Article

Journal of Educational Computing Research
0(0) 1–27

DOI: 10.1177/07356331211031287
journals.sagepub.com/home/jec
SAGE

Domain Knowledge and Adaptive Serious Games: Exploring the Relationship of Learner Ability and Affect Adaptability

Vipin Verma[1], Scotty D. Craig[1], Roy Levy[1], Ajay Bansal[1], and Ashish Amresh[1]

Abstract

Detection and responding to a player's affect are important for serious games. A method for this purpose was tested within Chem-o-crypt, a game that teaches chemical equation balancing. The game automatically detects boredom, flow, and frustration using the Affdex SDK from Affectiva. The sensed affective state is then used to adapt the game play in an attempt to engage the player in the game. A randomized controlled experiment incorporating a Dynamic Bayesian Network that compared results from groups with the affect-sensitive states vs those without revealed that measuring affect and adapting the game improved learning for low domain-knowledge participants.

Keywords

affect-sensitive, quantitative, learning environments, games, interactive, technology, assessment, artificial intelligence, serious game, boredom, flow, frustration

[1]Human Systems Engineering, Arizona State University, Tempe, Arizona, United States

Corresponding Author:

Vipin Verma, Human Systems Engineering, Arizona State University, Tempe, Arizona, United States.
Email: vverma9@asu.edu

Introduction

There is an increase in research that targets computer-enabled educational applications to assess the emotions of their learners and adapt the system accordingly (Harley, 2016). It is a challenging task to develop educational video games and online learning environments that can react to the emotions of their learners as it needs a dependable mechanism to detect their emotions accurately (Harley, 2016). Changes in emotions can affect the learning ability of a learner, and if these changes are not detected, these games and the online environment may fail to provide a way to tackle them (Hascher, 2010). Serious games need to be affect-sensitive if they plan to be effective like a human teacher as it is known that expert tutors are capable of identifying the emotions of their participants and respond accordingly (Goleman, 1995; Lepper & Woolverton, 2002). An affect-sensitive serious game should be able to assess the affective states of participants and provide strategies to keep them engaged in the learning process, arouse their interest, and possibly maximize learning.

Serious games are being used widely to support education. They are utilized at all the stages of education including primary (Papanastasiou et al., 2017), secondary (Earp et al., 2015), tertiary (Rodrıguez-Cerezo et al., 2014), as well as industrial training (Riedel & Hauge, 2011). It has been used to teach diverse topics in the fields such as science (Dias & Agante, 2011), mathematics (Kim & Chang, 2010), social science (Furió et al., 2013), and language (Vos et al., 2011). They promote inquisitiveness while providing a novel, fun, and challenging environment, presumably stimulating learning (Griffiths, 2002).

Affect in Serious Games

Affective states play a vivid role during the learning process. They may occur during the course of interaction with learning technologies such as simulated environments, animations, immersive serious games, intelligent tutoring systems, and simple computer interfaces. Affective states of boredom, flow, and frustration are quite common when learning with technology (D'Mello, 2013). While learning is positively correlated to flow, it is correlated negatively with boredom (Craig et al., 2004) which is in accord with the flow theory explained by (Csikszentmihalyi, 2014).

The learning medium should not only be able to sense the affective state of the learner but also aid them in regulating it (Calvo & D'Mello, 2011). Once sensed, the medium should provide intervention accordingly based on the affective state. For example, hints and explanations can be provided to alleviate frustration, and difficulty levels be reduced to tackle boredom (D'Mello, 2013). These interventions are built in several intelligent tutoring systems such as AutoTutor (Craig et al., 2008), ITSpoke (Forbes-Riley & Litman, 2011), and

Gaze Tutor (D'Mello et al., 2012). However, a serious game that implements these interventions does not exist yet.

Serious games help in maintaining the intrinsic motivation of participants because of their engaging and interactive nature (Amresh et al., 2014; Amresh et al., 2019; Freire et al., 2016; Grund, 2015; Yang, 2012). Aristotle said that people tend to look for pleasure and happiness (Bartlett & Collins, 2011). Thus, a serious game when used with individual personalization capabilities can elicit positive feelings among its players.

A game should maintain the player's skill in equilibrium with the difficulty level posed by the game, in order to keep the player in a state of flow. They should be kept in the Vygotsky's Zone of Proximal Development (Vygotsky, 1978) to maintain flow, in which the game difficulty and player's skill exist in a state of dynamic equilibrium with each other. If a player is not able to deal with the game difficulty level, then they may get frustrated. Contrarily, if the game is too easy for a player, then they may lose interest in the game. This is quite possible in educational video games, where the designers have to pay attention to the game play related goals, along with the player progress (Munz et al., 2007). This can be achieved in several ways. One way is to connect the game play and learning together (Van Eck, 2006). Another way could be to keep them loosely tied to each other (Baron, 2017; Baron & Amresh, 2015). Nonetheless, irrespective of the method chosen, a game should be able to adapt itself based on the affective states of boredom, flow, and frustration to keep the player within the bounds of the zone of proximal development.

Stealth Assessment

Player interactions and actions within the game can be mined to assess their performance in a stealth manner (Shute, 2011). This kind of assessment helps eradicate the test anxiety which is often associated with the non-stealth forms of assessment (Shute & Wang, 2015) while supporting the game to adapt itself to the skill level of the player. Detecting the affective state of a user and using it in a stealth manner adds to the existing plethora of stealth assessment techniques that exists (Verma et al., 2019). Not only the player interactions and actions, but their affective state can be mined to adapt their learning experience in a stealth manner and keep them in a state of flow (Hendrix et al., 2018). Therefore, affect detection can have a positive influence on the educational game design by keying out the game play elements that are causing frustration among the players or hindering the player's progress in the game.

Stealth assessment is free from any kind of equipment that uses sensors. Recently, many such models have been developed that do not use an intrusive sensor for detecting the affect of an individual. Such models have been used to measure boredom (Baker et al., 2012; D'Mello et al., 2008; Sabourin et al., 2011), engagement (Baker et al., 2012; D'Mello et al., 2008; Sabourin et al., 2011), confusion (Liu et al., 2013; Pardos et al., 2014), and frustration (Baker et al., 2012; D'Mello et al., 2008; Liu et al., 2013; Paquette et al., 2014; Pardos et al., 2014). These models were fairly successful in differentiating the several affective states and achieved an accuracy better than chance level.

Adaptability in Serious Games

There is no concrete evidence suggesting that adaptive serious games are better than their non-adaptive counterparts. Limited and conflicting research exists that compares the two. Sampayo-Vargas et al. (2013) manipulated the game difficulty in their study using responses to the game objective. Incorrect responses would cause the game difficulty to decrease while correct responses have an opposite effect in the adaptive version of the game. Their study indicated better learning outcomes in the adaptive version of the game. Holmes et al. (2009) evaluated the effect of adaptive games on the performance of the working memory. They matched the game difficulty to the player's performance in the adaptive version to keep them at the edge of their working memory limits and found substantial improvement in working memory as well as mathematical abilities. In another study by van Oostendorp et al. (2014), the adaptive version of the game led to better learning gains. They adapted the game to a complexity level based on the previous scores of the players. Similarly, Ali and Sah (2017) found that adapting the game based on the current knowledge level led to better and faster performance of participants.

On the other hand, there are research studies that say that adaptation makes no difference. Vanbecelaere et al. (2020) found that game adaptation did not cause any significant improvement in cognitive and non-cognitive factors. Their game involved various exercises, and they reduced or increased the number of exercises based on participant's performance in previous exercises. They also raised the threshold to be a 65% score in the adaptive version to proceed to the next exercise while there was no such threshold in the non-adaptive version of the game. In a recent study by Shute et al. (2020), the sequence in which the game levels were presented to the participant was determined using an algorithm in the adaptive version of the game. They found no significant difference of adaptation on participant learning either. Orvis et al. (2008) and Plass et al. (2019) corroborated these results and found no significant difference in their results.

There is contradictory evidence to the usefulness of game adaptation probably based on how the adaptation is built into the game. The current research adapts the game based on the affective states of the player. Although D'Mello et al. (2010) examined the part that affect plays in an Interactive Tutoring System, there is very limited research that examines the role of affect in serious game adaptation. The outcome of a study by D'Mello et al. (2010) suggested that the adaptive systems are more helpful to a learner who has lower domain knowledge compared to the ones who have higher knowledge. The current research expands on this finding in the context of a serious game.

Dynamic Bayesian Network

A Bayesian Network (BN) approach uses probabilistic graphical modeling in which several variables have conditional dependence on each other (Friedman et al., 1997). A BN represents a graphical structure made up of nodes and directional links or edges between them. The nodes represent continuous or discrete variables, and the link represents conditional dependence between them. Each node in the structure has a probability distribution attached to it dependent on the nodes that have directed links flowing into it. These probabilities can either be learned from data or assigned by an expert rater. A Dynamic Bayesian Network (DBN) is a kind of BN in which variables have probabilistic dependence over a period called lag or time-steps, allowing for modeling sequences and time-series (Reichenberg, 2018; Reye, 2004).

Figure 1 shows a simple DBN called knowledge tracing (Corbett & Anderson, 1994). The network shows a 2-quiz series that have four performance parameters and three nodes associated with it. These parameters are prior knowledge, slip rate, learn rate, and guess rate. The three nodes are: a participant node (S), a knowledge node (K), and a question node (Q). The participant node represents an individual learner and governs the prior knowledge parameter P(L). The prior knowledge parameter describes the initial knowledge level of a participant that they possess before playing the game. It can be obtained through a diagnostic assessment. The knowledge node depicts the state of participant knowledge at any point in time and is dependent upon the prior knowledge that the participant has. In typical applications of knowledge tracing, it is a discrete node that has two states, namely true and false which represent the possible states of the participant having or not having the knowledge,

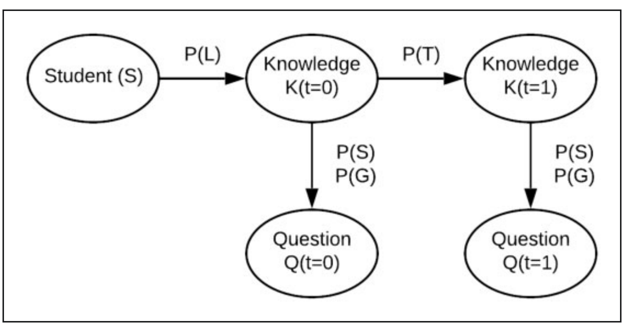

Figure 1. Bayesian Knowledge Tracing Adopted from Pardos & Heffernan (2010) Showing Two Time Slices.

respectively. The knowledge node is time-dependent, also called a temporal node, and is therefore replicated across both the time-steps. The knowledge node at any time-step is directly dependent on the knowledge node from the previous time-step. This conditional dependence of knowledge with itself based on time lag is represented using transition or learn rate P(T), which expresses the probability that a participant will transition from an unlearned to learned state in the next time-step. In the knowledge tracing model, it is typically assumed that it is not possible to lose knowledge on transition, and therefore the probability to go from a learned state to an unlearned state in the next time-step is zero (Corbett & Anderson, 1994). The question node denotes the question which is asked to gauge the knowledge of the participants. It has two states, true and false, which corresponds to the participant's answer being correct or incorrect and is modeled as being dependent on the time-specific knowledge level of the participant. The question node is also temporal and therefore replicated across both the time-steps as shown in Figure 1. It has two associated parameters, guess P(G) and the slip rate P(S). Guess rate models the probability of guessing correctly when a participant does not have the knowledge, while slip rate accounts for answering incorrectly despite having the required knowledge. The current study employs a more complex model, in which there are multiple observables at each time-step, linked to the time-specific knowledge node.

Current Study

Attempts have been carried out in the past to predict the cognitive-affective states (affective states) of flow, frustration, and boredom (Craig et al., 2008; Verma et al.,2020). The current study focuses on facial emotion tracking using Affdex Software Development Kit (SDK) from Affectiva. An easily available webcam is used to capture the facial features using the SDK which is then used to detect the emotions. A relatively high detection rate can be attained for facial features if the user is front- facing the camera and environment lighting is set up properly (Magdin & Prikler, 2018). Successful emotion detection can positively impact the fast-paced immersive environments by providing real-time adaptation capabilities based on the learners' emotions. This individual personalization becomes paramount during the times of crisis such as the era of COVID-19 because of an increase in usage of online education (Zhou et al.,2020).

The current study was formulated to gauge the effect of dynamic adaptation in educational video games on the skill acquisition of its players. The dynamic adaptation is carried out using the online assessment of player affect using facial emotion tracking. It is hypothesized that this dynamic adaptation will be more beneficial to players whose initial skill level is low compared to the ones who already have high skills before playing the game.

Material and Methods

The experiment involved a randomized control group design with a pre-test. Participants were randomly assigned to either the test group or the control group. In the treatment group, they played in a dynamic game difficulty setting, adapted using the facial emotion tracking, while participants in the control group played at a constant difficulty throughout the game irrespective of their affective states. The test group involved adaptation, while there was no adaptation in the control group. The game adaptation was manipulated to measure its effect on the increase in player knowledge as a result of game play.

Participants

A total of 107 undergraduate participants (29 female, 78 male, average age 18.9 years, a standard deviation of 2.6) were recruited from a university located in the southwestern United States for this study. The participation lasted up to 1.5 hours (mean game play time of 42.1 minutes, a standard deviation of 8.7 minutes) and the participants were offered a 1.5 course-credit for their participation. 57% of the participants reported that they have previously played games before taking part in this experiment, and their average game play time was 6 hours per week with a standard deviation of 8.82 hours.

Material

Content Agnostic Game Engineering (CAGE) architecture (Baron, 2017; Baron et al., 2016) was used to create a 2 D platformer game called “Chemo-o-crypt” in Unity3D (v2018.1.9f2). In the CAGE architecture, game play mechanics are designed to be independent of the educational content of the game, allowing the mechanics to be re-usable across content domains. A recent experiment by Atmaja et al. (2020) revealed the advantages of using CAGE architecture to save time and money and make the serious game development process more efficient.

Right and left player movement, jumping, and ladder climbing were the game mechanics employed in the game Chemo-o-crypt. The player’s health is lost partially on colliding with the enemies in the game, which are patrolling various areas. This health loss is dependent on the game difficulty which ranges from one to four. At the lowest difficulty setting, only 25% of the health is lost, but at the highest difficulty setting, all the health is lost. Further, higher difficulty levels feature more difficult game play elements such as moving platforms. The game consists of four content levels, which are separate from the game difficulty levels. Spikes and water, shown in Figure 2, are the environmental hazards present in the game that can reduce the available player health to zero when they come in contact with them. Heart-shaped items (1-up) and coins are distributed throughout the game. A player starts with three lives which can be raised by using a 1-up or accumulating one hundred coins.

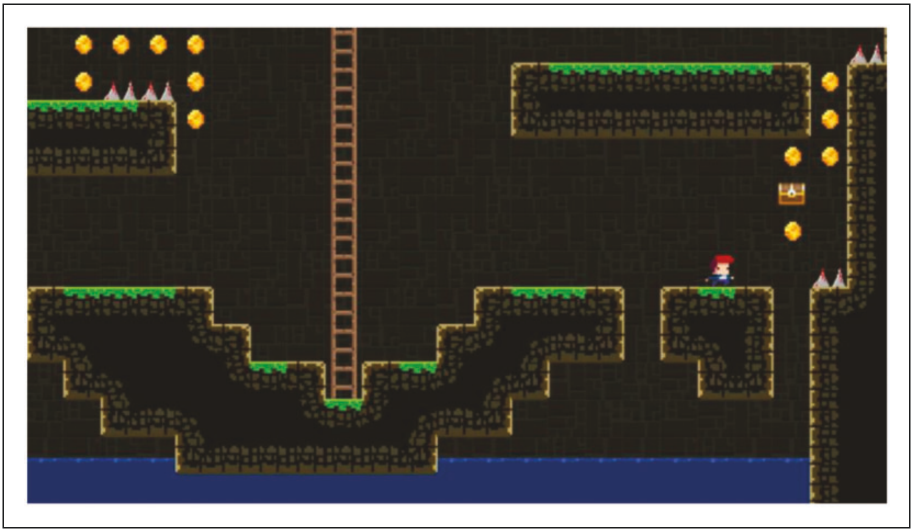

Figure 2. Screen Capture Showing the Spike and Water Hazard in Chem-o-crypt.

Figure 3. Screen Capture of the Goal of the Game Chem-o-crypt.

Each content level aims to balance the chemical equation by collecting the required number of elements and molecules that partake in the chemical reaction. 3 Oxygen (O_2) and 2 Ozone (O_3) molecules are needed to balance the chemical equation depicted in Figure 3. However, there will be an excess of

Figure 4. Screen Capture of the Level-End Task or Quiz.

these molecules present in the game environment, called distractors. There are 3 distractors present on every content level to make the game more challenging. These distractors when collected cause player kickback and may cause some health loss depending on the content level. When a player is near a collectible, it can randomly become either a distractor or a required molecule. All the distractors and molecules have an equal probability of showing up in the player proximity. When all the required molecules are collected to balance the equation, the player needs to collect the" GO" text upon which the same equation is presented to them (as a quiz) which they balanced through the game play mechanics (Figure 4). Once they hit the submit answer button on balancing the equation, the next context level loads irrespective of the wrong or right answer. There are four content levels and the balanced equation for each level is written below:

1. $2\,O_3 \rightarrow 3\ O_2$
2. $N_2 + 3\ H_2 \rightarrow 2\ NH_3$
3. $ZnS + 2\ HCl \rightarrow ZnCl_2 + H_2S$
4. $Al_2O_3 + 6\ HCl \rightarrow 2\ AlCl_3 + 3\ H_2O$

Each content level in Chemo-o-crypt is divided into 4 navigable chunks that lie next to each other in a sequence. Each chunk can have a game scene populated in it depending on the game difficulty. Therefore, each chunk has four possible scenes that it can be populated with. Consequently, there are 4×4, i.e. 16 maximum possible layouts for a level at any point in time depending on the

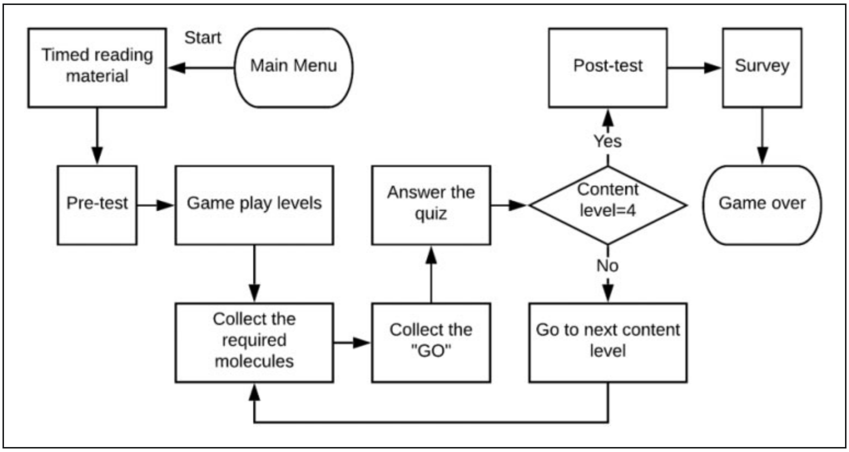

Figure 5. Flowchart Depicting a Typical Participant Workflow.

game difficulty. The chunks are continuous and therefore a player can easily navigate from one chunk to another. However, they can do so only when they are on the ground level. Players spawn in the first chunk during the first content level, the second chunk during the second content level, and so on. The default difficulty level is one when players start playing the game. Therefore, they begin in the first chunk when the game starts. A flowchart depicting the participant workflow during their participation is shown in Figure 5.

Affdex Software Development Kit

Affdex Software Development Kit (SDK) from Affectiva (Magdin & Prikler, 2018) was integrated into the Chemo-o-crypt game. The SDK outputs the probabilities for various emotions of the player by tracking their facial features. The output rate of the SDK was 20 Hz. The participant's face was captured using a template size of 640 by 480px (height by width). Seven basic Ekman emotions of Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Contempt were traced by the SDK in real-time. The output probability score increments from 0 (emotion absent) to 100 (emotion fully present) as the emotion appears and escalates.

Affect Detection and Adaptive Algorithm

During the game, player emotions were continually tracked. The tracked emotions were classified in real-time into the affective states of boredom, flow, and frustration using the following algorithm (Baron, 2017).

- If all the emotions are below the threshold, then the player is classified in a BORED state, unless they were in a state of FLOW previously.

- If any of the emotions is above the threshold, then the player is in a non-bored state.
  - If anger is above the threshold and happiness is below the threshold, then the player is classified to be in a FRUSTRATION state.
  - If surprise is above the threshold and sadness is below the threshold, then the player is classified to be in a FLOW state.
- If the above rules fail, then the player is classified to be in a state called NONE

For anger, a threshold of 10 was used, while for all other emotions a threshold of 20 was used. This was intentionally done due to the difficulty in detecting anger (Craig et al., 2008). This classification is then used to dynamically adjust the game difficulty depending on the test and control groups. If a state of boredom was detected in the test group, the difficulty is increased if it is less than 4 (since the maximum value of difficulty is 4). On detecting frustration, the difficulty level is reduced by 1 if it is higher than 1. There were four content levels as stated previously and the experiment ended when the player cleared all four levels.

Procedure

The experiment took place in a computer lab with maximum participation of 20 participants at any point in time. Participants signed the consent form electronically and sat approximately 60 cm away from the monitor. To avoid any hindrance in the facial emotion detection process, participants were requested to remove their caps and glasses and to abstain from masking their faces with their hands while playing the game. Upon consenting to the participation, they downloaded the game and started it. As the game started, it assigned the participants into either test or the control group randomly. Then they played the game as per the workflow depicted in Figure 5 until they finish it and were rewarded course-credits upon game completion.

Bayesian Tracking

The DBN employed in the game Chem-o-crypt is shown in Figure 6. The network was created using Bayes Server 8.17 (BayesServer, 2020). There is a total of eleven nodes, of which five are temporal or time-series nodes. A description of each node is available in Table 1. Prior node models the prior knowledge that the participant has before starting the game play. A pre-test consisting of 20 questions about balancing chemical equations is used to assign the value to the Prior node. As an example, if a player correctly answered 13/20 questions, then their Prior node will be assigned a state of priorScore6 out of the 11 possible states [0-10] for this node. This will then act as evidence for the latent node

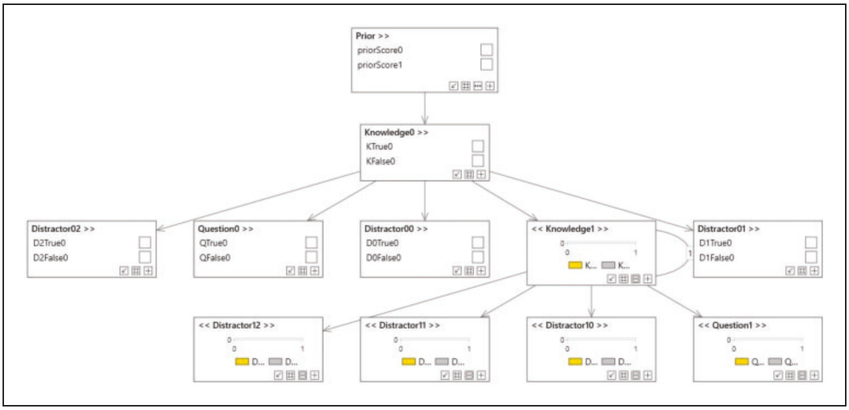

Figure 6. DBN Employed in the Game.

Knowledge0, which governs the probabilities of the nodes which are dependent on Knowledge0. Knowledge0 represents the probability that the participant has the knowledge required to balance the chemical equations at time-step $t = 0$, i.e. on the content level 1. It has two states, true and false, indicating whether the participant has the required knowledge or not. Distractor00, Distractor01, Distractor02 represent the probability that the participant picks up the three respective distractors on the 1st level ($t = 0$). Question0 denotes the probability that the participant answered the level 1 quiz correctly or incorrectly. In Figure 6, the Knowledge1 node is linked to itself with a temporal order of 1, which means that the knowledge on the subsequent level (current time-step) is dependent on the knowledge gained in the last level ($t = t\ 1$). Distractor10, Distractor11, Distractor12 represent the probability that the participant picks up the three respective distractors on later levels ($t = 1, 2, 3$). Question1 denotes the probability that the participant answered the level 2,3, and 4 quizzes correctly or incorrectly.

A simplified version of this DBN is shown in Figure 7. It is similar to the one in Figure 6 except that level 1 ($t = 0$) nodes are not depicted. They are not missing but merged in their temporal counterparts. In doing so, the probability of knowledge, picking up distractors, answering questions is the same across all four levels while it is different for the first level in the complex version of the network. In the network shown in Figure 6, the probabilities are different for the first level ($t = 0$), compared to the rest of the levels ($t = 1, 2, 3$), by design. This is intentionally done to allow the player to learn the mechanics of the game during the first level. When they start playing the game, participants are not aware that they are not supposed to collect distractors. However, upon doing so, they get a kickback and a health loss depending on the content level along with a message that says that they do not need any more molecules of that type. As an example,

Table 1. Description of Nodes in the DBN Shown in Figure 6.

Node name	Type	Time-steps	Conditional dependency	States	Description
Prior	Initial	$t = 0$	None	11 states indicating score in the range [0-10]	A state of 0 denotes a participant who scored a 0 in the pre-test while a state of 10 represents a participant who scored 100% in the pre-test
Knowledge0			Prior	True and False	State of true denotes the possibility that the participant has the required knowledge at timestep $t = 0$
Distractor00 Distractor01 Distractor02			Knowledge0		True denotes the evidence that the participant has collected this distractor at timestep $t = 0$
Question0					True denotes the evidence that the participant answered the quiz correctly at timestep $t = 0$
Knowledge1	Temporal	$t = 1, 2, 3$			True denotes the possibility that the participant has the required knowledge at timestep $t = 1, 2, 3$
Distractor10 Distractor11 Distractor12			Knowledge1		True denotes the evidence that the participant has collected this distractor at timestep $t = 1, 2, 3$
Question1					True denotes the evidence that the participant answered the quiz correctly at timestep $t = 1, 2, 3$

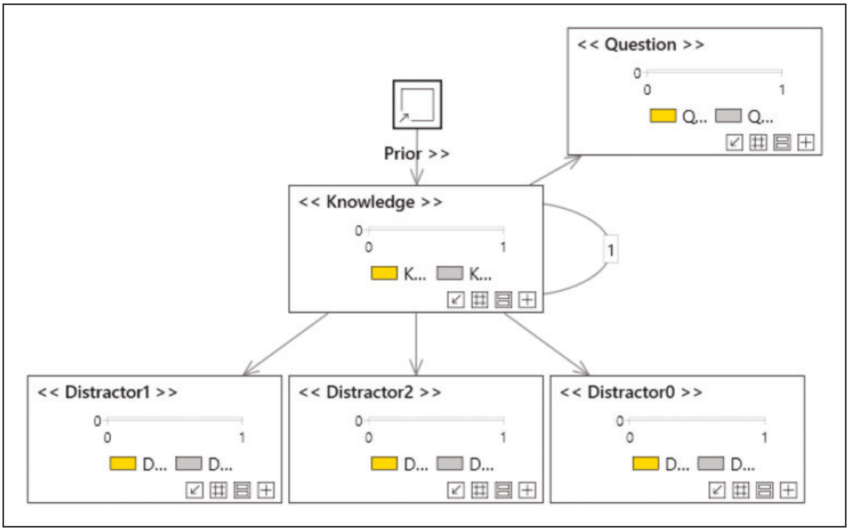

Figure 7. A Simplified Version of the DBN Used in the Game.

Figure 8. The Message that Appears when the Player Collects a Distractor.

Figure 8 shows a screenshot of the message that appears when the player collects excess of the Ozone (O_3) molecules. There is no health loss on content level one, and instant player death on content level 4 if they pick up the distractors.

Table 2. Learned Conditional Probabilities for the Prior Node.

Data source	Prior score (pre-test score)										
	0	1	2	3	4	5	6	7	8	9	10
Overall data	.00	.00	.00	.03	.04	.05	.04	.07	.27	.44	.07
Test group	.00	.00	.00	.02	.02	.02	.06	.07	.27	.45	.09
Control group	.00	.00	.00	.04	.06	.08	.02	.08	.27	.42	.04

Therefore, at content level 1, while they are learning the mechanics of the game, the probability of picking up the distractors is higher as compared to the rest of the levels.

Analysis

The affect classification carried out using the algorithm stated previously, pre-test score (prior), evidence of collecting distractors, and the quiz responses of the participants were collected in a log file. This data is then used for parameter learning in the Bayes

Server 8.17 using Log-Likelihood as the convergence method and a rolling time-series mode (BayesServer,2020). This allowed for learning the conditional probabilities of the nodes which are part of the DBN which can then be compared across the test and the control groups to evaluate the hypothesis. The analysis is done for the cumulative data set, as well as the test and the control groups.

Results

The test group (n = 55) played the adaptive version of the game and 52 in the control group played a static version of the game. In the test group, a total of 41 participants exhausted the available player lives and therefore did not complete all the four content levels. However, 31 participants from the test group still managed to reach content level

4. Exhausting the available lives instead of finishing all the four content levels would mean that less data is available for parameter learning. Due to the presence of the latent nodes (Knowledge0 and Knowledge1), two equivalent solutions were obtained as a result of parameter learning. These solutions suggested the phenomenon of label switching (Jasra et al., 2005). However, the most interpretable solution is presented in this paper.

The parameter learning for the entire data set using Log-Likelihood converged in 23 iterations. Conditional probabilities thus obtained from the parameter learning are summarized in Tables 2 to 4. The probabilities of the Prior node (Table 2) show that most of the participants scored 80% and above in the

Table 3. Learned Conditional Probabilities for Knowledge Node at $t = 0$.

	Knowledge0					
	Overall data		Test group		Control group	
Prior	True	False	True	False	True	False
0	.50	.50	.50	.50	.50	.50
1	.50	.50	.50	.50	.50	.50
2	.50	.50	.50	.50	.50	.50
3	.02	.98	.04	.96	.02	.98
4	.73	.27	.96	.04	.60	.40
5	.58	.42	.96	.04	.45	.55
6	.01	.99	.01	.99	.04	.96
7	.62	.38	.50	.50	.74	.26
8	.69	.31	.74	.26	.65	.35
9	.78	.22	.75	.25	.83	.17
10	.86	.14	.82	.18	.98	.02

Table 4. Learned Conditional Probabilities for Overall Data.

	Distractor00		Distractor01		Distractor02		Question0		Knowledge1	
Knowledge0	True	False	True	False	True	False	True	False	True	False
True	.52	.48	.01	.99	.00	1.00	.97	.03	.53	.47
False	.99	.01	.94	.06	.28	.72	.58	.42	.34	.66
	Distractor10		Distractor11		Distractor12		Question1		Knowledge1	
Knowledge1	True	False	True	False	True	False	True	False	True	False
True	.34	.66	.00	1.00	.00	1.00	.92	.08	.80	.20
False	1.00	.00	.64	.36	.20	.80	.75	.25	.11	.89

pre-test, with almost none answering more than fourteen questions incorrectly. However, there were only a handful of participants who answered all the twenty questions correctly. Therefore, the data set consisted of more people who had a high level of initial knowledge according to the pre-test scores.

Table 4 shows the conditional probabilities of various nodes given the knowledge node at the first content level ($t = 0$). As an example, the probability that a participant picked up the second distractor (Distractor01) given they have the knowledge of the content is 0.01. Similarly, the probability that they answered the quiz incorrectly despite having the knowledge is 0.58. Further, the probability that they have the knowledge on the next level (Knowledge 1 is true), given

Table 5. Learned Conditional Probabilities for the Test Group.

	Distractor00		Distractor01		Distractor02		Question0		Knowledge1	
Knowledge0	True	False	True	False	True	False	True	False	True	False
True	.53	.47	.02	.98	.01	.99	.97	.03	.62	.38
False	.99	.01	.98	.02	.31	.69	.57	.43	.71	.29
	Distractor10		Distractor11		Distractor12		Question1		Knowledge1	
Knowledge1	True	False	True	False	True	False	True	False	True	False
True	.62	.38	.01	.99	.00	1.00	.82	.18	.76	.24
False	.99	.01	.98	.02	.47	.53	.94	.06	.35	.65

Table 6. Learned Conditional Probabilities for the Control Group.

	Distractor00		Distractor01		Distractor02		Question0		Knowledge1	
Knowledge0	True	False	True	False	True	False	True	False	True	False
True	.53	.47	.03	.97	.01	.99	.97	.03	.61	.39
False	.98	.02	.88	.12	.26	.74	.56	.44	.32	.68
	Distractor10		Distractor11		Distractor12		Question1		Knowledge1	
Knowledge1	True	False	True	False	True	False	True	False	True	False
True	.30	.70	.00	1.00	.00	1.00	.93	.07	.78	.22
False	.99	.01	.67	.33	.14	.86	.70	.30	.14	.86

that they possess the knowledge on level 1 (Knowledge 0 is true) is 0.53. Table 4 also shows the results for the higher content levels (two and above). For example, the probability that the participant will pick up the third distractor (Distractor12) above level 1 given that they do not have the knowledge on that level is 0.20. Similarly, looking at the last column of data, the probability that they will lack the skill on the next level (Knowledge 1 is false) given that they possess the skill on the current level (Knowledge 1 is true) is 20%. Table 5 shows the learned probabilities for the test group, while Table 6 is for the control group.

Parameter learning of the structured DBN revealed the probabilities associated with various events happening in the game. Results of the analysis at time $t = 0$ for level 1 are as expected. There is a 52% chance to pick up the first distractor despite having the knowledge, and a 99% chance if the knowledge is missing. On picking up the first distractor, players get a kickback and a feedback message not to pick them up, and therefore the probability of picking up the second distractor went down to 1%, provided they have the knowledge. However, it remained high (0.94) for the less skilled ones who do not have the knowledge yet. The probability of collecting the third distractor went even further down to almost 0% for knowledgeable ones, suggesting the possible effectiveness of the feedback system. However, it could also be possible that the third distractor did not show up in the player's surroundings before they could finish the level, since it is a random process. Slip rate for answering the question incorrectly when players possess the knowledge remained low (3%), while the guess rate for guessing the answer correctly despite no knowledge was high (58%). However, the conditional probabilities of having the knowledge on level 2, given the knowledge on level 1 were unexpected. Ideally, it is assumed that once a player has gained knowledge, they are going to retain it 100% (Pardos & Heffernan, 2010). But the results show sustenance of about 53% only, a loss of 47%. Results show a low probability of 34% that a player who is not skilled on level 1 will transition to get the skill on the next level.

Conditional probabilities obtained for $t > 0$ were as expected. The probability of picking up the first distractor, given the player has the knowledge, decreased to 34%. For the second and third distractor, they are almost zero. However, for the players with a low knowledge level, the probabilities of picking up the distractor remained high at 100%, 64%, and 20% respectively. The slip rate for answering the question increased from 3 to 8%, while the guess rate also increased from 58% to 75%. The knowledge retention rate increased from 53% to 80%, while the transition rate decreased from 34% to 11%.

Adaptive Condition

The parameter learning for the test data using Log-Likelihood converged in 54 iterations. Conditional probabilities thus obtained from the parameter learning are summarized in Tables 2, 3, and 5. The probabilities of the Prior node (Table 2) show that 80% of the participants in the test group scored 80% and above in the pre-test.

Non-Adaptive Condition

The parameter learning for the control data using Log-Likelihood converged in 25 iterations. Conditional probabilities thus obtained from the parameter learning are summarized in Tables 2, 3, and 6. The probabilities of the Prior node (Table 2) shows that 72% of the participants in the control group scored 80% and above in the pre-test.

Does Dynamic Adaptation Affect Players?

Separate parameter learning for the test and the control groups revealed some differences between the two. Conditional probabilities obtained for $t = 0$ were similar for the test and control groups, as compared to the overall data, the only difference being the Knowledge1 node. While the probability of retaining the knowledge from level 1 to level 2 was comparable, the probability of transition from no knowledge state to knowledge state was much higher in the test group (71%) compared to the control group (32%). However, the probability of picking up the first distractor at $t > 0$ was higher (62%) in the test group compared to the control group (30%). At $t > 0$, the slip rates and guess rates were also higher for the test group (18% and 94%) in comparison to the control group (7% and 70%). The transition rate at $t > 0$ was also higher for the test group (35%) than the control group (14%).

Discussion

The current study investigated adaptivity within a serious game to determine its effectiveness for higher and lower domain learners. The rate for the transition from no knowledge state to knowledge state supports the hypothesis that adapting the game will be more beneficial for low domain learners. The transition rate was better in the test group (71%) in comparison to the control group (32%). This suggests that the game adaptation was effective in improving the knowledge levels in the test group, and in absence of the adaptation, the transition rate remained relatively low. The adaptive game offered better learning to the individuals with low skill levels, compared to the non-adaptive version of the game in the control group. These findings are important and indicate the importance of interventions to mitigate the negative affective states such as boredom and frustration.

Within an Interactive Tutoring System, learning has been shown to correlate positively with the flow and negatively with the boredom (Craig et al., 2004). The current study supports this finding and expands it to a serious game. Similar to previous research with interactive affective systems, the current work also showed that adaptivity based of off affect can impact learning. Affect adaptation is more important for learners who have low domain knowledge. Present research shows affect adaptability is also useful for improving the performance of low domain learners within a serious game. Therefore, it is important to detect these affective states and provide a way to treat them in a manner that gets a learner more engaged in the learning process. Results from this study are also in agreement with those obtained by D'Mello et al. (2010) using AutoTutor and suggests the importance of affect in the learning process of a learner who has low initial domain knowledge. Although the adaptive system was beneficial for beginners, it did not have any impact on the learners who had high prior knowledge.

Conflicting evidence exists regarding the effect of adaptability on the learning imparted by a serious game. This largely depended on how the adaptation was built into the game (Ali & Sah, 2017; Vanbecelaere et al.,2020). Ali and Sah

(2017) used user ontology and semantic rules to adapt the game and found better learner performance for the adaptive version of the game. On the other hand, Vanbecelaere et al. (2020) adapted the game to show a different number of exercises based on the learner's performance in previous exercises. They found no significant improvement as a result of adaptation. Current research uses the affective states of the learner to adapt the game by altering the game difficulty. When boredom kicks in, game difficulty is increased. This causes the game environment layout to change in a way that makes it harder to navigate around posing a challenge to the learner. Further, the health loss from collisions with the enemy and the enemy movement speed is increased to ramp up the challenge. Very finite research exists that adapts the game in such a manner. Present results may explain the current divide in the literature regarding the effectiveness of adaptation in serious games suggesting that it is effective for low domain learners only.

Previous theories such as Zone of Proximal Development (Vygotsky, 1978) and Knowledge Space Theory (Craig et al., 2013; Falmagne et al., 1990) state that adaptation based on the learner's domain knowledge supports learning because prior knowledge indicates what the learner is ready to learn next. However, the outcome of the current study indicates that the adaptation based on learner's affect is useful as well, especially for low domain learners. These findings are useful to keep the learners at the edge of their abilities by detecting their affect unobtrusively. Affect detection can be combined with other forms of stealth assessment to adapt the game play. For example, it can be used along with Dynamic Bayesian Network to assess the current knowledge level of the learner and provide remediation if the knowledge level falls below a certain threshold. It can be used in conjunction with the mouse tracking, player log data, and other forms of stealth assessment indicated in Verma et al. (2019).

Limitations

A limitation is evident from the prior score distribution in Table 2. The prior score is the evaluation based on the pre-test which determines the participant's knowledge level before the game play. Participants are not evenly distributed across all the groups and most of them have a prior score of 80% and above. The study should involve some participants who have a low level of initial knowledge to prevent the bias that may occur because of this reason. There were not many participants who had extreme scores, i.e. either 0 or 20/20. Therefore, parameter learning did not return the expected probabilities for extreme cases. Table 3 suggests that the probability of knowledge when the participant scored 0 is 50%, which is unexpected. Therefore, these results must be interpreted with caution.

Conclusion

Game adaptation using affect detection supported a considerable improvement in the player learning as evident from the learned probabilities for the test and the control group. Current study provided insight regarding when the educational game should be adapted. It was hypothesized that the dynamically adaptive gaming environment will be more beneficial to players whose initial skill level is low prior to the game play. The learned probabilities affirm this hypothesis. The transition rate for players to shift from low knowledge state to high knowledge state was more than twice in the test group in comparison to the control group. Adaptation based on affect appeared to be effective for low ability learners and therefore the game should be adapted for them. Future research should investigate the low ability learner's affect within these adaptive systems.

Based on current findings, the adaptive games can support learning for low ability learners. However, in a traditional classroom, students are exposed to a uniform curriculum that is not always tailored to individual needs (Zhang et al., 2004). Using the current findings, educational games and online learning systems can provide personalized learning environments that support participant's skill acquisition.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Vipin Verma https://orcid.org/0000-0001-9423-9861
Scotty D. Craig https://orcid.org/0000-0002-4103-5547

References

Ali, A., & Sah, M. (2017). Adaptive game-based e-learning using semantic web technologies. In *2017 International conference on open source systems technologies (ICOSST)* (pp. 15–23). https://doi.org/10.1109/ICOSST.2017.8278999

Amresh, A., Clarke, D., & Beckwith, D. (2014). Gamescapes and simapps: New techniques for integrating rich narratives with game mechanics. In *Proceedings of the European conference on games based learning* (Vol. 1, pp. 18–25). https://www.proquest.com/conference-papers-proceedings/gamescapes-simapps-new-techniques-integrating/docview/1674172735/se-2?accountid=4485

Amresh, A., Verma, V., Baron, T., Salla, R., Clarke, D., & Beckwith, D. (2019). Evaluating gamescapes and simapps as effective classroom teaching tools. In *13th*

European Conference on games based learning (p. 8). http://dx.doi.org/10.34190/GBL.19.151

Atmaja, P. W., Muttaqin, F., & Sugiarto, S. (2020). Facilitating educational contents of different subjects with context-agnostic educational game: A pilot case study. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, *6*(1), 53–65.

Baker, R. S., Gowda, S., Wixon, M., Kalka, J., Wagner, A., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., & Rossi, L. (2012). *Sensor-free automated detection of affect in a cognitive tutor for algebra* [Paper presentation]. Educational Data Mining 2012. https://www.semanticscholar.org/paper/Sensor-free-automated-detection-of-affect-in-a-for-Baker-Gowda/1c7a1e4bff56cf6e9851037ca53c754e8961ce92

Baron, T. (2017). *An Architecture for Designing Content Agnostic Game Mechanics for Educational Burst Games* [Unpublished doctoral dissertation]. Arizona State University.

Baron, T., & Amresh, A. (2015). Word towers: Assessing domain knowledge with non-traditional genres. In *European conference on games based learning* (p. 638).

Baron, T., Heath, C., & Amresh, A. (2016). Towards a context agnostic platform for design and assessment of educational games. In *European conference on games based learning* (pp. 3–4).

Bartlett, R. C., & Collins, S. D. (2011). *Aristotle's Nicomachean ethics*. University of Chicago Press. https://www.google.com/books/edition/Aristotle_s_Nicomachean_Ethics/1t3SMJaTUlwC?hl = en&gbpv = 0

BayesServer (2020). Dynamic Bayesian networks – An introduction. https://www.bayesserver.com/docs/introduction/dynamic-bayesian-networks

Calvo, R. A., & D'Mello, S. K. (2011). *New perspectives on affect and learning technologies* (*Vol. 3*). Springer Science & Business Media.

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction*, *4*(4), 253–278.

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with autotutor. *Journal of Educational Media*, *29*(3), 241–250.

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with autotutor: Applying the facial action coding system to cognitive – Affective states during learning. *Cognition & Emotion*, *22*(5), 777–788.

Craig, S. D., Hu, X., Graesser, A. C., Bargagliotti, A. E., Sterbinsky, A., Cheney, K. R., & Okwumabua, T. (2013). The impact of a technology-based mathematics after-school program using aleks on student's knowledge and behaviors. *Computers & Education*, *68*, 495–504.

Csikszentmihalyi, M. (2014). Toward a Psychology of Optimal Experience. In Flow and the Foundations of Positive Psychology (pp. 209–226). Dordrecht: Springer. https://doi.org/10.1007/978-94-017-9088-8_14

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.

D'Mello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, *18*(1–2), 45–80.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Perkins, L., & Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In Aleven V., Kay J., & Mostow J. (Eds.), *Intelligent Tutoring Systems*. ITS 2010. Lecture Notes in Computer Science (vol. 6094). Springer. https://doi.org/10.1007/978-3-642-13388-6_29

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, *70*(5), 377–398.

Dias, M., & Agante, L. (2011). Can advergames boost children's healthier eating habits? A comparison between healthy and non-healthy food. *Journal of Consumer Behaviour*, *10*(3), 152–160.

Earp J., Catalano C. E., & Mortara M. (2015). Investigating the Deployment of Serious Games in Secondary Education: A Pilot Study Inspired by Design-Based Research. In De Gloria A. (Ed.), *Games and Learning Alliance*. GALA 2014. Lecture Notes in Computer Science (vol. 9221). Springer. https://doi.org/10.1007/978-3-319-22960-7_2

Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., & Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test, and search them. *Psychological Review*, *97*(2), 201–224.

Forbes-Riley, K., & Litman, D. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication*, *53*(9–10), 1115–1136.

Freire, M., Serrano-Laguna, Á., Manero, B., MartÍnez-Ortiz, I., Moreno-Ger, P. & Fernández-Manjón, B. (2016). Game learning analytics: Learning analytics for serious games. In *Learning, design, and technology* (pp. 1–29). Springer Nature Switzerland AG.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*(2/3), 131–163.

Furió, D., GonzáLez-Gancedo, S., Juan, M.-C., SegúI, I., & Rando, N. (2013). Evaluation of learning outcomes using an educational iphone game vs. traditional game. *Computers & Education*, *64*, 1–23.

Goleman, D. (1995). Emotional Intelligence. Bantam Books.

Griffiths, M. D. (2002). The educational benefits of videogames. *Education and Health*, *20*(3), 47–51.

Grund, C. K. (2015). How games and game elements facilitate learning and motivation: A literature review. In Cunningham, D. W., Hofstedt, P., Meer, K. & Schmitt, I. (Hrsg.), *INFORMATIK 2015* (pp. 1279–1293). Bonn: Gesellschaft für Informatik e.V. https://dl.gi.de/handle/20.500.12116/2112

Harley, J. M. (2016). Measuring emotions: A survey of cutting edge methodologies used in computer-based learning environment research. In Tettegah S. Y., & Gartmeier M. (Eds.), *Emotions, technology, design, and learning* (pp. 89–114). Elsevier.

Hascher, T. (2010). Learning and emotion: Perspectives for theory and research. *European Educational Research Journal*, *9*(1), 13–28.

Hendrix, M., Bellamy-Wood, T., McKay, S., Bloom, V., & Dunwell, I. (2018). Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*, *11*(4), 320–327.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–F15.

Jasra, A., Holmes, C. C., & Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, *20*(1), 50–67.

Kim, S., & Chang, M. (2010). Computer games for the math achievement of diverse students. *Journal of Educational Technology & Society*, *13*(3), 224–232.

Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In Aronson J. (Ed.), *Improving academic achievement* (pp. 135–158). Elsevier.

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. (2013). *Sequences of frustration and confusion, and learning* [Paper presentation]. Educational Data Mining, 2013. https://www.semanticscholar.org/paper/Sequences-of-Frustration-and-Confusion%2C-and-Liu-Pataranutaporn/a899b3caf4bb983e015bfed88531e5bab570c4a6

Magdin, M., & Prikler, F. (2018). Real time facial expression recognition using webcam and SDK affectiva. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(1), 7–15.

Munz, U., Schumm, P., Wiesebrock, A., & Allgower, F. (2007). Motivation and learning progress through educational games. *IEEE Transactions on Industrial Electronics*, *54*(6), 3141–3144.

Orvis, K. A., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior*, *24*(5), 2415–2433.

Papanastasiou, G., Drigas, A., & Skianis, C. (2017). Serious games in preschool and primary education: Benefits and impacts on curriculum course syllabus. *International Journal of Emerging Technologies in Learning (iJET)*, *12*(1), 44–56.

Paquette, L., Baker, R. S., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. In Trausan-Matu S., Boyer K.E., Crosby M., & Panourgia K. (Eds.), *Intelligent Tutoring Systems*. ITS 2014. Lecture Notes in Computer Science (vol. 8474). Springer. https://doi.org/10.1007/978-3-319-07221-0_1

Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a Bayesian networks implementation of knowledge tracing. In De Bra P., Kobsa A., & Chin D. (Eds.), *User Modeling, Adaptation, and Personalization*. UMAP 2010. Lecture Notes in Computer Science (vol. 6075). Springer. https://doi.org/10.1007/978-3-642-13470-8_24

Plass, J. L., Homer, B. D., Pawar, S., Brenner, C., & MacNamara, A. P. (2019). The effect of adaptive difficulty adjustment on the effectiveness of a game to develop executive function skills for learners of different ages. *Cognitive Development*, *49*, 56–67.

Reichenberg, R. (2018). Dynamic Bayesian networks in educational measurement: Reviewing and advancing the state of the field. *Applied Measurement in Education*, *31*(4), 335–350.

Reye, J. (2004). Student modelling based on belief networks. *International Journal of Artificial Intelligence in Education*, *14*(1), 63–96.

Riedel, J. C., & Hauge, J. B. (2011). State of the art of serious games for business and industry. In *2011 17th International conference on concurrent enterprising* (pp. 1–8). IEEE.

Rodrıguez-Cerezo, D., Sarasa-Cabezuelo, A., GóMez-AlbarráN, M., & Sierra, J.-L. (2014). Serious games in tertiary education: A case study concerning the comprehension of basic concepts in computer language implementation courses. *Computers in Human Behavior*, *31*, 558–570.

Sabourin, J., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). When Off-Task is On-Task: The Affective Role of Off-Task Behavior in Narrative-Centered Learning Environments. In Biswas G., Bull S., Kay J., & Mitrovic A. (Eds.), *Artificial Intelligence in Education*. AIED 2011. Lecture Notes in Computer Science (vol. 6738). Springer. https://doi.org/10.1007/978-3-642-21869-9_93

Sampayo-Vargas, S., Cope, C. J., He, Z., & Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers & Education*, *69*, 452–462.

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C.-P., Kuba, R., Liu, Z., Yang, X., & Sun, C. (2020). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*, *37*(1), 127–141.

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

Shute, V. J., & Wang, L. (2015). Measuring problem solving skills in portal 2. In Isaías P., Spector J., Ifenthaler D., & Sampson D. (Eds.), *E-learning systems, environments and approaches* (pp. 11–24). Springer.

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE Review*, *41*(2), 16.

van Oostendorp, H., Van der Spek, E. D., & Linssen, J. (2014). Adapting the complexity level of a serious game to the proficiency of players. *EAI Endorsed Transactions on Game-Based Learning*, *1*(2), e5.

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *Journal of Computer Assisted Learning*, *36*(4), 502–513.

Verma V., Baron T., Bansal A., & Amresh A. (2019). Emerging Practices in Game-Based Assessment. In Ifenthaler D., & Kim Y. (Eds.), Game-Based Assessment Revisited. Advances in Game-Based Learning. Springer. https://doi.org/10.1007/978-3-030-15569-8_16

Verma V., Rheem H., Amresh A., Craig S. D., & Bansal A. (2020). Predicting Real-Time Affective States by Modeling Facial Emotions Captured During Educational Video Game Play. In Marfisi-Schottman I., Bellotti F., Hamon L., & Klemke R. (Eds.), *Games and Learning Alliance*. GALA 2020. Lecture Notes in Computer Science (vol 12517). Springer. https://doi.org/10.1007/978-3-030-63464-3_45

Vos, N., Van Der Meijden, H., & Denessen, E. (2011). Effects of constructing versus playing an educational game on student motivation and deep learning strategy use. *Computers & Education*, *56*(1), 127–137.

Vygotsky, L. (1978). Zone of proximal development: A new approach. In *Mind in Society: The Development of Higher Psychological Processes* (pp. 84–91). https://www.google.com/books/edition/Mind_in_Society/Irq913lEZ1QC?hl=en&gbpv=0

Yang, Y.-T C. (2012). Building virtual cities, inspiring intelligent citizens: Digital games for developing students' problem solving and learning motivation. *Computers & Education*, *59*(2), 365–377.

Zhang, D., Zhao, J. L., Zhou, L., & Nunamaker, J. F. Jr. (2004). Can e-learning replace classroom learning? *Communications of the ACM*, *47*(5), 75–79.

Zhou, L., Li, F., Wu, S., & Zhou, M. (2020). "School's out, but class's on", the largest online education in the world today: Taking China's practical exploration during the covid-19 epidemic prevention and control as an example. *Best Evidence of Chinese Education*, *4*(2), 501–519.

Author Biographies

Vipin Verma completed his PhD in Human Systems Engineering from Arizona State University (2021). His work focused on integrating content-agnostic stealth assessment in educational video games. His research interests include serious games, affective games, dynamic adaptation, Bayesian networks, and learning assessment.

Scotty D. Craig is an Associate Professor of Human Systems Engineering at Arizona State University and Director of the DoD Advanced Distributed Learning Initiative Partnership Laboratory at Arizona State University. He has a dual affiliation with the Ira A. Fulton Schools of Engineering and the Mary Lou Fulton Teachers College. Dr. Craig has expertise within cognitive psychology, usability, and the science of learning (specifically learning technology). He has published widely in the intersecting areas of Psychology, Education, and Technology. He has served on the program committee for AIED, iFEST, and Cognitive Science and serves on the editorial board for Journal of Educational Psychology, Computers & Education (open), International Journal of STEM Education, and the journal Technology, Mind, & Behavior.

Roy Levy is a Professor in the T. Denny Sanford School of Social & Family Dynamics at Arizona State University. His scholarly interests include methodological investigations and applications in psychometrics and statistical modeling, focusing on item response theory, structural equation modeling, Bayesian networks, and Bayesian approaches to inference and modeling.

Ajay Bansal is an Assistant Professor in Ira A. Fulton Schools of Engineering at Arizona State University. His primary research interests include Knowledge Representation and Reasoning for AI, Machine Learning, CS Education and Logic Programming. His work on Coinductive Logic Programming won the Test of Time Award at Intl. Conference on Logic Programming (ICLP) in 2016. He has also worked in the area of Semantic Web services and was involved in the design and development of Universal Service-Semantics Description

Language (USDL) that received the Best Paper award at the European Conference on Web Services (ECOWS 2005). He has received several teaching awards and was awarded the 2019 Teaching Excellence Award by the Fulton Schools of Engineering at ASU, He obtained a PhD in Computer Science from the University of Texas at Dallas in 2007. Prior to joining ASU, he was a visiting assistant professor at Georgetown University in Washington, D.C. He has also worked in the industry for four years as a software engineer at SIEMENS India, Metallect Corp. and Tyler Technologies.

Ashish Amresh is a Research Scientist at the Center for Human, AI and Robot Teaming and the Decision Theater, Dr. Amresh conducts research in the development of Serious Games and previously he led the Computer Gaming curriculum initiatives at Arizona State University, where he founded the Computer Gaming Certificate, Camp Game and UTBC summer programs. He was the founding program chair for the Simulation Science Games and Animation program at Embry Riddle Aeronautical University and his industry pursuits included working for Ronin Entertainment as a Graphics Software Engineer known for Star Wars: Force Commander and Bruce Lee Quest of the Dragon games. He is the author of Unreal Game Development, a popular book used for teaching game development skills for high school students. He has given numerous talks on using games in the class room for enhancing math and science learning. He is currently researching the development of software tools to improve game based learning outcomes.