Chinese and French, among other languages and cultures. Each module includes three major components: Skill Builder, Arcade Game, and Mission Game. In the Skill Builder, learners practice vocabulary and phrases, and complete assessment quizzes that require mastery of spoken language. In the Arcade Game, learners give voice commands in the target language to control the behaviors of game characters. In the Mission Game, learners speak on behalf of their characters to interact with in-game NPCs. All *TLCTS* content is highly task-based and promotes hands-on practice, while leveraging game elements to help learners overcome the motivational barrier of language learning.

TLCTS incorporates AI dialog models to support robust spoken dialog with NPCs; these models can interpret the learner's verbal and non-verbal gestures and control the NPCs' responses accordingly. More specifically, the game uses automated speech recognition to provide immediate feedback, both on simple repeat-and-recall tasks and on more complex dialog exercises. The speech recognition is enabled through an underlying acoustic model, trained on a combination of native speech and learner speech to ensure an acceptable recognition rate. *TLCTS* also uses AI to maintain student models to track students' mastery of targeted language skills throughout game play; these models provide learners with recommendations on the lessons and game episodes to undertake next (thus, this game is also in the "AI-based Decision Support" category).

Three evaluations have been conducted on *TLCTS,* all of Iraqi Arabic language and culture. The first study involved 89 marines, each of whom underwent four hours of self-paced computer-based training each week over three months. The second study recruited eight (8) military personnel who spent eight hours of *TLCTS* training per day, over five days. The third study had 268 participants, who had 28 hours of training with *TLCTS*. All three studies yielded significant increases in knowledge of Iraqi Arabic language, as measured by an independently constructed posttest; in two of the studies, participants also reported significant increases in speaking and listening self-efficacy.

Figure 6: An active dialog in *TLCTS*. Trainees learn how to use language and act culturally appropriate when interviewing Iraqi civilians. © Alelo Inc. Reprinted with permission

The final game we'll discuss in detail in the "AI Character Interaction" category is *TurtleTalk* (Jung et al., 2019), a single player, web-based programming game that interacts with learners through voice recognition and speakers (See Figure 7). The children's' utterances are converted directly into code. The game focuses on the programming topics of sequencing and iteration, both of which are fundamental programming constructs. In the case of sequencing, child users learn that a program runs in a sequence of steps, where an action, or step, leads to the next ordered action. In the case of iteration, children learn that programs often contain loops of steps that are executed repeatedly. The child and *TurtleTalk* communicate with one another in a turn taking manner to control the conversation and help the child focus on their given tasks. *TurtleTalk* provides hints to help the child decide what to do, but it does not appear that, as of this writing, the authors have grounded the pedagogical model underlying the game in established learning science.

Although there have been other computer gamified environments for learning programming, such as *Scratch* (Resnick et al., 2009) and *Blockly* (Fraser et al., 2013), *TurtleTalk* brings a new twist: an AI-infused NPC – the turtle – that guides and helps children through its voice and voice understanding. The Turtle appears on the screen and moves onto blocks according to voice commands, which are converted into programming code by a neural network.

Thus far, however, only a small pilot study of *TurtleTalk* has been conducted. Eight participants, 6 to 9 years of age, were recruited for the study. Players were surveyed after playing the game to learn about their playing experience (there was no pretest or posttest). The general results (which are, of course, very preliminary and based on minimal data) are: (1) Children appear to learn programming easily, enjoyably, and confidently with *TurtleTalk;* (2) Children understand the key constructs of programming through *TurtleTalk;* and (3) Voice interaction allows children to become more immersed in the game.

Figure 7: A child listening to TurtleTalk's questions and then commanding the turtle NPC to act through voice command. (Figure from Jung et al., 2019, reproduced by permission from the authors)

Other digital learning games in the "AI Character Interaction" category include *ELIA, ECHOES*, *Squares Family, TALENT* (already discussed), and *Betty's Brain. ELIA* (*E*motions for *L*earning and *I*ntelligent *A*ssessment – Kaczmarek & Petroviča, 2018) is a single player game targeted at university students for learning about artificial intelligence. The game is patterned after the famous "Who wants to be a Millionaire" television game show in which students try to win a "million-dollar prize" by correctly answering a series of questions. *ELIA* integrates game mechanics, most importantly an on-screen NPC that supports the student, with an underlying intelligent tutor. The game uses facial expression recognition software, supported by a neural network, to help in identifying student emotions and deciding on a tutoring approach. *ELIA* analyzes and classifies students through a pre-game questionnaire, subsequently selecting an appropriate supporting NPC (friend, expert, coach, or evaluator) and teaching strategy (mastery or performance). During game play questions to the student are adapted with the aid of the *Emotion API,* a commercially available API from Affectiva[2], which is based on more than 6 million face videos collected from 87 countries. Each of the supporting NPCs have their own (different) rules for tutoring based on different perceived emotions and student progress. In a study with 240 university students, a game-based condition in which 87 students used *ELIA*, compared to a paper-based, business-as-usual condition with the same content, showed that the game led to lower posttest scores. However, the game also led to increased motivation and engagement for lower-performing students, prompting the authors to suggest that the game still could be beneficial for that class of students. *ECHOES* (Bernardini et al., 2014) is a digital learning game designed to train young children with Autism Spectrum Conditions (ASCs) to acquire social communication skills. An intelligent virtual character, Andy, guides and teaches children, by playing the role of both a peer and a tutor, through a multi-touch LCD display with eye tracking. All activities take place in a sensory garden populated by Andy with interactive magic objects that react in unusual ways when touched by either the children or the agent. Children practice two forms of activities: *goal-oriented*, where they follow steps with an identifiable end-goal, and *cooperative turn-taking*, where they take turns exploring the garden with Andy to promote social reciprocity. Andy is also responsible for fostering the children's initiative and exposing them to positive peer interactions. Andy employs a domain-independent architecture FAtiMA (Dias & Paiva, 2005) to display one of 22 emotions, based on its assessment of the current events. The agent's back-end has a pedagogical component that monitors its interaction with the child, and a "child model," which assesses in real time the child's cognitive and emotional state. *ECHOES* was tested in a small study of twenty-nine (29) children 4 to 14 years old from the U.K. with Autism Spectrum Conditions (ASC) and/or other disabilities. While no significant transfer of increased social responsiveness or initiations to real-world contexts were observed across all children, the experimental results showed that the number of initiations made by the children when first using *ECHOES* wass significantly less than the number of initiations made during the final session.

The last two games in the "AI Character Interaction" category, *Squares Family* and *Betty's Brain*, employ an instructional paradigm known as "teachable agents" (TA - Brophy et al., 1999; Chase et al., 2009; Leelawong & Biswas, 2008). Teachable agents technology draws on the social metaphor of a student teaching a computer agent, which in turn can help the student themselves learn. TA is based on the theories of learning-by-teaching (Bargh & Schul, 1980; Palthepu et al., 1991) and the protégé effect (Chase et al., 2009), in which students are coaxed to work harder to learn for their TAs than on their own. *Squares Family* (Pareto, 2009; 2014; Sjödén et al., 2017) is a two-player math card game (either two humans or a human and computer player) designed to teach elementary school students a conceptual understanding of mathematics and mathematical reasoning, in particular of positive and negative integer operations. The game prompts discussion between pairs, using a graphical metaphor of colored squares that denote 100s, 10s, and 1s and collected squares that go into virtual boxes. A key feature of the game is that students train an on-screen teachable agent (TA) as they play the game (e.g., Pareto, 2014; Sjödén et al., 2017). The AI component of *Squares Family* lies in how the TA "learns" as the game is played. Just as a human learner, the teachable agent asks questions of the student during game play to complement and build its knowledge. The TA will only ask questions that are, according to its understanding, within the student's *zone of proximal development* (Vygotsky, 1978). Pareto (2014) reports a three-month study in which 314 students played *Squares Family* and taught their teachable agents, while 129 took a standard math class without the game (i.e., the control). Results showed that there was a significant learning advantage in playing *Squares Family* compared to the control, also suggesting that young students can act as successful tutors. In another study, Sjödén and colleagues (2017) analyzed the anonymous log data of 163 fourth-graders playing *Squares Family* competitively

---

[2] https://www.affectiva.com/

against the computer from nine classes in a school in Sweden (age 10–11, total of 3,983 games). Results showed that students who tutored a TA had higher performance than students who played without a TA.

We note that the final game in the "AI Character Interaction" category, *Betty's Brain* (Biswas et al., 2016), unlike *Squares Family*, is more purely a TA learning technology than a game. However, because of the rich interaction between students and agents, as well as the goals of *Betty's Brain*, which arguably fit our earlier definition of digital learning games, for the purposes of this chapter, we include it as a learning game[3]. *Betty's Brain* trains students in modeling chains of cause-and-effect relationships using a concept map that students develop. Students actively engage with an agent named Betty in three activities: teaching Betty to answer hypothetical questions about their concept maps, Betty reasons in real-time visually in the concept map (which helps to remediate student knowledge), and Betty is quizzed by a mentor NPC at the end of a session. By interacting with and guiding Betty, the idea is for students to update their concept maps and, along the way, learn more about cause and effect. Betty's use of AI includes a breadth-first search of the concept map to deduce relationships in the student's concept map (Leelawong & Biswas, 2008). Betty also displays AI-controlled affect – e.g., happiness, disappointment – as she solves problems. Other uses of AI with Betty's Brain include the application of learning analytics to improve the game (Segedy et al., 2014; Kinnebrew et al., 2014) – Hence, *Betty's Brain's* inclusion in the fourth and final category of AI in digital learning games, "Use of Learning-Analytics (LA) *..."*). *Betty's Brain* has been subject to a variety of studies including one by Segedy et al. (2015) in which 98 6th grade students used *Betty's Brain* to learn about two topics: climate change and human thermo-regulation. Results demonstrated learning gains by students in multiple-choice and short-answer item questions, but not on causal reasoning, causal link extraction, or quiz evaluation items.

It can be concluded that AI Character Interaction research has had some clear successes (Johnson, 2010; Pareto, 2014; Sawyer et al., 2017) in making games more realistic, engaging, and motivating, as well as leading to better learning outcomes. However, AI-supported characters in a game context have not always led to better learning outcomes. For instance, as earlier mentioned, in one *Crystal Island* study students playing the game did not learn more than from a slideshow of the same material (Adams et al., 2012). As another example, *ELIA* motivated students more than a business-as-usual condition with the same content but led to worse learning outcomes (Kaczmarek & Petroviča, 2018). The challenge in creating interactive, engaging, and helpful NPCs that lead to learning is clearly very high – perhaps higher than any other challenge in AIED digital learning games – thus leading to mixed results thus far. We'll return to this issue, and how it might be addressed, in the "Future Directions" section of this chapter. Another interesting and unique aspect of research in this area is how it can support the investigation of affect during instruction, both the affect of NPCs and that of students interacting with the game. For instance, some of the work with *Crystal Island* has been aimed, at least in part, at assessing student affect while game playing (Sabourin et al., 2011). Arroyo and colleagues (2013; 2014) have shown that an NPC can promote positive affect in students. In short, AI Character Interaction research is likely to continue to be a very interesting and fruitful direction for AIED digital learning game research.

3.4. Use of Learning Analytics (LA) and/or Educational Data Mining (EDM) for Game Analysis and Improvement

Learning analytics and educational data mining have led to important insights into learner behavior and affect while playing games (e.g., Alonso-Fernandez et al., 2019; Baker et al., 2007; Nguyen et al., 2020; 2019; O'Rourke, Haimovitz et al., 2014; Wang et al., 2019). For instance, in the learning game *Decimal Point*, EDM has helped in identifying students' learning difficulties (Nguyen et al., 2020), as well

[3] We contacted Gautam Biswas, the leader of the group that developed *Betty's Brain*, and he also noted anecdotally that during talks he has given, audiences have suggested that *Betty's Brain* could be viewed as a game.

as the relationship between in-game learning and posttest / delayed posttest performance (Nguyen et al., 2019). An example finding is that while problems involving a number line tend to be the most difficult to master, performance on these problems is predictive of performance on the delayed posttest. Engagement and affect have also been a focus of AIED analyses in the use of digital learning games, given the hypothesis that these constructs are mediators to learning with games. As students interact with games, they typically display a variety of affective responses, including delight, engaged concentration, happiness, boredom, confusion, and frustration (Graesser et al., 2014; D'Mello, 2013), and learners' affective state has often been shown to be correlated with their learning processes (e.g., D'Mello, 2013; Shute et al., 2015).

Learning analytics and educational data mining have also helped in designing, redesigning, and extending digital learning games. Often, this involves using machine learning as a means of analyzing and improving games (Shute et al., 2015; Harpstead & Aleven, 2015). Many of the techniques used for player modeling, analysis, and game improvement in AIED digital learning games are derived from research on intelligent tutoring systems (Mousavinasab et al., 2021). For example, learning curve analysis, which is based on the power law of practice (Card et al., 1983) and is often used in intelligent tutoring system research (Martin et al, 2005), has been used to model the acquisition of fluent skills in the game *Zombie Division*. This analysis provided insights into student skills that suggests (re)design of the game and its content (Baker et al., 2007). Likewise, Harpstead and Aleven (2015) used learning curves to help in analyzing student strategies and guiding redesign of the *Beanstalk* game.

In what follows, we describe in detail two games – *Refraction* and *Zombie Division* – for which post-gameplay learning analytics and/or data mining helps us both better understand student behavior and learning in game play and suggest ideas on how to revise, improve, and/or extend the games. We also discuss the specific way learning analytics and data mining was applied to data collected from those games. We then summarize and briefly discuss other games in the "Use of Learning Analytics (LA) …" category. Finally, we discuss what learning analytics and data mining has revealed to us more generally about learning from digital games. Note that some digital learning games for which learning analytics and/or EDM have been applied – namely, *Betty's Brain*, *Crystal Island*, and *Decimal Point* – are not further discussed here; information about how learning analytics or data mining was used in analyzing these games is provided in earlier sections.

The first game we describe in detail is *Refraction* (O'Rourke et al., 2015; 2016; O'Rourke, Ballweber et al., 2014; O'Rourke, Haimovitz et al., 2014), a single player, puzzle game designed to teach fraction concepts to elementary school students. To play, a child interacts with a grid that contains laser sources, target spaceships, and asteroids (See Figure 8). The goal of the game is to split a laser shooter into correct fractional amounts to shoot target spaceships. The student must, at the same time, avoid moving asteroids. To win, a player must accurately shoot all the target spaceships at the same time (i.e., correctly satisfy the fractions). Since its release in 2012, *Refraction* has been played hundreds of thousands of times on the educational website Brainpop.com. O'Rourke and colleagues modified the game to experiment with the concept of a *growth mindset*, the theory that intelligence is malleable and can be improved with effort (Dweck, 2006; Heyman & Dweck, 1998). A "fixed mindset," on the other hand, is a theory stating that intelligence is set and unchangeable.

O'Rourke and colleagues extended *Refraction* to reward players with "brain points" when they exhibited growth mindset behavior, such as effort, use of strategy, and incremental progress (O'Rourke, Haimovitz et al. 2014). Heuristics were used in real time play to detect and reward positive behavioral patterns. The heuristics were developed by first observing various patterns in student play. O'Rourke and colleagues then developed heuristic rules to identify those patterns during actual gameplay. For example, one rule identifies when a player solves a math sub-problem (incremental progress). Developing hand-authored rules -- versus machine-learned or statistical rules -- was key, because the rules needed to be semantically meaningful, providing feedback to students. The work of O'Rourke, Bellweber et al. (2014) also developed different "hints" versions of *Refraction*, in the intelligent tutoring tradition (VanLehn, 2006;

2011). (Surprisingly, in an experiment involving over 50,000 students, they found that students who used a no-hint version of the game learned more than students in four hint conditions - hint content (concrete versus abstract) and hint presentation (by level versus reward).)

The critical way that Refraction uses AI, however, is in its post-game quantitative analyses of student behavior. Use of *Refraction* was evaluated through an iterative series of studies. First, the “brain points” version of *Refraction* was compared to a control version that awarded “level points” for each completed level (performance). Using educational data mining techniques, this study showed that players in the “brain points” version persisted longer and exhibited more productive learning behaviors than those in the control (O’Rourke, Haimovitz et al. 2014). A second study, again comparing the “brain points” and “level points” versions of *Refraction*, showed, again applying data mining, that students who were younger, male, and of higher income were more engaged in the game (O’Rourke et al., 2015). In a third study, five different versions of “brain points” *Refraction* were compared, each of which removed features to isolate the impact of a particular aspect of the intervention (e.g. the brain points, growth mindset animation, a summary screen). The results of this study showed that brain points are not effective when awarded randomly, demonstrating that brain points are successful precisely because they are given to students in response to productive learning behaviors (O’Rourke et al., 2016). In short, this series of studies shows how quantitative and data mining analyses can be used to better understand how game features impact an educational game’s effectiveness – and, in turn, point to how to improve the game.

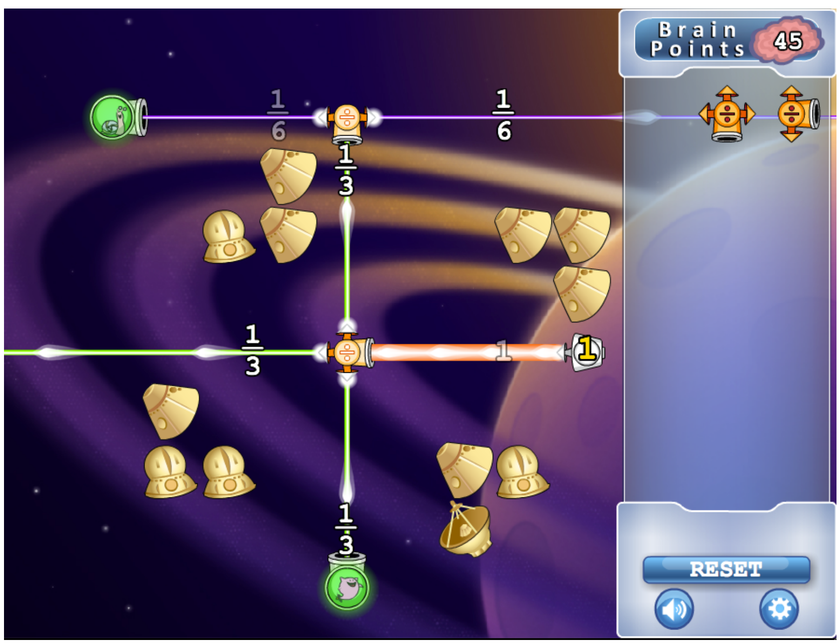


Figure 8: A level of *Refraction*. The goal of the game is to use the elements on the right to split lasers into fractional pieces and redirect them to satisfy the target spaceships. All spaceships must be satisfied at the same time to win. (Figure from O’Rourke et al., 2014, reproduced by permission from the first author)

Zombie Division (Habgood & Ainsworth, 2011) is a single player, 3D adventure game developed to help children 7 to 11 years old learn division and number patterns. The player in the game uses different weapons to kill attacking zombies by mathematically dividing into whole parts the numbers displayed on the zombies (See Figure 9). The game has different levels, as in many games, with approximately 20 attacking zombies per level. The game also includes game mechanics in which players explore a dungeon and collect keys; however these are not integrated with the mathematical content of the game. Players can get two types of in-game help: a magical book of times tables (the multiplication grid) and an NPC that provides helpful feedback.

Habgood and Ainsworth (2011) report two studies that explore whether intrinsic or extrinsic integration of academic content into a digital learning game leads to better learning and which version of the game students prefer. In the *intrinsic* version of the game (Figure 9), the player attacks zombies through the game mechanic of whole number division, as described above. In an *extrinsic* version of the game, the game is played *without* the numbered weapons and zombies. Instead, the same division problems are presented to students to solve *between* levels (i.e., completely outside of the game). A control condition was also included in which students played *Zombie Division* without the numbered weapons and numbered zombies or division problems presented between levels. In study 1, a total of 58 kids, 7 to 11 years old, participated, 20 in the extrinsic, 20 in the intrinsic, and 18 in the control condition. The intrinsic condition led to significantly better learning outcomes on both a posttest and delayed posttest than both the extrinsic and control conditions. In study 2, 16 children had the free choice to play either the intrinsic or extrinsic version of *Zombie Division* and for as long as they desired.This study demonstrated an overwhelming preference for the intrinsic version of the game; the kids in the intrinsic version spent on average more than 7 times longer playing than the extrinsic version. Essentially, the Habgood and Ainsworth results demonstrate that integrating math content directly into a game leads to both better enjoyment and learning outcomes than playing the same game, yet doing math separately.

While the *Zombie Division* game itself has no AI functionality, a learning curve analysis by Baker et al. (2007) was used to model skill fluency in the game and to suggest useful modifications. In particular, this analysis found, among other things, that students quickly and successfully gained fluency in divisibility by 2, 4 and 5, but not by 3, implying that future versions of the game should include extra support for the division-by-3 skill. In short, this is a stereotypical example of using EDM as a means for both assessing student behavior *and* subsequently suggesting changes to the game.

Figure 9: This is a screenshot of the intrinsic version of the *Zombie Division* game. The player's avatar is in the foreground and chooses weapons from the top of the screen to attack zombies as they appear. The goal is to choose weapons that are equal divisors of the zombies, which have numbers on their chests. The "magical book of times tables" is shown in the upper right. The hearts indicate the health of the player, the skull on the bottom right shows how many zombies are yet to appear in this level. (Figure provided by M. P. Jacob Habgood, reproduced by permission)

In general, a wide variety of AI, data mining, and statistical techniques have been used to investigate and model students' engagement and affective states. For example, Shute et al. (2015) used structural equation modeling to discover that in-game performance playing the *Physics Playground* (discussed earlier) can be predicted by pretest data, frustration, and engaged concentration. In the game *Prime Climb* (also discussed earlier), Conati and Zhou (2002) experimented with using a Dynamic Decision Network to model students' emotional reaction during their interaction with the game, based on a cognitive theory of emotion. Data from the BROMP protocol -- in which observational data is used to build machine-learned models of affect (Ocumpaugh et al., 2015) -- has been used to develop affect detectors for several digital learning games or game-like instructional systems, including *TC3Sim* (Henderson et al., 2020a, 2020b), *Physics Playground* (Shute et al., 2015), and *Reasoning Mind* (Ocumpaugh et al., 2013). Eye-tracking and other sensor technologies have been used to capture students' physiological data (e.g., in *Prime Climb* - Conati et al., 2013), which are increasingly used to create deep learning models that can infer students' affective states (Henderson et al., 2020a, 2020b; Loderer et al., 2019; Wiggins et al., 2018a; 2018b).

Other learning games for which learning analytics and data mining have been used for analyzing game behavior and/or game improvement include *Battleship Numberline, Heroes of Math Island, ST Math, Zoombinis, Beanstalk*, and *Downtown: A Subway Adventure*. *Battleship Numberline* (Lomas et al., 2011; 2012; 2013), a single-player game that gives players practice in estimating where whole numbers, fractions and decimals fall on a number line, has been, for instance, used to examine student engagement. A large-scale study with the game, involving approximately 70,000 players, found that players were more engaged and played longer when the game was easier, contradicting the common assumption that maximum engagement occurs at a moderate difficulty level (Lomas et al., 2013). Conati and Gutica (2016) developed a variety of machine learned detectors of affect for players of the game *Heroes of Math Island*, a narrative math game for middle schools students targeted at learning about divisibility, prime numbers, and number decomposition. Their ultimate aim was to detect student affect while playing the game, using a camera, and then to adjust the affect of a supporting NPC (a monkey) to help students while learning. Based on a small study of 15 students, Conati and Gutica were able to identify the frequencies of a wide range of emotions, but also noted the steep challenges of labeling game data to identify emotions. In the area of using EDM and statistical analysis for helping to improve a game and its content, Peddycord-Liu and colleagues (2017; 2019) experimented with *ST Math*, a single player learning game that uses spatial puzzles to teach mathematical concepts. They uncovered predictive relationships between different objectives and math concepts. For instance, they used linear regression to identify the most predictive prior objectives. In a study of 1,565 3rd grade students, Peddycord-Liu et al were able to use their techniques to identify the skills that students needed more practice on, a redesign of lessons within the game, and clusters of difficult problems that should be separated to balance students' pacing through the game. Rowe and colleagues (2020; 2021) conducted EDM analyses of *Zoombinis* (Hancock & Osterweil, 1996), a single-player game designed to help students learn computational thinking (CT), to better understand student behaviors in playing the game. For instance, Rowe et al. (2021) investigated the frequency of systematic testing and trial and error, common CT constructs, by hand labelling the data of 194 students, building machine-learned detectors, and then validating the detectors with 54 additional students. Harpstead and Aleven (2015) applied learning analytics to learn more about and to modify *Beanstalk,* a single person learning game for 5-to-8-year-olds based on the folktale of Jack and the Beanstalk and designed to teach the physical properties of a balance beam. They used learning curves to analyze the in-game behavior of 177 students playing *Beanstalk*. Their analysis suggested a game redesign related to a previously unidentified shallow game strategy that "wins" but does not reflect an understanding of the underlying balance beam principle. Finally, Cano and colleagues (2016; 2018) used learning analytics to investigate *Downtown: A Subway Adventure*, a spy game where a single player must discover the location and time of their enemy's gathering by navigating the Madrid subway and collecting clues along the way. The goal of the game is to train students with intellectual disabilities to use the public subway system. In a study of 51 adults with varying intellectual disabilities, they found, for instance, that students' prior experience with transportation training did not have an impact on their game performance, but those who played video games regularly performed better than their counterparts. In summary, a wide variety of techniques have been used to analyze student gameplay and affect in the use of AIED digital learning games, including machine-learned detectors, learning curve analysis, BKT analysis, deep learning, eye-tracking, video, and linear regression. All of these tools in the AIED "toolbox" have uncovered important findings about how students interact with learning games, as well as how we could make the games better, both in game mechanics and content. Note also that this category of AIED digital learning games is the most active, with the largest number of games having been analyzed (15).

4. Summary of What Have We Learned From Research with AIED Digital Learning Games

This chapter has presented a variety of AIED digital learning games that employ AI in various ways and that have been applied to many different topic areas, from math to science, computer science, cultural understanding, language learning, and social communication. Many of the games that have been presented in this chapter use AI to adapt game play, content, and help in a manner similar to what is often seen in intelligent tutoring systems (e.g., *iStart-2*, *MathSpring*, *Policy World, Prime Climb*). Other learning games use AI to provide useful real-time information, in the form of learning analytics, to help students make their *own* instructional decisions (e.g., *Decimal Point*, *Physics Playground, TALENT*) and to explore self-regulated learning. Still others include AI-supported non-player characters (NPCs) to guide and motivate students in their learning and social well-being (e.g., *Crystal Island*, *MathSpring*, *TLCTS, Turtle Talk*). Finally, perhaps the most important and prolific area of AI applied to digital learning games has been that of using learning analytics, data mining, and statistics to analyze student behavior and affect to help in better understanding student use of games and in redesigning the games (e.g., *Battleship Numberline, Beanstalk, Crystal Island, Physics Playground, Prime Climb, Refraction*, *Zombie Division, Zoombinis*).

AIED digital learning games have also demonstrated a considerable amount of success in a variety of studies, uncovering various aspects of learning from games that were not previously known. For instance, *Decimal Point* has been shown to be more effective and engaging in helping kids reinforce their understanding of decimals than a comparable decimal tutoring system (McLaren et al., 2017). *Decimal Point* has also been shown to be more effective in helping females learn (McLaren, Farzan et al., 2017). *AutoThinking (*Hooshyar et al., 2021) and *Zoombinis* (Rowe et al., 2021), through well designed and reasonably sized studies, have shown that computational thinking can be learned through game play. *Crystal Island*, a narrative game that is perhaps the longest standing and most studied of any AIED learning game, has shown in a variety of studies that students can learn microbiology from an NPC and a narrative driven game (e.g., Lee et al., 2012; Sawyer et al., 2017). A variety of studies have shown that students can "learn by teaching" an NPC, through interaction with *Betty's Brain* (Biswas et al., 2016) and *Squares Family* (Pareto, 2014; Sjödén et al., 2017). A classroom study of *Zombie Division* demonstrated an important and previously unknown aspect of digital learning games: that integrating academic content and game play is more effective for learning than separating the two (Habgood & Ainsworth, 2011).

A critical realization of the AIED digital learning game community has been the need for AIED researchers, who are predominantly learning scientists and educational technologists, to collaborate with game designers to be successful. The development of games with AI is inherently a multi-disciplinary undertaking; it requires not only the usual experts in learning science, educational technology and human-computer interaction, but also those versed in the intricacies of game mechanics and game design. An indication of this movement was the convening of a CHI 2018 workshop, "CHI 2018

Workshop: Data-Driven Educational Game Design” (McLaren et al., 2018), in which game designers, learning scientists and educational data miners met to discuss the various disciplines and issues that need to be addressed in order to create AIED digital learning games. More such workshops and events, with participation across a variety of disciplines, will benefit the further development of AIED learning games.

5. Future Directions

While there has been much work done in the area AIED digital learning games, also with much success, as outlined in the prior section, there are also several areas of potential work with digital learning games that have been under-explored.

For instance, AI could play a more prominent role in providing more realistic and compelling non-player characters and/or pedagogical agents within digital learning games. While games such as *Crystal Island* (Lester et al., 2013; Sawyer et al., 2017), *MathSpring* (Arroyo et al., 2014; 2013), and *TLCTS* (Johnson, 2010) have shown that even relatively primitive AI-infused non-player characters can make a significant difference to learning, there have been more recent developments in man-machine communication that could lead to an even greater impact. For instance, the emergence of relatively sophisticated virtual assistants such as Alexa (Lopatovska et al., 2019) has demonstrated the possibilities of more natural and effective communication between humans and technology. AI, and in particular natural language processing (NLP), could have a much bigger and direct impact on students’ learning experience when incorporated into pedagogical agents that accompany the student, either as learning companions or competitors (see, for instance, the chapter by Rus et al. in this handbook). There is ample opportunity to apply state-of-the-art methods in NLP (e.g., Crossley et al., 2017; Howard & Ruder, 2018; Ruseti et al., 2018; Young et al., 2018) to this area. Of course, a challenge with many of the natural language machine learning techniques is their need for a large amount of in-domain training data, often thousands or more examples. This requirement is difficult for many AIED digital learning games, which have typically been tested in traditional randomized controlled experiments in labs or classrooms with, at most, a few hundred students (with a couple notable exceptions: *Battleship Numberline* (Lomas et al., 2013) and *Refraction* (O’Rourke, Bellweber, et al., 2014), both of which managed to reach more than 50,000 players through Internet presence). Yet, with the pandemic of 2020-2022 and the sudden and significant move to online learning more AIED games may be positioned to reach much larger, out-of-school audiences. Working with organizations like Brainpop.com is perhaps a wise move for more AIED game researchers.

Equity and inclusiveness have rarely been considered in digital learning games (Buffum et al., 2016). Most prior work with AIED learning games – and in fact in AI instructional systems more generally – has treated student populations as though all individual learners are the same, or very similar. However, it is likely that different sub-populations learn and play in different ways. For instance, Ogan and colleagues (2015) found that the help seeking behavior of students using intelligent tutors can vary considerably across different cultures; they specifically found differences between Costa Rican students and students from the United States and the Philippines. In digital learning game research more specifically, McLaren and colleagues have consistently found that female students learn better than males with *Decimal Point* (Hou et al., 2020; McLaren, Farzan et al., 2017). Likewise, Arroyo and colleagues found that *MathSpring* improved mathematics learning, engagement and other affective outcomes for both female and low achieving students (Arroyo et al., 2013; 2014). As another example, the collaborative game *Engage* was explicitly designed to support females and students with less prior game experience in learning computing science (Buffum et al., 2016). A couple of the games reviewed in this chapter have reached to under-served populations, e.g., people with disabilities (*ECHOES* – Bernardini et al., 2014; *Downtown: A Subway Adventure* – Cano et al., 2016; 2018). Identifying and reacting to these learner differences is an important step towards more equity, better personalization, and, presumably, the increased effectiveness of games. AI could certainly play a role in better adapting games for these different and diverse populations. On the other hand, just as with AI research more generally, designers of AIED digital learning games must be extremely careful to avoid AI bias (Manyika et al., 2019; Prates et al., 2019). On this issue, the AIED community should engage with the larger AI community, following breakthroughs that might occur in machine learning algorithms and approaches, to help in eradicating AI bias in AIED digital learning games (and other AIED systems).

Another way that digital learning games could better employ AI is in the use of more sophisticated student models and adaptation approaches, for instance, to track enjoyment in addition to domain understanding and skills. While various adaptive algorithms have been used to modify gameplay to support increasing difficulty and help students reach mastery (Arroyo et al., 2013; 2014; Conati et al., 2013; Hooshyar et al., 2021), these algorithms have only been used in a relatively small number of learning games and have almost exclusively focused on learning objectives and domain skills. However, such models could also be used to modify gameplay to maximize fun and/or engagement for students, as has been implemented using dynamic difficulty adjustment in digital game research (i.e., video game research that is not focused on instruction – See e.g., Ang & Mitchell, 2019; Baldwin et al., 2016; Frommel et al., 2018; Zohaib, 2018). There have been some steps in this direction, most notably, the research with *Decimal Point* in which students self-reported their enjoyment in playing specific mini-games within *Decimal Point*, a student model of that enjoyment was displayed to them on a dashboard and was used by students to select further mini-games to play (Hou et al., 2020; 2021). While this research is a step in the direction of an "enjoyment" or "engagement" student model, there are still a lot of interesting and related research questions that may involve the support of AI. For instance, how do students react to different representations of their enjoyment? Is enjoyment more about encouraging what has already been enjoyed or seeking new and unknown opportunities?

An exciting new direction that a few AIED researchers have started to pursue, but is still in an embryonic stage, is supporting tangible game-based learning. These types of games involve a synergy of real-world physical play with computer-based instruction. For instance, Arroyo and colleagues (2017) have developed a game environment for learning mathematics that involves the use of "wearable tutors" in the form of smart phones and smart watches. The wearable devices act as mobile tutors, while students are given real world tasks – e.g., a scavenger hunt where students search for physical objects – and provide input to the devices on what they have done, while the tutors provide hints to help them in the physical environment. Students manipulate, measure, estimate, and find mathematical objects that satisfy certain constraints to help in playing the games. Another example is Yannier et al. (2016; 2020; 2021) who added the manipulation of physical objects, such as blocks and towers, to an AI-based mixed reality game to help 4-to-8-year-old children learn basic physics (see Figure 10).

Figure 10: Students playing with the *Intelligent Science Station (norilla.org).* (Figure provided by Nesra Yannier and Trinity Area School District, reproduced by permission)

A specialized AI computer vision algorithm is used to track what children do in the physical environment and then to provide personalized feedback in the form of an on-screen character (Yannier et al., 2020; 2022). They were able to show that adding the physical objects to the game – in essence, bridging the physical and virtual worlds in a mixed-reality setting – can lead to more learning compared to an equivalent solely screen-based game . There are a number of exciting questions that AIED could tackle in this space, such as: How can actions taken in the physical world be input to adaptive AI algorithms that provide pedagogical advice on what to do in the physical world? Can the physical objects themselves – for instance, an item being searched for in a scavenger hunt – be imbued with AI to add to the "smarts" and engagement of the games?

It should be noted that most of the games reviewed in this chapter are single player games, with just a few exceptions (*TALENT* – Maragos, 2013; *Prime Climb* – Conati et al., 2013; *Squares Family* – Pareto, 2009; 2014; *Physics Playground* – Sun et al., 2020; 2022). Thus, there may be learning opportunities that are being missed by not developing more collaborative AIED learning games. Working collaboratively could help students help one another by bringing their strengths, weaknesses, knowledge, and misconceptions to the game, thus complementing one another so that together group members can solve the problem at hand and learn (Housh et al., 2021; see also the chapter Martinez-Maldonado et al. in this volume). As previously mentioned, Buffum and colleagues (2016) explicitly designed for and found that their collaborative game, *Engage*, better supported female and less experienced students in collaborative play versus single player use of the game. Feedback from students in their study revealed potential benefits to collaboration, such as how a partner in gameplay can be a "sounding board" for what to try next or when one player is stuck in gameplay they can rely on their partner to take over. Perhaps AI could be supportive in such circumstances by having an AI agent that explicitly prompts the collaborating students to help one another in these, and other ways. As with *Squares Family* (Pareto, 2009) and AI programs in general, such as AI chess playing programs (Warwick, 2017), having students alternately play with or against an AI or a human might provide instructional opportunities not otherwise possible. For instance, the concept of learning by teaching (Biswas et al., 2005; Brophy et al., 1999; Palthepu et al., 1991) could be exercised in such collaborative gameplay circumstances, prompting students to teach *both* human and AI confederates in a game context.

Finally, the ultimate goal of research on AIED digital learning games should be to transfer more of the evidence-based learning games to the real-world. In this way, we not only *study* AI in learning games at schools and with students, but also provide those schools and students with substantial, on-going use of games that have been proven in rigorous empirical tests. Leading the way – and potentially providing a blueprint for others to follow – are games such as *Zoombinis* and *TLCTS*. TERC has sold almost 40,000 copies of the non-web version of *Zoombinis* and last year sold approximately 13,000 new licenses to the Internet version of the game[4]. Lewis Johnson has spun off a company, Alelo, that sells an updated version of TLTS to the U.S. military[5]. Hundreds of thousands of trainees have used Alelo's cultural games and tools to get cultural training. The military has required this training for people deployed in over 80 countries. While there are these glimmers of success in the real-world, there need to be more efforts made to transfer thoroughly tested learning games to real and wide-spread use in education and training contexts. To learn more about real-world transfer and commercialization of AI-based learning technology see chapters by Ritter and Koedinger, Luckin and Cukurova in this volume.

In conclusion, while there has already been a lot of impressive research done with AIED digital learning games, the future is bright with exciting new directions before us. By creating multi-disciplinary teams of learning scientists, educational technologists, game designers, and AI experts, we foresee a future in which AIED digital learning games will become an ever increasing and important component of education.

Acknowledgements

This work has been influenced by conversations and collaborations with many of our colleagues including Ryan Baker, Rosta Farzan, Jodi Forlizzi, Jessica Hammer, Richard Mayer, Michael Mogessie, Eleanor O'Rourke, and Liz Richey.

References

Adams, D.M., Mayer, R.E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*(1), 235-249. https://psycnet.apa.org/doi/10.1037/a0025595.

Aleven, V., Dow, S., Christel, M., et al. (2013). Supporting social-emotional development in collaborative inquiry games for K-3 science learning. In: *Proceedings of Games+Learning+Society Conference (GLS 9.0)*, ETC Press, 53–60.

Alonso-Fernandez, C., Martínez-Ortiz, I., Caballero, R., Freire, M., & Fernández-Manjón, B. (2019). Predicting students' knowledge after playing a serious game based on learning analytics data: A case study. *Journal of Computer Assisted Learning 2019*;1–9. doi: 10.1111/jcal.12405

[4] From personal correspondence with Elizabeth Rowe of TERC.
[5] From personal correspondence with Lewis Johnson of Alelo.

Ang, D. & Mitchell, A. (2019). Representation and frequency of player choice in player-oriented dynamic difficulty adjustment systems. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '19*, 589–600. https://doi.org/10.1145/3311350.3347165.

Arroyo, I., Burleson, W., Tai, M., Muldner, K., & Woolf, B. P. (2013). Gender differences in the use and benefit of advanced learning technologies for mathematics. *Journal of Educational Psychology*, 105(4), 957-969. https://psycnet.apa.org/doi/10.1037/a0032748.

Arroyo, I., Woolf, B. P., Burleson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education, 24*(4), 387-426. https://doi.org/10.1007/s40593-014-0023-y.

Arroyo, I., Micciollo, M., Casano, J., Ottmar, E., Hulse, T., & Rodrigo, M. (2017). Wearable Learning: Multiplayer Embodied Games for Math. *Proceedings of the ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY 2017)*. Amsterdam, Netherlands. https://doi.org/10.1145/3116595.3116637.

Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., & Graesser, A.C. (2010). Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. International Journal of Human-Computer Studies, 68 (4), 223-241. https://doi.org/10.1016/j.ijhcs.2009.12.003.

Baker R.S.J.D., Habgood M.P.J., Ainsworth S.E., Corbett A.T. (2007). Modeling the acquisition of fluent skill in educational action games. In: Conati C., McCoy K., Paliouras G. (eds) *User Modeling 2007 (UM 2007).* Lecture Notes in Computer Science, vol 4511. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-73078-1_5.

Baldwin, A., Johnson, D. & Wyeth, P. (2016). Crowd-Pleaser: Player perspectives of multiplayer dynamic difficulty adjustment in video games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, 326–337. https://doi.org/10.1145/2967934.2968100.

Banawan, M.P., Rodrigo, M.M.T. & Andres, J.M.L. (2017). Predicting student carefulness within an educational game for physics using Support Vector Machines. In: *Proceedings of the 25th International Conference on Computers in Education (ICCE 2017).*

Bargh, J., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology. 72(5)*, 593–604. https://psycnet.apa.org/doi/10.1037/0022-0663.72.5.593.

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin, 144*(1), 77–110. https://psycnet.apa.org/doi/10.1037/bul0000130

Benton, L., Mavrikis, M., Vasalou, M., Joye, N., Sumner, E., Herbert, E., Revesz, A., Symvonis, A. & Raftopoulou, C. (2021). Designing for "challenge" in a large-scale adaptive literacy game for primary school children. British Journal of Educational Technology. 52, 5: 1862-1880. Wiley. https://doi.org/10.1111/bjet.13146.

Bernardini, S., Porayska-Pomsta, K., & Smith, T. J. (2014). ECHOES: An intelligent serious game for fostering social communication in children with autism. *Information Sciences*, *264*, 41-60. https://doi.org/10.1016/j.ins.2013.10.027.

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. O. D. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning, 5*(02), 123-152. http://dx.doi.org/10.1142/S1793206810000839.

Biswas, G., Leelawong, K., Belynne, K., & Adebiyi, B. (2005). Case studies in learning by teaching behavioral differences in directed versus guided learning. In: *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 828-833).

Biswas, G., Segedy, J.R., & Bunchongchit, K. (2016). From design to implementation to practice – A learning by teaching system: Betty's Brain. *International Journal of Artificial Intelligence in Education, 26*(1), 350-364. https://doi.org/10.1007/s40593-015-0057-9.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems*, 6(2), 17.11-17.31. https://doi.org/10.1145/2946837.

Brophy, S., Biswas, G., Katzlberger, T., Bransford, J., & Schwartz, D. (1999). Teachable agents: Combining insights from learning theory and computer science. In S. P. Lajoie & M. Vivet (Eds.), *Proceedings of Artificial Intelligence in Education* (pp. 21–28). Amsterdam: IOS Press.

Buffum, P. S., Frankosky, M., Boyer, K. E., Wiebe, E. N., Mott, B. W., & Lester, J. C. (2016). Collaboration and gender equity in game-based learning for middle school computer science. *Computing in Science & Engineering, 18*(2), 18-28. https://doi.org/10.1109/MCSE.2016.37.

Bull, S. (2020). There are open learner models about!. *IEEE Transactions on Learning Technologies*, *13*(2), 425-448. https://doi.org/10.1109/TLT.2020.2978473.

Bull, S. & Kay, J. (2008). Metacognition and open learner models. In: The *3rd workshop on meta-cognition and self-regulated learning in educational technologies, at ITS2008*. pp. 7–20 (2008).

Cano, A.R., Fernández-Manjón, B., & García-Tejedor, Á.J. (2016). Downtown, a subway adventure: using learning analytics to improve the development of a learning game for people with intellectual disabilities. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)* (pp. 125-129). IEEE. https://doi.org/10.1109/ICALT.2016.46.

Cano, A.R., Fernández-Manjón, B., & García-Tejedor, Á.J. (2018). Using game learning analytics for validating the design of a learning game for adults with intellectual disabilities. *British Journal of Educational Technology*, *49*(4), 659-672. https://doi.org/10.1111/bjet.12632.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Castell, S. D., & Jenson, J. (2003). Serious play: Curriculum for a post-talk era. *Journal of the Canadian Association for Curriculum Studies, 1(1)*.

Chase, C., Chin, D.B., Oppezzo, M, & Schwartz, D.L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology, 18*(4), 334-352. doi: 10.1007/s10956-009-9180-4.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145-182. https://doi.org/10.1207/s15516709cog1302_1.

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477. https://doi.org/10.1207/s15516709cog1803_3.

Clark, D. B., Tanner-Smith, E., & Killingsworth, S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79-122. doi: 10.3102/0034654315582065.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale: Lawrence Erlbaum Associates.

Conati, C. & Gutica, M. (2016). Interaction with an edu-game: A detailed analysis of student emotions and judges' perceptions. *International Journal of Artificial Intelligence in Education, 26*, 975-1010. doi: 10.1007/s40593-015-0081-9

Conati, C., Jaques, N. & Muir, M. (2013). Understanding attention to adaptive hints in educational games: An eye-tracking study. *International Journal of Artificial Intelligence in Education, 23*: 131-161. doi: 10.1007/s40593-013-0002-8.

Conati, C. & Zhou, X. (2002). Modeling students' emotions from cognitive appraisal in educational games. In: *Proceedings of the International Conference on Intelligent Tutoring Systems (ITS 2002)*. 944-954. . http://dx.doi.org/10.1007/3-540-47987-2_94.

Corbett, A. T. & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*. *4* (4): 253–278. https://psycnet.apa.org/record/1996-29060-001.

Crocco, F., Offenholley, K., & Hernandez, C. (2016). A proof-of-concept study of game-based learning in higher education. *Simulation & Gaming, 47*(4), 403-22. https://doi.org/10.1177%2F1046878116632484.

Crossley, S., Liu, R., & McNamara, D. (2017). Predicting math performance using natural language processing tools. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 339-347). ACM. https://doi.org/10.1145/3027385.3027399.

Czikszentmihalyi, M. (1975). *Beyond boredom and anxiety,* San Francisco: Jossey Bass.

Czikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105*(4), 1082-1099. https://psycnet.apa.org/doi/10.1037/a0032674.

Dean, T., & Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence 5*(3): 142–150. https://doi.org/10.1111/j.1467-8640.1989.tb00324.x.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, Tampere, Finland (pp. 9-15). ACM. https://doi.org/10.1145/2181037.2181040.

Dias, J. & Paiva, A. (2005). Feeling and reasoning: A computational model for emotional characters. In: the *Proceedings of the Portuguese conference on Artificial Intelligence* (pp. 127-140). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11595014_13.

Dweck, C.S. (2006). *Mindset: The new psychology of success.* New York: Random House.

Easterday, M.W., Aleven, V., Scheines, R. & Carver, S.M. (2017). Using tutors to improve educational games: A cognitive game for policy argument, *Journal of the Learning Sciences, 26*:2, 226-276, doi: 10.1080/10508406.2016.126928.

Eggen, T. J., & Verschoor, A. J. (2006). Optimal testing with easy or difficult items in computerized adaptive testing. *Applied Psychological Measurement*, *30*(5), 379-393. https://psycnet.apa.org/doi/10.1177/0146621606288890.

Elo A. (1978).*The rating of chess players, past and present.* New York: Arco Publishers.

Embretson, S.E. & Reise, S.P. (2000). *Item Response Theory for psychologists.* Psychology Press. ISBN 9780805828191.

Fraser, N. et al. (2013). Blockly: A visual programming editor. URL: https://blockly-games.appspot.com.

Frommel, J., Fischbach, F., Rogers, K. & Weber, M. (2018). Emotion-based dynamic difficulty adjustment using parameterized difficulty and self-reports of emotion. In the *Proceedings of the Annual Symposium on Computer-Human Interaction in Play Extended Abstracts - CHI PLAY '18*, 163–171. https://doi.org/10.1145/3242671.3242682.

Geden, M., Emerson, A., Carpenter, D., Rowe, J., Azevedo, R., & Lester, J. (2021). Predictive student modeling in game-based learning environments with word embedding representations of reflection. *International Journal of Artificial Intelligence in Education*, *31*(1), 1-23. https://doi.org/10.1007/s40593-020-00220-4.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave/Macmillian.

Gee, J. P. (2007). Good video games and good learning Collected essays on video games, learning and literacy, 2nd Edition. Peter Lang International Academic Publishers. doi: https://doi.org/10.3726/978-1-4539-1162-4.

Gilbert, N. (2021). *Number of gamers worldwide 2021/2022: Demographics, statistics, and predictions*. FinancesOnline-Reviews for Business. https://financesonline.com/number-of-gamers-worldwide/

Graesser, A. C., D'Mello, S. K., & Strain, A. C. (2014). *Emotions in advanced learning technologies.* In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *Educational psychology handbook series. International handbook of emotions in education* (p. 473–493). Routledge/Taylor & Francis Group.

Gros, B. (2007). Digital games in education: The design of games-based learning environments. *Journal of Research on Technology in Education, 40*, 1-23.

Ha, E. Y., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). Goal recognition with Markov Logic Networks for player-adaptive games. In: *Proceedings of the Seventh International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 32–39. Menlo Park, CA: AAAI Press.

Habgood, M. P. J., & Ainsworth, S. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*(2), 169–206. doi:10.1080/10508406.2010.50802.

Hancock, C. & Osterweil, S. (1996). *Zoombinis* and the art of mathematical play. *Hands On! 19*, 1 (Spring 1996), pp. 17–19.

Harpstead, E. & Aleven, V. (2015). Using empirical learning curve analysis to inform design in an educational game. In: *Proceedings of CHI Play 2015*. doi: http://dx.doi.org/10.1145/2793107.2793128 .

Harpstead, E., Richey, J.E., Nguyen, H., & McLaren, B.M. (2019). Exploring the subtleties of agency and indirect control in digital learning games. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK'19).* (pp. 121-129). ACM. doi: https://doi.org/10.1145/3303772.3303797.

Henderson, N., Min, W., Rowe, J., & Lester, J. (2020a). Enhancing affect detection in game-based learning environments with multimodal conditional generative modeling. In: *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 134-143).

Henderson, N., Min, W., Rowe, J., & Lester, J. (2020b). Multimodal player affect modeling with auxiliary classifier generative adversarial networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 16, No. 1, pp. 224-230).

Heyman, G.D., & Dweck, C.S. (1998). Children's thinking about traits: Implications for judgments of the self and others. *Child Development 64*, 2, 391–403.

Hocine, N. (2019). Personalized serious games for self-regulated attention training. In: *Proceedings of the 27th Conference on User Modeling, Adaptation and Personalization (UMAP '19).* https://doi.org/10.1145/3314183.3323458.

Hocine, N., Ameur, M. & Ziani, W. (2019). Keep Attention: A personalized serious game for attention training. In: *Proceedings of the 3rd International Symposium on Gamification and Games for Learning (GamiLearn '19)*.

Homer, B.D., Hayward, E.O., Frye, J., & Plass, J.L. (2012). Gender and player characteristics in video game play of preadolescents. *Computers in Human Behavior, 25*(5), 1782-1789.

Honey, M. A. & Hilton, M. L. (2011). *Learning science through computer games and simulations*. The National Academies Press. (http://www.nap.edu/openbook.php?record_id=13078&page=R1).

Hooshyar, D., Malva, L., Yang, Y., Pedaste, M., Wang, M., & Lim, H. (2021). An adaptive educational computer game: Effects on students' knowledge and learning attitude in computational thinking. *Computers in Human Behavior, 114*.

Hou, X., Nguyen, H.A., Richey, J.E., & McLaren, B.M. (2020). Exploring how gender and enjoyment impact learning in a digital learning game. In: I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.) *Proceedings of the 21st International Conference on Artificial Intelligence in Education. AIED 2020.* Lecture Notes in Computer Science (LNCS, Vol. 12163). Springer, Cham. https://doi.org/10.1007/978-3-030-52237-7_21.

Hou, X., Nguyen, H. A., Richey, J. E., Harpstead, E., Hammer, J., & McLaren, B. M. (2021). Assessing the effects of open models of learning and enjoyment in a digital learning game. *International Journal of Artificial Intelligence in Education*, 1-31. https://doi.org/10.1007/s40593-021-00250-6.

Housh, K., Saleh, A., Phillips, T., Hmelo-Silver, C.E., Glazewski, K., Lee, S., Mott, B., & Lester, J. (2021). Designing for equitable participation in collaborative game-based learning environments. In: *Proceedings of the International Conference of the Learning Sciences (ICLS 2021)*.

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (Vol. 1, pp. 328-339).

Jackson, G. T., Guess, R. H., & McNamara, D. S. (2010). Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science*, *2*(1), 127-137. https://doi.org/10.1111/j.1756-8765.2009.01068.x.

Jackson, G. T. & McNamara, D.S. (2011). Motivational impacts of a game-based intelligent tutoring system. In: *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference*.

Jacovina, M. E., Jackson, G. T., Snow, E. L., & McNamara, D. S. (2016). Timing game-based practice in a reading comprehension strategy tutor. In: *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 59-68). Springer, Cham.

Johnson, W. L. (2010). Serious use of a serious game for language learning. *International Journal of Artificial Intelligence in Education 20*(2): 175–195.

Johnson, W.L. & Lester, J. (2018). Pedagogical agents: Back to the future. *AI Magazine 39.* Doi: https://doi.org/10.1609/aimag.v39i2.2793

Jung, H., Kim, H.J. So, S., Kim, J. & Oh, C. (2019) TurtleTalk: An educational programming game for children with voice user interface. In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts),* May 4–9, 2019, Glasgow, Scotland U.K. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3290607.3312773.

Juraschka, R. (2019, September 29). *How digital game-based learning improves student success*. Prodigy Math Blog. https://www.prodigygame.com/main-en/blog/digital-game-based-learning/.

Kaczmarek, S. & Petroviča, S. (2018). Promotion of learning motivation through individualization of learner-game interaction. In: *Proceedings of the 2018 IEEE Conference on Computational Intelligence and Games (CIG 2018)*. doi: https://ieeexplore.ieee.org/document/8490371.

Kaufman, G. & Flanagan, M. (2015). A psychologically "embedded" approach to designing games for prosocial causes. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *9*(3). doi: https://doi.org/10.5817/CP2015-3-5

Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development, 64*(2), 219-44. https://eric.ed.gov/?id=EJ1094502.

Ke, F., Shute, V., Clark, K.M., & Erlebacher, G. (2019). *Interdisciplinary design of game-based learning platforms: A phenomenological examination of the integrative design of game, learning, and assessment.* Springer. ISBN 978-3-030-04339-1.

Kinnebrew, J. S. & Biswas, G. (2012). Identifying Learning Behaviors by Contextualizing Differential Sequence Mining with Action Features and Performance Evolution. *International Educational Data Mining Society*. https://eric.ed.gov/?id=ED537192.

Kinnebrew, J.S., Loretz, K. M., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, *5*(1), 190-219. https://doi.org/10.5281/zenodo.3554617.

Kinnebrew, J. S., Segedy, J.R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and learning*, *9*(2), 187-215.

Kinnebrew, J.S., Segedy, J.R. & Biswas, G. (2017). Integrating model-driven and data-driven techniques for analyzing learning behaviors in open-ended learning environments. *IEEE Transactions on Learning Technologies, 10*(2), 140-153. https://eric.ed.gov/?id=EJ1146025.

Klinkenberg, S., Straatemeier, M., & van der Maas, H.L.J. (2011). Computer adaptive practice of Maths ability using a new item response model for on-the-fly ability and difficulty assessment. *Computers & Education 57*, 1813-1824. https://doi.org/10.1016/j.compedu.2011.02.003.

Landers, R., Armstrong, M., & Collmus, A. (2017). How to use game elements to enhance learning: Applications of the theory of gamified learning. *Serious games and edutainment applications*, 457-483. doi:10.1177/1046878114563660.

Landers, R. & Landers, A. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming*. 45(6): p. 769-785. https://doi.org/10.1177/1046878114563662.

Lee, S.Y., Mott, B.W., & Lester, J.C. (2011). Modeling narrative-centered tutorial decision making in guided discovery learning. In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Education*, 163–170. Berlin: Springer.

Lee, S. Y., Mott, B. W., & Lester, J. C. (2012). Real-time narrative-centered tutorial planning for story-based learning. In *Proceedings of the Eleventh International Conference on Intelligent Tutoring Systems*, 476–481. Berlin: Springer. https://dl.acm.org/doi/10.1007/978-3-642-30950-2_61.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education*, *18*(3), 181-208. https://psycnet.apa.org/record/2008-15202-002.

Lester, J. C., Ha, E. Y., Lee, S. Y., Mott, B. W., Rowe, J. P., & Sabourin, J. L. (2013). Serious games get smart: Intelligent game-based learning environments. *AI Magazine*, *34*(4), 31-45. https://doi.org/10.1609/aimag.v34i4.2488.

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://dl.acm.org/doi/abs/10.1016/j.ins.2013.09.005.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human‐Computer Interaction*, *7*(1), 57-78. https://doi.org/10.1080/10447319509526110.

Lindberg, R. S. & Laine, T. H. (2016). Detecting play and learning styles for adaptive educational games. In: *Proceedings of the International Conference on Computer Supported Education (CSEDU 2016).* (pp. 181-189).

Lindberg, R. S., Hasanov, A., & Laine, T. H. (2017). Improving Play and Learning Style Adaptation in a Programming Education Game. In *Proceedings of the International Conference on Computer Supported Education (CSEDU 2017).* (pp. 450-457).

Lindberg, R. S., & Laine, T. H. (2018). Formative evaluation of an adaptive game for engaging learners of programming concepts in K-12. *International Journal of Serious Games*, *5*(2), 3-24. https://doi.org/10.17083/ijsg.v5i2.220.

Lobel, A., Engels, R.C.M.E., Stone, L.L., Burk, W.J., & Granic, I. (2017). Video gaming and children's psychosocial well-being: A longitudinal study. *Journal of Youth and Adolescence, 46*(4), 884–897. doi: 10.1007/s10964-017-0646-z.

Loderer, K., Pekrun, R., & Lester, J. C. (2018). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002.

Loderer, K., Pekrun, R., & Plass, J. (2019). Emotional foundations of game-based learning. In J.L. Plass, R.E. Mayer, & B.D. Homer (Eds.), *Handbook of game-based learning*. (pp. 111-151). MIT Press. Cambridge, Massachusetts.

Lomas, D., Ching, D., Stampfer, E., Sandoval, M., & Koedinger, K. (2011). Battleship Numberline: A Digital Game for Improving Estimation Accuracy on Fraction Number Lines. *Society for Research on Educational Effectiveness.* https://eric.ed.gov/?id=ED528880.

Lomas, D., Patel, K., Forlizzi, J.L., & Koedinger, K.R. (2013). Optimizing challenge in an educational game using large-scale design experiments. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 89–98. ACM. https://doi.org/10.1145/2470654.2470668.

Lomas, D., Stamper, J., Muller, R., Patel, K., & Koedinger, K. R. (2012). The effects of adaptive sequencing algorithms on player engagement within an online game. In: *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 588-590). Springer, Berlin, Heidelberg. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.353.186&rep=rep1&type=pdf.

Long, Y. & Aleven, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. In: S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS 2014).* (pp. 378-387). New York: Springer. doi:10.1007/978-3-319-07221-0_47.

Long, Y. & Aleven, V. (2018). Educational game and intelligent tutoring system: A classroom study and comparative design analysis. *Proceedings of CHI 2018*. doi:10.1145/3057889

Lopatovska, I., Rink, K., Knight, I., Raines, I., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q., & Martinez, A. (2019). Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science, 51*(4), https://doi.org/10.1177/0961000618759414.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*, 333-369. doi: 10.1207/s15516709cog0504_2.

Manske, M., & Conati, C. (2005). Modelling Learning in an Educational Game. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, *pp. 411-418*.

Manyika, J., Silberg, J. & Presten, B. (2019). What do we do about the biases in AI? *Harvard Business Review.* https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai.

Maragos, K. (2013). Web-based adaptive educational games-Exploitation in computer science education. *The Committee of Research and Development*, University of Athens, 81, 2013.

Martin, S.M., Casey, J.R., & Kane, S. (2021). *Serious games in personalized learning: New models for design and performance*. Routledge. ISBN 9780367487508.

Martin, B., Koedinger, K.R., Mitrovic, A., & Mathan, S. (2005). On using learning curves to evaluate ITS. In: *Proceedings of the 12th International Conference on Artificial Intelligence in Education, AIED 2005.* 419-426. ISBN 978-1-58603-530-3.

Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach.* Cambridge, MA: MIT Press. ISBN: 9780262027571

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology. 70:* 531–49. https://doi.org/10.1146/annurev-psych-010418-102744.

Mayer, R. E. & Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research, 42*(3), 241–265. https://journals.sagepub.com/doi/10.2190/EC.42.3.a.

McLaren, B. M., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2017). A computer-based game that promotes mathematics learning more than a conventional approach. *International Journal of Game-Based Learning (IJGBL), 7*(1), 36-56. https://doi.org/10.4018/IJGBL.2017010103.

McLaren, B. M., Asbell-Clarke, J. & Hammer, J. (2018). “CHI 2018 Workshop: Data-Driven Educational Game Design” CHI 2018 Workshop at the CHI conference in Montreal, Quebec Canada.

McLaren, B.M., Farzan, R., Adams, D.M., Mayer, R.E., & Forlizzi, J. (2017). Uncovering gender and problem difficulty effects in learning with an educational game. In: E. André, R. Baker, X. Hu, M.M.T. Rodrigo, and B. du Boulay (Eds.). *Proceedings of the 18th International Conference on Artificial Intelligence in Education (AIED 2017).* LNAI 10331 (pp. 540-543). Springer: Berlin.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://eric.ed.gov/?id=ED602400

McNamara, D. S., Jackson, G. T., & Graesser, A. C. (2010). Intelligent tutoring and games (ITaG). In: Y. K. Baek (Ed.), Gaming for classroom-based learning: Digital role-playing as a motivator of study (pp. 44–65). Hershey, PA: IGI Global.

Min, W., Mott, B., Rowe, J., Taylor, R., Wiebe, E., Boyer, K., & Lester, J. (2017). Multimodal goal recognition in open-world digital games. In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 13, No. 1).

Mitrovic, A. (2012). Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User Modeling and User-Adapted Interaction*, 22(1-2), 39-72. https://doi.org/10.1007/s11257-011-9105-9.

Mousavinasab, E., Zarifsanajey, N., Niakan Kalhori, S.R., Rakhshan, M., Keikha, L, & Ghazi Saeedi, M (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments. https://doi.org/10.1080/10494820.2018.1558257.

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling learners’ cognitive and affective states to scaffold SRL in open ended learning environments. In: *Proceedings of the International Conference on User Modelling, Adaptation and Personalization (UMAP)*, Singapore, pp. 131-138. https://doi.org/10.1145/3209219.3209241

Nguyen, H., Harpstead, E., Wang, Y. & McLaren, B.M. (2018). Student agency and game-based learning: A study comparing low and high agency. In: C. Rosé, R. Martínez-Maldonado, H.U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren and B. du Boulay (Eds.). *Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED 2018).* LNAI 10947 (pp. 338-351). Springer: Berlin.

Nguyen, H.A, Hou, X., Stamper, J., & McLaren, B.M. (2020). Moving beyond test scores: Analyzing the effectiveness of a digital learning game through learning analytics. In: *Proceedings of the 13th International Conference on Educational Data Mining* (EDM 2020), pp. 487–495.

Nguyen, H., Wang, Y., Stamper, J., & McLaren, B.M. (2019). Using knowledge component modeling to increase domain understanding in a digital learning game. In: *Proceedings of the 12th International Conference on Educational Data Mining* (EDM 2019). (pp. 139-148). https://eric.ed.gov/?id=ED599221.

The NPD Group (2019). Retail Tracking Service, 2019 Entertainment Survey. https://igda-website.s3.us-east-2.amazonaws.com/wp-content/uploads/2019/10/16161928/NPD-2019-Evolution-of-Entertainment-Whitepaper.pdf

Ocumpaugh, J., Baker, R. S., Gaudino, S., Labrum, M. J., & Dezendorf, T. (2013). Field observations of engagement in Reasoning Mind. In: *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 624-627). Springer, Berlin, Heidelberg.

Ocumpaugh, J., Baker, R.S., Rodrigo, M.T.: Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual. (2015).

Ogan, A., Walker, E., Baker, R., Rodrigo, M.M.T., Soriano, J.C. & Castro, M.J. (2015). Towards understanding how to assess help-seeking behavior across cultures. *International Journal of Artificial Intelligence in Education, 25*, 229-248. https://doi.org/10.1007/s40593-014-0034-8.

O'Rourke, E,. Ballweber, C., & Popović, Z. (2014). Hint systems may negatively impact performance in educational games. *Proceedings of the First Annual ACM Conference on Learning @ Scale* (L@S '14). (pp. 51-60). dos: https://doi.org/10.1145/2556325.2566248.

O'Rourke, E., Chen, Y., Haimovitz, K., Dweck, C.S., & Popović, Z. (2015). Demographic differences in a growth mindset incentive structure for educational games. In: *Proceedings of the Second Annual ACM Conference on Learning at Scale Works in Progress (L@S WIP 2015).*

O'Rourke, E., Haimovitz, K., Ballweber, C., Dweck, C.S., & Popović, Z. (2014). Brain points: A growth mindset structure boosts persistence in an educational game. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2014)*. https://doi.org/10.1145/2556288.2557157.

O'Rourke, E., Peach, E., Dweck, C.S., & Popović, Z. (2016). Brain points: A deeper look at a growth mind-set incentive structure for an educational game. In: *Proceedings of the ACM Conference on Learning at Scale (L@S 2016)*. https://doi.org/10.1145/2876034.2876040.

Palthepu, S., Greer, J., & McCalla, G. (1991). Learning by teaching. In: *Proceedings of the International Conference on the Learning Sciences*, AACE (pp. 357–363).

Pareto, L. (2009). Teachable agents that learn by observing game playing behaviour. In: *Proceedings of the Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education* (pp. 31-40).

Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *International Journal of AI in Education, 24*, 251-283. https://doi.org/10.1007/s40593-014-0018-8.

Parong, J., Mayer, R. E., Fiorella, L., MacNamara, A., Homer, B. D., & Plass, J. L. (2017). Learning executive function skills by playing focused video games. *Contemporary Educational Psychology*, *51*, 141-151. doi: 10.1016/j.cedpsych.2017.07.002.

Parong, J, Wells, A., & Mayer, R. E. (2020). Replicated evidence towards a cognitive theory of game-based training. *Journal of Educational Psychology*, *112*(5), 922-937. https://doi.org/10.1037/edu0000413.

Pashler H., McDaniel M., Rohrer D., & Bjork R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*. 2008;9(3):105-119. doi:10.1111/j.1539-6053.2009.01038.x

Peddycord-Liu, Z., Cody, C., Kessler, S., Barnes, J.T., & Lynch, C.F. (2017). Using serious game analytics to inform digital curricular sequencing: What math objective should students play next? In: *CHI Play '17: Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. pp. 195–204. https://doi.org/10.1145/3116595.3116620.

Peddycord-Liu, Z., Catete, V., Vandenberg, Barnes, J.T., Lynch, C., & Rutherford, T. (2019). A field study of teachers using a curriculum-integrated digital game. In: *Proceedings of CHI Conference on Human Factors in Computing Systems (CHI '19), May 4–9, 2019, Glasgow, Scotland UK.* ACM, New York, NY, USA 12 Pages. https://doi.org/10.1145/3290605.3300658.

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

Prates, M.O.R., Avelar, P.H.C., & Lamb, L. (2019). Assessing gender bias in machine translation -- A case study with Google Translate. *Computers and Society.* doi: 10.1007/s00521-019-04144-6*.*

Prensky, M. (2006). *Don't bother me, Mom, I'm learning! How computer and video games are preparing your kids for 21st century success and how you can help*. St. Paul: Paragon House.

Prensky, M. (2004). *Digital game-based learning*. McGraw-Hill. ISBN:0071454004; ISBN13:9780071454001

Prizant, B. M., Wetherby, A. M., Rubin, E., & Laurent, A. C. (2003). The SCERTS model: A transactional, family-centered approach to enhancing communication and socioemotional abilities of children with autism spectrum disorder. *Infants & Young Children*, *16*(4), 296-316.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B. et al. (2009). Scratch: Programming for all. *Communications of the ACM 52*, 11, 60–67. https://doi.org/10.1145/1592761.1592779.

Rollings, A. & Morris, D. (2000). *Game architecture and design: Learn the best practices for game design and programming*. Coriolis. ISBN-13: 978-1576104255; ISBN-10: 1576104257.

Rowe, E., Asbell-Clarke, J., Bardar, E., Almeda, M.V., Baker, R.S., Scruggs, R., & Gasca, S. (2020). Advancing research in game-based learning assessment: Tools and methods for measuring implicit learning. In: *Advancing Educational Research With Emerging Technology.* doi: 10.4018/978-1-7998-1173-2.ch006.

Rowe, E., Almeda, M.V., Asbell-Clarke, J., Scruggs, R., Baker, R., Bardar, E., & Gasca, S. (2021). Assessing implicit computational thinking in Zoombinis puzzle gameplay. *Computers in Human Behavior,* *120*. Special Issue on “Towards Strengthening Links between Learning Analytics and Assessment: Challenges and Potentials of a Promising New Bond.” https://doi.org/10.1016/j.chb.2021.106707.

Rowe, J.P. & Lester, J.C. (2010). Modeling user knowledge with dynamic Bayesian Networks in interactive narrative environments. In: *Proceedings of the Sixth Annual Artificial Intelligence and Interactive Digital Entertainment Conference,* 57–62. Palo Alto, CA: AAAI Press.

Rowe, J.P., Shores, L.R., Mott, B.W., & Lester, J.C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education 21*(2): 115–133. http://dx.doi.org/10.3233/JAI-2011-019.

Ruseti, S., Dascalu, M., Johnson, A.M., Balyan, R., Kopp, K.J., McNamara, D.S., Crossley, S.A., & Trausan-Matu, S. (2018). Predicting question quality using recurrent neural networks. In: *Proceedings of the International Conference on Artificial Intelligence in Education (AIED 2018)* (pp. 491-502). Springer, Cham. https://doi.org/10.1007/978-3-319-93843-1_36.

Sabourin, J.L., Mott, B.W., & Lester, J.C. (2011). Modeling learner affect with theoretically grounded dynamic Bayesian Networks. In: *Proceedings of the Fourth International Conference on Affective Computing and Intelligent Interaction*, 286–295. Berlin: Springer. doi:10.1007/978-3-642-24600-5_32

Sabourin, J.L., Shores, L.R., Mott, B.W., & Lester, J.C. (2013). Understanding and predicting student self-regulated learning strategies in game-based learning environments. *International Journal of Artificial Intelligence in Education*, *23*: 94-114. doi 10.1007/s40593-013-0004-6. https://doi.org/10.1007/s40593-013-0004-6.

Salen, K. & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press. ISBN: 9780262240451.

Sawyer, R., Smith, A., Rowe, J., Azevedo, R., & Lester, J. (2017). Is more agency better? The impact of student agency on game-based learning. In: Andŕe, E., Baker, R., Hu, X., Rodrigo, M.M.T., du Boulay, B. (eds.) *Proceedings of the International Conference on AI in Education (AIED 2017)*. LNCS, vol. 10331, pp.335–346. Springer, Cham. https://doi.org/10.1007/978-3-319-61425-0_28.

Schöbel, S., Saqr, M., & Janson, A. (2021). Two decades of game concepts in digital learning environments–A bibliometric study and research agenda. *Computers & Education*, *173*, 104296. https://doi.org/10.1016/j.compedu.2021.104296.

Segedy, J. R., Biswas, G., Blackstock, E. F., & Jenkins, A. (2013a). Guided skill practice as an adaptive scaffolding strategy in open-ended learning environments. In: *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 532-541). Springer, Berlin, Heidelberg. http://dx.doi.org/10.1007/978-3-642-39112-5_54.

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2013b). The effect of contextualized conversational feedback in a complex open-ended learning environment. *Educational Technology Research and Development*, *61*(1), 71-89. http://dx.doi.org/10.1007/s11423-012-9275-0.

Segedy, J.R., Biswas, G., & Sulcer, B. (2014). A model-based behavior analysis approach for open-ended environments. *Journal of Educational Technology & Society*, *17*(1), 272-282.

Segedy, J.R., Kinnebrew, J.S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviors in open-ended learning environments. *Journal of Learning Analytics, 2*(1), 13-48. https://doi.org/10.18608/jla.2015.21.3.

Self, J. (2016). The birth of IJAIED. *International Journal of Artificial Intelligence in Education 26*: 4-12. doi 10.1007/s40593-015-0040-5. https://link.springer.com/article/10.1007/s40593-015-0040-5.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and Instruction* 55, 2 (2011), 503-524.

Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., Ventura, M. & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education 86* (2015) 224-235. https://psycnet.apa.org/doi/10.1016/j.compedu.2015.08.001.

Shute, V. J., Rahimi, S., & Smith, G. (2019). Chapter 4: Game-based learning analytics in physics playground. In: A. Tlili and M. Chang (eds.), *Data Analytics Approaches in Educational Games and Gamification Systems*, Smart Computing and Intelligence, https://doi.org/10.1007/978-981-32-9335-9_4.

Shute, V. J., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C.-P., Kuba, R., Liu, Z., Yang, X. & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*, 127-141. https://doi.org/10.1111/jcal.12473.

Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment*. Cambridge, MA: The MIT Press.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in Newton's playground. The Journal of Educational Research, 106(6), 423-430. https://doi.org/10.1080/00220671.2013.832970.

Sjödén, B., Lind, M., & Silvervarg, A. (2017). Can a teachable agent influence how students respond to competition in an educational game? In: *Proceedings of the 18th International Conference on Artificial Intelligence in Education (AIED 2017).* https://doi.org/10.1007/978-3-319-61425-0_29.

Spann, C. A., Shute, V. J., Rahimi, S., & D'Mello, S. K. (2019). The productive role of cognitive reappraisal in regulating affect during game-based learning. *Computers in Human Behavior, 100*, 358–369. https://doi.org/10.1016/j.chb.2019.03.002.

Squire, K. & Jenkins, H. (2003). Harnessing the power of games in education. *Insight, 3*(7), 5-33.

Sun, C., Shute, V. J., Stewart, A. E. B., Beck-White, Q., Reinhardt, C. R., Zhou, G., Duran, N., & D'Mello, S. K. (2022). The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. *Computers in Human Behavior*, *128*, [107120]. https://doi.org/10.1016/j.chb.2021.107120

Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers & Education, 143,* [103672]. https://doi.org/10.1016/j.compedu.2019.103672.

Taub, M., Sawyer, R., Smith, A., Rowe, J., Azevedo, R. & Lester, J. (2020). The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment. *Computers & Education 147,* 1-19. https://doi.org/10.1016/j.compedu.2019.103781.

Tahir, F., Mitrovic, A., & Sotardi, V. (2020) Investigating the Effects of Gamifying SQL-Tutor. In: So, H.J. et al. (Eds.) *Proceedings of the 28th International Conference on Computers in Education*, pp. 416-425. Asia-Pacific Society for Computers in Education. ISBN978-986-97214-5-5.

Tobias, S. & Fletcher, J. D. (2011). *Computer games and instruction.* Charlotte NC: Information Age. https://eric.ed.gov/?id=ED529495.

Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, *35*(3), 407-420. https://doi.org/10.1111/jcal.12347.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227-265. https://psycnet.apa.org/record/2007-08276-002.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist, 46*(4), 197-221. https://doi.org/10.1080/00461520.2011.611369.

Vygotsky, L.S. (1978). In: M. Cole, V. John-Steiner, S. Scribner and E. Souberman, (Eds.), *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press. ISBN 9780674576292

Waalkens, M., Aleven, V., & Taatgen, N. 2013. Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems. *Computers & Education, 60*, 159-171. https://doi.org/10.1016/j.compedu.2012.07.016.

Wang, P., Rowe, J. P., Min, W., Mott, B. W., & Lester, J. C. (2017). Interactive narrative personalization with deep reinforcement learning. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (pp. 3852-3858). https://doi.org/10.24963/ijcai.2017/538.

Wang, Y., Nguyen, H., Harpstead, E., Stamper, J. & McLaren, B.M. (2019). How does order of gameplay impact learning and enjoyment in a digital learning game? In: Isotani S., Millán E., Ogan A., Hastings P., McLaren B., Luckin R. (eds). *Proceedings of the 20th International Conference on Artificial Intelligence in Education (AIED 2019)*. LNAI 11625 (pp. 518-531). Springer, Cham. https://doi.org/10.1007/978-3-030-23204-7_43.

Warwick, K. (2017). "A Brief History of Deep Blue, IBM's Chess Computer". *Mental Floss*. From July 29, 2017. Retrieved November 9, 2020. https://www.mentalfloss.com/article/503178/brief-history-deep-blue-ibms-chess-computer.

Wiggins, J., Kulkarni, M., Min, W., Mott, B., Boyer, K., Wiebe, E., & Lester, J. (2018a). Affect-based early prediction of player mental demand and engagement for educational games. In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 14, No. 1). Retrieved from https://ojs.aaai.org/index.php/AIIDE/article/view/13047.

Wiggins, J. B., Kulkarni, M., Min, W., Boyer, K. E., Mott, B., Wiebe, E., & Lester, J. (2018b). User affect and no-match dialogue scenarios: An analysis of facial expression. In: *Proceedings of the 4th International Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction* (pp. 6-14). https://doi.org/10.1145/3279972.3279979.

Wouters, P., & van Oostendorp, H. (Eds.). (2017). *Instructional techniques to facilitate learning and motivation of serious games.* New York: Springer.

Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning* (p. 413–432). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.021.

Yannier, N., Hudson, S.E., Stampfer Wiese, E., & Koedinger, K.R. (2016). Adding physical objects to an interactive game improves learning and enjoyment: Evidence from earthshake. *ACM Transactions Computer-Human Interaction 23*, 4, Article 26 (September 2016), 31 pages. doi: http://dx.doi.org/10.1145/2934668.

Yannier, N., Hudson, S.E., & Koedinger, K.R. (2020). Active learning is about more than hands-on: Mixed-reality AI system to support STEM education. *International Journal of Artificial Intelligence in Education, 30*, pages 74–96. doi: 10.1007/s40593-020-00194-3.

Yannier, N., Hudson, S.E., Crowley, K., Do, Y., & Koedinger, K.R. (2022). Intelligent Science Exhibits: Transforming Hands-on Exhibits into Mixed-Reality Learning Experiences. To appear in the *Journal of the Learning Sciences.*

Yannier, N., Hudson, S. E., Koedinger, K. R., Hirsh-Pasek, K., Golinkoff, R. M., Munakata, Y., ... & Brownell, S. E. (2021). Active learning: "Hands-on" meets "minds-on". *Science*, *374*(6563), 26-30. DOI: 10.1126/science.abj99.

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine, 13*(3), 55-75. http://dx.doi.org/10.1109/MCI.2018.2840738.

Zeng, J., Parks, S., & Shang, J. (2020). To learn scientifically, effectively, and enjoyably: A review of educational games. *Human Behavior and Emerging Technologies* 2(2), 186-195. https://doi.org/10.1002/hbe2.188.

Zimmerman, B.J. & Schunk, D.J. (2008). *Self-regulated learning and academic achievement: Theoretical perspectives.* Second Edition. Lawrence Erlbaum Associates. ISBN: 978-1-4612-8180-1.

Zohaib, M. (2018). Dynamic difficulty adjustment (DDA) in computer games: A review. *Advances in Human-Computer Interaction*, 2018, 1-12. https://doi.org/10.1155/2018/5681652.