Chapter 16
Emerging Practices in Game-Based Assessment

Vipin Verma, Tyler Baron, Ajay Bansal, and Ashish Amresh

16.1 Introduction

Serious and educational games have been a subject of research for a long time. They usually have game mechanics, game content, and content assessment all tied together to make a specialized game intended to impart knowledge of the associated content to its players (Van Eck, 2006). While this approach is good for developing games for teaching highly specific topics, it consumes a lot of time and money. Being able to re-use the same mechanics and assessment methods for creating games that teach different contents would lead to a lot of savings in terms of time and money. The Content Agnostic Game Engineering (CAGE) Architecture mitigates the problem by disengaging the content from game mechanics (Baron, 2017). Moreover, the content assessment in games is often quite explicit in the sense that it interrupts the flow of the players and thus hampers the learning process, as it is not integrated into the game flow. Stealth assessment can be beneficial in such cases to keep the player engagement intact while assessing them at the same time (Shute, 2011). Integrating stealth assessment into the CAGE framework in a content-agnostic way will increase its usability while also decreasing the time and cost of developing in-game assessment.

The word "agnostic" has Greek origin which translates to "not known". The word content agnostic in the context of an educational video game emphasizes the fact that the game mechanics are independent of the target content domain of the game. In the following sections, this chapter will dive into the theory of motivation, followed by the definition of game mechanics, content, and assessment. Then the emerging need for content-agnostic assessment will be discussed, and how the motivation can help in effective learning. It will be followed by the approaches to make the assessment unobtrusive and then methods to quantify the learning gains.

V. Verma (✉) · T. Baron · A. Bansal · A. Amresh
Arizona State University, Chandler, AZ, USA
e-mail: vverma9@asu.edu; tjbaron@asu.edu; ajay.bansal@asu.edu; amresh@asu.edu

D. Ifenthaler, Y. J. Kim (eds.), *Game-Based Assessment Revisited*, Advances in Game-Based Learning, https://doi.org/10.1007/978-3-030-15569-8_16

16.2 Self-Determination Theory

Motivation is to be moved to do something and can be categorized as intrinsic or extrinsic (Ryan & Deci, 2000). Intrinsic motivation involves an innate desire to achieve an outcome while extrinsic motivation uses external rewards to drive a person towards the desired outcome. Since people learn better while acting on their natural tendencies, intrinsic motivation can actuate better and higher-quality learning (Ryan, LaGuardia, & Rawsthorne, 2005). Inherent interactivity, challenge, fantasy, and curiosity in the video games help in sustaining the intrinsic motivation of the players during the game-play (Freire et al., 2016; Malone, 1981). Avatar customization in the game *Zombie Apocalypse* is an example of intrinsic motivation (Birk, Mandryk, & Atkins, 2016). Extrinsic motivation such as a grade, on the other hand, can be detrimental to learning.

Self-Determination theory (SDT) specifies the degree to which a person is intrinsically motivated to improve themselves (Chatzisarantis, Biddle, & Meek, 1997). Unfamiliar gaming environment motivates players to master the environment and learn new skills in the process. As shown in Fig. 16.1, it has three components: autonomy, relatedness, and competence. The need for autonomy is related to the sense of control over one's surroundings (Deci & Vansteenkiste, 2004). Video games present autonomy by providing its players with a set of choices and allowing its players to follow their own path towards an objective. Customization of player avatar in *Second Life* (Linden Labs, 2003) and branching narratives in *Dragon Age: Origins* (BioWare, 2009) are some examples of autonomy manipulation within games. The need for relatedness revolves around a person's desire to have a sense of belongingness among their peers, competitors, and instructors. Multiplayer games allow the need for relatedness to be fulfilled by allowing a person to play with others. Multiplayer group (clan) play in *League of Legends* (Riot Games, 2009) and

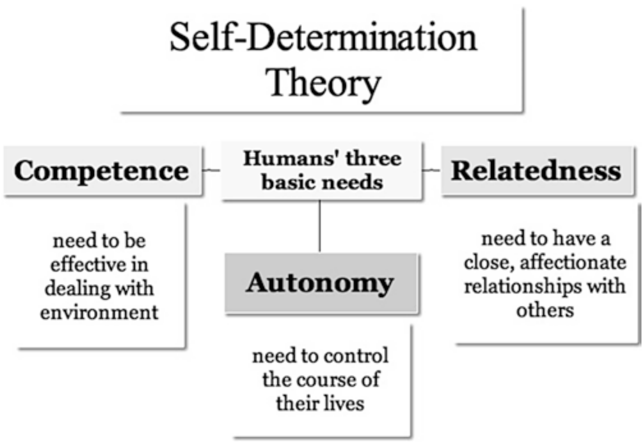

Fig. 16.1 The three components of SDT (Deci & Ryan, 2008)

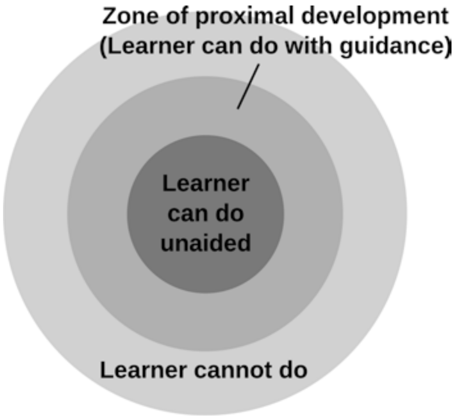

Fig. 16.2 Zone of proximal development (Vygotsky, 1978)

match-making algorithms in online multiplayer games like *Brawl Stars* (Supercell, 2017) are some ways to keep relatedness intact. The need for competence relates to a person's ability to attain learning objectives. Video games can promote competence by providing incremental objectives with an increasing difficulty level. Ryan, Rigby, and Przybylski (2006) used SDT to explain the motivation pull in video games. Their experiments suggested that a video game which is autonomy-friendly, relatedness-invoking, and competence-evoking could help sustain the motivation levels in a video game (Sørebø & Hæhre, 2012). Situations that thwart these three needs undermine the intrinsic motivation of an individual. However, high autonomy makes it difficult to compare the evidence gathered from two different players, and increased relatedness can lead to construct-irrelevant variances, thus thwarting the assessment process. So, a delicate balance is required to keep both engagement and assessment intact, simultaneously.

Tasks that lead a learner to the cusp of their abilities affect their engagement and motivation positively (Gee, 2003) and help them remain in flow (Csikszentmihalyi, 1975). Vygotsky (1978) used the term called the zone of proximal development to describe this edge of abilities. The zone of proximal development is the difference between what the learners can do without any assistance and what they cannot do even if they had help. This zone contains the skills that the learner can attain when guided properly. A learner with high skill level when presented a low-level challenge will get bored, while introducing a difficult task to an unskilled learner will make them anxious or frustrated. Thus, it is advisable to keep the learner in the zone of proximal development by keeping the optimum level of challenge suitable for their current skill level (Fig. 16.2).

16.3 Game Mechanics, Content, and Assessment

Sicart (2008) defined the term game mechanics as the ways in which players interact with the gaming environment. A game mechanic can be understood as a verb, for example, climbing, running, whistling, shooting, grabbing, and switching weapons

(Järvinen, 2008). Mechanics are a means to overcome the challenges encountered during the game-play or any desired outcome that requires an effort (Sicart, 2008). For example, stabbing is a basic mechanic found in the game *Shadow of the Colossus* which involves plunging a weapon into the body of the colossus to injure them (Team Ico, 2005).

The content domain of a game is the topic which the game is trying to teach its players (Baron & Amresh, 2015). For example, consider a game designed to teach encryption methods to its players. The content domain for this game would be Cryptography. Unlike game mechanics, which are important pieces in any video games, content domain is defined only in educational video games. Commercial entertainment games are not meant for teaching purposes; hence they do not need to define a content domain. Defining a content domain is a crucial part in the design of an educational video game because its aim is to impart skills pertaining to that domain. It is thus a common practice to specify a content domain and then design the educational game around it.

Assessment is a process which uses data to determine if the learning goals are met (Chin, Dukes, & Gamson, 2009). Consider the game from the previous example in which the content domain is Cryptography. Then the purpose of the in-game assessment would be to find out if the player has learned how to use basic encryption mechanisms taught by the game such as the Caesar cipher. Assessment is critical to the growth of serious games and the quantification and validation of learning so that their benefits can be justified over other instructional strategies (Ritterfeld, Cody, & Vorderer, 2009). Assessment and learning should happen simultaneously in an educational game so that the players are aware of their current skill and can progress towards the learning outcome accordingly. Setting up the assessment is equally important as defining the content and mechanics for an educational video game. In level-based games, the level progression will be governed by the assessment, as players will be allowed to progress further in the game only if they demonstrate the ability to clear the previous set of challenges. In the absence of an assessment, the level of game progress will not be an indicator of the skill level of the player.

The two most pertinent questions while designing any assessments are: what and why (Plass et al., 2013). That is, what variables need to be measured and why they need to be measured in order to accurately assess student progress. In educational games, learning outcomes are the variables that are measured to gauge the effectiveness of learning employed in the game. Three categories of variables exist during an educational assessment: general trait variables, general state variables, and situation-specific variables. Trait variables (such as executive functions, verbal ability, and spatial abilities) are relatively stable and are usually not targeted in educational video games. Typically, the aim of educational games is to improve the state variables (such as subject-specific knowledge and meta-cognition) while keeping the situational variables at their optimal level for maximum learning to occur. Situational variables (such as emotional state, engagement, and cognitive load) will change as a result of the player's interactions with the gaming environment. Game design affects the situational variables to a greater extent, and thus it is important to follow game design principles that optimize these variables to keep the player in a zone of proximal development.

Confounding results may occur during an assessment procedure due to several reasons (Plass et al., 2013). Motor skills, content irrelevant skills, and emotions are several potential confounding variables. For example, a game that requires its learners to tilt a tablet device in order to guide a ball to the correct answer could lead to an incorrect observation if the learner tilted the device too quickly and guided the ball to the wrong location despite having the required skill to answer it correctly. Similarly, a game which involves chemical equation balancing may be confounded by the need to know about basic algebra. Further, situations that lead to different results when people respond differently under different emotional states could present a potential confound to the assessment process. It is important that these variables be taken care of during the assessment process. It is problematic if a student is answering incorrectly because of these reasons despite having the required level of competency.

16.4 Disconnecting the Mechanics and Assessment from Content

Previously, commercial games have been used for educational purposes (Van Eck, 2006). Using commercial off-the-shelf (COTS) games for learning is cost-effective and thus gaining acceptance owing to its practicality. However, they pose various challenges as commercial games were not designed for learning. Very limited topics can be taught using COTS games, which might be neither complete nor accurate. These games may cover a large range of content, as a breadth approach, or they may focus on a narrow and specialized topic, as a depth approach. Games that take a depth approach to the content may have missing contents, while the games that take the breadth approach may have missing topics within the content. The depth approach focuses on few topics with lots of detail, while the breadth approach focuses on several topics generally. However, the absence of relevant topics and contents causes a state of cognitive disequilibrium which promotes the thinking and learning of its players in order to attain equilibrium (Kibler, 2011). This persistent cycle of cognitive disequilibrium and equilibrium helps the players engage to the game-play and maintain flow (Csikszentmihalyi, 1975). However, the missing content needs to be addressed using either the traditional classroom activities or through the game itself. But the flow will be interrupted if the players are asked to stop the game to be educated on the missing content. Thus, COTS-based games are detrimental to the flow experience of the player (Van Eck, 2006). This suggests that the ideal solution is to link the game content domain with the game mechanics in order to obtain an optimal flow experience. However, linking the two may cause another problem. For example, imagine that you developed an educational video game which is designed to teach chemical equation balancing with an embedded assessment to evaluate the learning progress. Over time, a developer may decide to create a new game to teach basic cryptographic encryptions. The problem that you will find is that if you can use the same game to teach encryption as well, it would be really difficult to teach and assess the learning of encryptions using it. You may need to make many modifications to the game to teach the encryptions which would need a substantial amount of time and effort. As an alternative, you can also develop an entirely new game from scratch, which after a certain point may be easier than trying to modify the original game.

To mitigate this problem, one can design game mechanics which are content-agnostic, i.e. mechanics which are independent of the content being taught by the game. However, this may cause several other problems. The first problem is the same which is encountered when using COTS games for learning, as it can lead to inaccurate and incomplete content (Van Eck, 2006). However, this problem can be reduced if the learning and assessment strategies are taken into consideration during the early stages of game design. Baron (2017) has provided a game development framework called CAGE which helps in creating a content-agnostic game. The second problem that may arise is the issue of generalizability. It may be boring to play multiple games for learning different contents, all of which employ the same game mechanics, as the mechanics will become difficult to enjoy after a while. Further, there exist some specialized skills which require highly specific training that could be very difficult to fit to other content. So, it is difficult to create a single game which can address multiple content domains. However, this should be kept in mind while developing a game and accommodated using the adaptive game design and feedback capabilities to palliate this problem to a considerable extent. Moreover, it will be better over the current state where a specific game is required for each type of content and assessment.

16.5 Stealth Assessment

There are three types of assessments depending on the time when assessment takes place (West & Bleiberg, 2013). They are diagnostic, formative, and summative. Diagnostic assessments occur prior to delivering instruction to measure the prior knowledge of a student. It can be used to design the delivery of information before a student starts learning. Formative assessments monitor the student's understanding during the learning and can be used to plan the subsequent learning strategy according to the changing level of the player. Based on the continuous evaluation of the student, it can be used to provide ongoing feedback, remediate misconceptions, and dynamically adapt the learning as the learning progresses. Its purpose is to improve student learning by keeping them in the zone of proximal development. Summative assessment occurs after the learning process to evaluate overall achievement summary of a student's performance. Summative assessments inform whether the student has attained the required knowledge or not. Summative assessments are usually high stakes and answer questions such as whether the employee should be promoted, should a player be allowed to progress to the next level, or what grade or SAT score should be assigned to a student. Formative assessments provide an opportunity to rectify mistakes without any grave penalties, while summative assessments do not give a chance to correct errors.

Christel Moors, head of a middle school in Atheneum, Bree, dreams of a school devoid of grades (Renard, 2016a). Her school has removed all the exams and is striving towards a system free from grades and tests, which helps reduce the stress and anxiety levels of students. They believe in formative assessments instead of the grades calculated via summative assessments. The school also thinks that self-determination theory is the way to implement it, and they only talk about a student in terms of his/her strength and weakness instead of grades. To achieve autonomy for students, the instructional strategy needs to move from traditional methods to interactive ones with choices (Renard, 2016a). Students should be allowed to be themselves with the learning activities that fit their world. By doing this, students will be more engaged to the learning material, as they own their learning process. The process involves many challenges for students to accomplish their goals, and they are free to decide which pathway to follow at their own pace. A student should feel connected to his/her peers and teachers in order to be able to make mistakes and learn from them, which follows the principle of relatedness. Further, every student should have a positive self-image and feel competent enough to take on new challenges to obtain satisfactory results. This way each student will have their own success story with a boost in self-confidence. A student who is self-driven, connected with peers, and confident will be better motivated to learn (Renard, 2016a).

Bellotti, Kapralos, Lee, Moreno-Ger, and Berta (2013) suggest incorporating the assessment into the game itself, known as stealth assessment which aims to remove the demarcation between learning and assessment (Moreno-Ger, Martinez-Ortiz, Freire, Manero, & Fernandez-Manjon, 2014). Also, Shute and Ventura (2013) proposed learning games as an alternative to traditional learning with a benefit of adjusting the learning to the level of a struggling student with the help of an embedded stealth assessment. They argued that the classroom learning progresses at its own pace with little regard to a single struggling student. However, student interaction with the gaming world can be analysed at run-time or later to quantify the learning gains. Run-time analysis can be used for personalizing the learning of an individual student by augmenting the game with the help of dynamic adaptation and actionable feedback to improve learning. Formative stealth assessment improves the accessibility for the customized learning to happen (Renard, 2016b). It helps in obtaining the current standing of the student and the objective that they are working towards while helping them thrive towards it.

Stealth assessment is based on Evidence-centred design (ECD), which itself consists of five layers where the assessment design decisions take place (Mislevy, Almond, & Lukas, 2003). Information about the content domain of interest is gathered in the first layer, called the *Domain Analysis* layer. Thus, information is then used to build assessment arguments in the second layer, which is the *Domain Modelling* layer. These assessment arguments are converted into the specific tasks in the third layer, called the *Conceptual Assessment Framework* layer. In the fourth layer, which is the *Assessment Implementation* layer, the tasks are presented to the students, and their responses are analysed. *Assessment Delivery* layer is the last one where the assessment is reported. All these layers are guided by the third layer of *Conceptual Assessment Framework*, which consists of three models: a competency model, a task model, and an evidence model. The competency model, also called the student model, composes of variables representing student skills and knowledge that need to be assessed (Mislevy, Behrens, Dicerbo, & Levy, 2012). The task model consists of the situations and scenarios used to elicit the behaviours that can reveal the skills under observation. It usually relates the unobservable skills with the observable missions in games (Shute & Spector, 2008). The evidence model is responsible for updating the competency model on the basis of evidence gathered from the task model and is the bridge between the two models (Conrad, Clarke-Midura, & Klopfer, 2014).

16.5.1 Stealth Assessment Techniques

There are various ways a stealth assessment can be incorporated in a video game. Some of them are mouse-tracking (Rheem, Verma, & Becker, 2017), emotion tracking, log analysis, Bayesian modelling, along with several other Educational Data Mining techniques (Baker et al., 2012). The strength of all these techniques is that they provide rich information without the use of any expensive intrusive equipment, such as eye-tracker, galvanic skin response sensor, EEG, and other biometric instruments.

16.5.1.1 Mouse-Tracking

Educational video games that involve the use of a computer mouse or a touchscreen device can use mouse or touch-tracking as a stealth measure to assess situational specific variables, such as cognitive load (Rheem et al., 2017). Figure 16.3 shows a sample mouse-tracking plot depicting the trajectories for mouse-movement from the start location to the target. The process involves tracking the mouse-coordinates with time, and it is used to make inferences about the state or intent of the player. Mouse-tracking has been used in the past for inferring positive and negative emotions (Yamauchi & Xiao, 2018), memory strength (Papesh & Goldinger, 2012), gender stereotypes (Freeman & Ambady, 2009), numerical representation (Faulkenberry, 2016), perceptual decision making (Lepora & Pezzulo, 2015), and cognitive load (Rheem, Verma, & Becker, 2018). The inferences can then be used to alter the game-play to suit the player. For example, if it is observed that the player is experiencing a high cognitive load, then relevant steps should be taken to reduce the extraneous load by adapting the game in a suitable manner. While mouse-tracking is beneficial, collecting mouse-tracking data is a resource-intensive process and may demand extensive computer memory depending on the required temporal

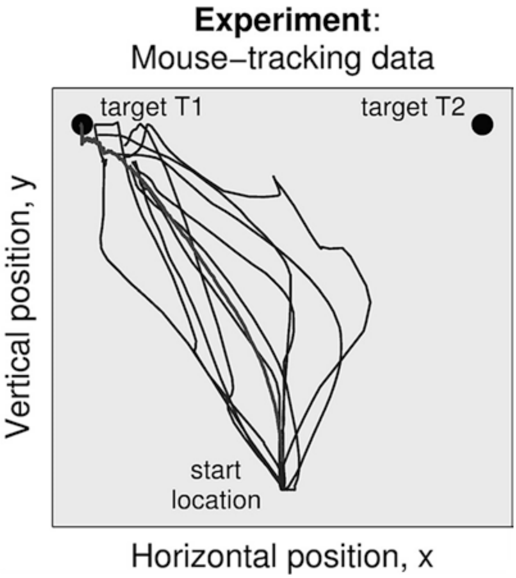

Fig. 16.3 Sample plot showing mouse-trajectory data (Lepora & Pezzulo, 2015)

resolution. For example, tracking mouse coordinates every 200 ms is less expensive compared to collecting it every 50 ms.

16.5.1.2 Emotion-Tracking

Emotion tracking involves tracking the mood of the player during the game-play so that it can be used to adjust the game for an optimal experience. A person might get bored if the game difficulty is too low, or they may get frustrated if it is too high. Thus, the game difficulty should be kept at such a level that keeps them in a state of flow (Csikszentmihalyi, 1975). The process requires facial tracking to detect the mood of the player. There are various methods available for the affect detection using facial tracking that use the Facial Action Coding System (Ekman & Friesen, 1978). Visage|SDK (Visage) from Visage Technologies and Affdex (Affectiva) from Affectiva are two software development kits which can be embedded in a video game for affect detection. Figure 16.4 shows the facial tracking snapshot, with the action units highlighted using white dots.

16.5.1.3 Log Analysis

Player data such as the number of lives remaining, number of player deaths, player level, time spent on a level and during a task, hint usage, quiz responses, score, and anything else that can be assigned to an observable variable can be collected and stored in a log file. A sample log file shown in Fig. 16.5 can then be analysed later

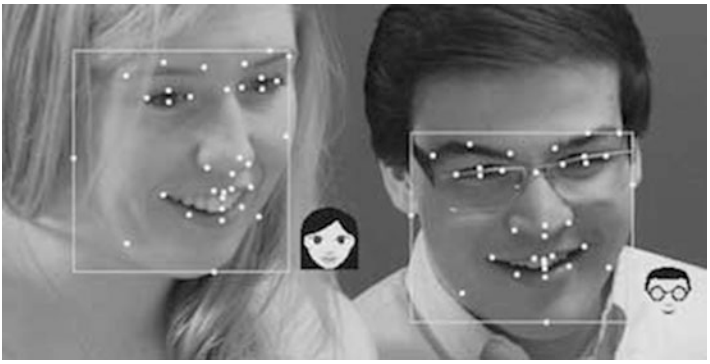

Fig. 16.4 Snapshot of emotion tracking using Affectiva (Metrics, 2019)

for a summative assessment or used for a runtime formative analysis. Wang, Shute, and Moore (2015) has incorporated the best practices to be used for a logging system. In short, they suggested to keep the log files customizable, manageable, well organized, usable, and include only the relevant data in it.

16.5.1.4 Bayesian Modelling

Bayesian modelling is a probabilistic approach to model the conditional dependence of a variable on several other variables (Friedman, Geiger, & Goldszmidt, 1997). García, Amandi, Schiaffino, and Campo (2007) used a Bayesian network to predict the learning styles of students in a web-based learning system. Figure 16.6 depicts a simple Bayesian network called knowledge tracing for a two-quiz sequence that incorporates the four performance parameters called prior knowledge P(L), guess rate P(G), slip rate P(S), and learn rate P(T) (Corbett & Anderson, 1994). Prior knowledge is obtained using diagnostic assessment and probabilistically influences all the other parameters. Guess rate is to account for the correct answers despite not having the knowledge required to do so, while slip rate is for the incorrect response by a skilled student. Learn rate is the probability that the learning will occur in the second quiz based on the response of the previous quiz. Bayesian networks can be used to model complex student models and will be discussed in more detail in the following sections.

16.5.1.5 Educational Data Mining

Educational Data Mining (EDM) consists of methods which are used to discover patterns in high volumes of educational data gathered during the student game-play interactions (Scheuer & McLaren, 2012). As a non-stealth measure, EDM has been used by D'Mello and Graesser (2010) to predict the affective states of students

```
3.483282 timeSpentInMenu 3.483282
4.441414 FoundTheFace
13.10056 CorrectAnswersDuringPreTest [1,2,5]
13.10056 timeSpentInPreTest 8.844877
219.8173 timeSpentInReading 206.7167
220.789 FoundTheFace
229.12 StateDetected Boredom
269.58 CollectedEverything
277.4767 timeSinceLastDeath 87.34433
296.36 P(skill(t=0)|evidence 0.321
309.1737 selfReportedFrustration
400.3962 timeOnDiffLevel1 235.5336
400.3962 SettingTheDifficultyLevel Two
422.8518 cumulativeTimeSpentOnLevels 260.8243
429.7115 timeOnUESSurvey 5.928741
434.5058 timeOnDemoSurvey 4.794312
447.4459 CorrectAnswersDuringPostTest [1,4]
447.4459 timeSpentOnPostTest 12.94009
447.4459 totalTimeSinceStart 508.0759
447.4459 totalHelpCount 0
447.4459 totalCoinsPickedUp 69
447.4459 totalLivesPickedUp 1
447.4459 totalDistractorsPickedUp 3
447.4459 enemyCollisionCount 11
447.4459 hazardCollisionCount 0
447.4459 timesDied 4
447.4459 skillLevel 0.9858083
```

Fig. 16.5 A sample log file

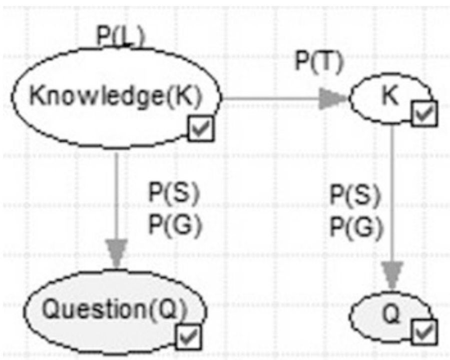

Fig. 16.6 Bayesian knowledge tracing (Pardos & Heffernan, 2010)

while they were sitting on a chair. They investigated the affective states and posture patterns of 28 students while they were learning with the help of an interactive tutoring system. Application of binary logistic regression associated the leaning back on a seat with boredom and disengagement and leaning forwards to frustration or delight depending on the angle of inclination while leaning forward. As a stealth measure, EDM was used by Baker and colleagues (2012) to predict the affective states of players using interaction logs and obtained a better than chance performance. EDM has also been used in the past to measure the degree of agency with which a student exerts control over their choice patterns (Snow, Jacovina, Varner, Dai, & McNamara, 2014). There is a wide array of EDM methods available such as, clustering, classification, regression, support vector machine, and reinforcement learning. Hence a great deal of care should be taken to pick the right one. Further, all the assumptions (if any) should be kept in mind while using that method.

16.5.2 Student Model

There are various aspects of a student that may need modelling while they are interacting with an educational video game. It can comprise trait variables, state variables, situation-specific variables, or any combination of them. The student model is a representation of the corresponding student assessment variable(s) at any point in time during the assessment. The student model can be potentially used to personalize the student learning to keep them in the zone of proximal development and provide necessary remediation if required.

Figure 16.7 above shows an example of a student model for an educational video game which uses the Dynamic Bayesian Network of knowledge tracing adapted from Pardos and Heffernan (2010). It is similar to the network in Fig. 16.6, except it is more complex and dynamic. The network shown in Fig. 16.6 consists of two nodes: a student node (S), a knowledge node (K), and a question node (Q). The prior knowledge parameter P(L) depicts the initial skill level of a student. The knowledge node corresponds to the state of the student knowledge, i.e. whether the skill has been attained or not. While the question node depicts whether they answered the quiz correctly or incorrectly. Figure 16.6 contains more nodes such as student node (S) and Distractor nodes (D). The student node represents an individual student. The arc below the Knowledge node depicts the conditional dependence of Knowledge at time step $t + 1$ on the Knowledge at previous time step t. This is shown clearly in the unrolled Dynamic Bayesian Network in Fig. 16.8.

Consider a game which is designed to teach encryption methods to its players using the basic Caeser cipher. The aim of any level in the game is to encrypt a plain text using a given key. To achieve this, the player is tasked with collecting the letters which appear in the resultant cipher-text when plain text is encrypted using the given key. For example, in Fig. 16.9, the resultant cipher-text for the given plain text “ATTACK AT DAWN” using the encryption key 2 will be “CVVCEM CV FCYP”. So the task of the player is to collect the letters ‘C’,‘V’,‘V’,‘C’,‘E’,‘M’,‘C

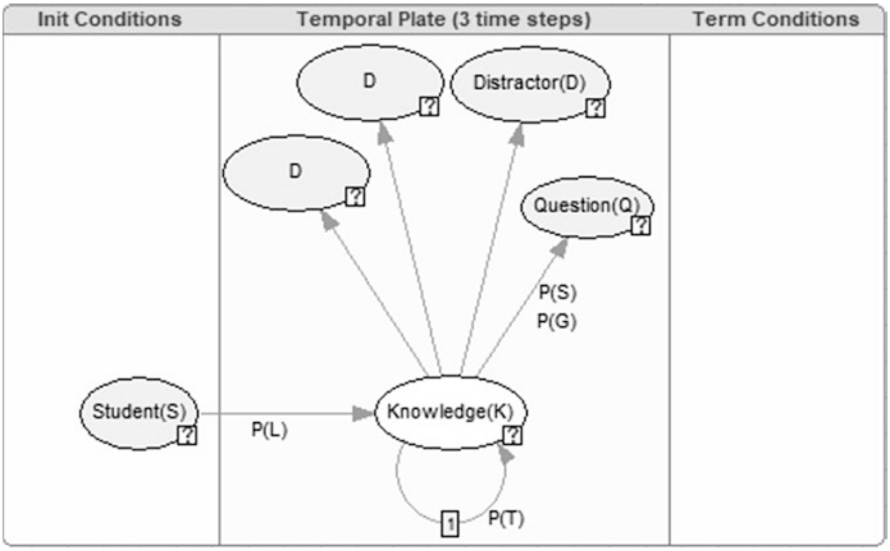

Fig. 16.7 An example of a dynamic Bayesian network

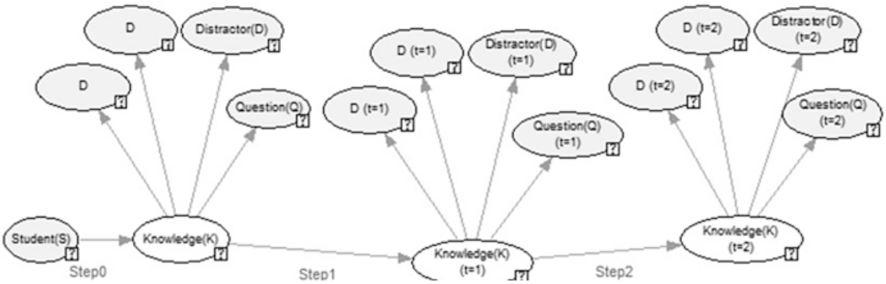

Fig. 16.8 Unrolled version of dynamic Bayesian network from Fig. 16.7

','V','F','C','Y','P' which are scattered throughout the level. The student node, in this case, would represent an individual player and their initial knowledge about encoding text using the Caeser cipher. Knowledge node would correspond to the state of their encoding skill, and question node would represent whether they achieved the task successfully or not. In addition to these three, Figs. 16.7 and 16.8 have several other nodes called distractor (D) which represents various distractors laid out around the level to check student skill and potential guessing. In the example game shown above, a distractor could be a letter which does not appear in the resultant cipher-text and therefore not supposed to be collected. Figure 16.4 displays a distractor letter 'H' which does not appear in the resultant cipher-text "CVVCEM CV FCYP". Collecting these distractors while not having the required skill could suggest guessing. All the performance parameters which represent the conditional probabilities at various nodes can be used for Bayesian inference while

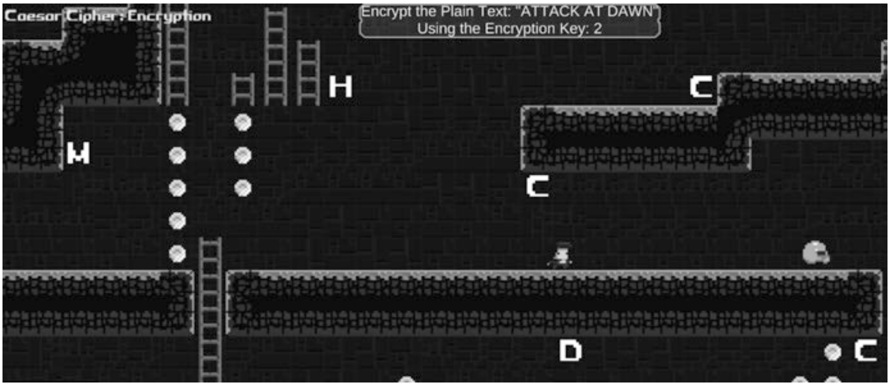

Fig. 16.9 An example of a game for student modelling

the game-play is in progress. The inference can be used to gauge the current skill level of student given various pieces of evidence. This, in turn, can serve as a formative assessment of the student skill and can be used for personalizing the learning of an individual by taking appropriate measures in accordance with the student model.

16.6 Content Agnostic Game Engineering

Educational video games have been shown to be effective for learning, but the learning gains are not generalizable (Cheng, Rosenheck, Lin, & Klopfer, 2017; Fletcher & Tobias, 2011; Freeman & Higgins, 2016). The results are often limited to the games used for research, and they are not content-agnostic. CAGE is an architecture for designing educational video games and assessment in which the game mechanics are independent of the game content while keeping the educational value of the game intact (Baron, 2017). It follows a game design approach and helps keep the players engaged to the game-play and learning. Being content-agnostic, it facilitates making the subsequent versions of the game and thus accelerating the development process. Only the first game will require the full-scale expenses; the following games will need some minor changes to accommodate the new content leading to reduced time and cost requirements.

CAGE has been proven to be effective in reducing the time spent while developing subsequent versions of the same game for different content (Baron, 2017). Baron (2017) did a study based on 11 students from a game-based learning class in Arizona State University. Participants were asked to make two games using the CAGE framework. On an average, they reported writing 70% lesser code and spending about 55% less time in developing the second game when using the CAGE framework. The results also indicated that the participants perceived the CAGE framework to be helpful in speeding up the game development process. However, it led to a decrease in cognitive load and engagement for players, when playing the second content right after the first one. For the first version of the game, the mechanics are new to the player and need to be learned. However, for the second version, the mechanics are the same and thus not required to be learned, hence the expected decrease in cognitive load and engagement.

The CAGE model depicted in Fig. 16.10 essentially consists of a one-way loop which begins with the player input to the game (Baron, 2017). The input is passed from the system hardware to the mechanics component which converts them into in-game action. The actions are then analysed by the content component, evaluating the action and passing the evaluation to student model which then accumulates the evaluation and passes the feedback to the player. Player then incorporates the feedback in their subsequent action.

CAGE architecture is component based and consists of the mechanics component, the content component, the student model, and the framework which binds them all together.

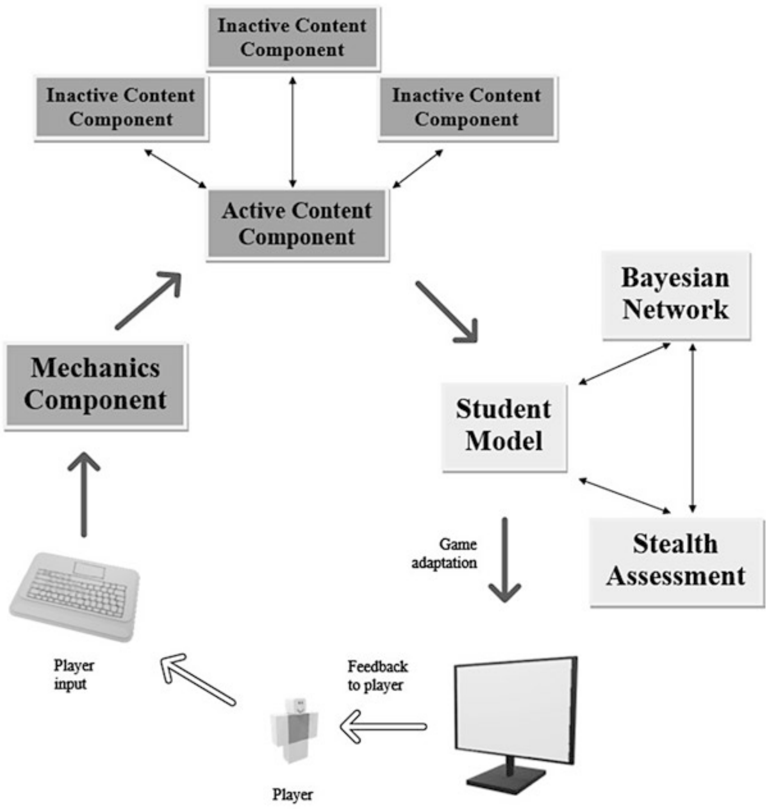

Fig. 16.10 The CAGE model (Baron, 2017)

16.7 Cage Architecture

The architecture framework is built in Unity game engine built by Unity Technologies (Baron, 2017). The framework utilizes generic messages called Hooks which are activated during the game-play events invoked by the player input. These Hooks are passed to the content component and processed if they are relevant to the content domain being played; otherwise they are ignored. This allows the mechanics component to send out the hooks to content component without knowing which content is active at present. The content component selectively implements the relevant Hooks. If an unknown Hook is received by the content component, it is ignored, and the player action is marked as invalid by the content component for that Hook.

16.7.1 Framework

The Framework is the skeleton that keeps all the components tied together (Baron, 2017). It connects the external input of the player to the game mechanics. The evaluation of the input is passed to the content component, and then to student model, which returns the feedback to the player via the framework part of the architecture. The player then incorporates the feedback into their next action, and the cycle is repeated. The Framework is static and consistent across all the version of the game developed using the architecture.

16.7.2 Mechanics Component

This component processes the input received from the player and converts it into in-game action. In CAGE architecture, this component is designed to be content-agnostic (Baron, 2017). Usually, game mechanics and content domain are either deeply connected as in traditional games, or poorly connected when using COTS games (Van Eck, 2006). However, in CAGE architecture they will be independent of each other and thus facilitate the mechanics to be content-agnostic.

16.7.3 Content Component

In CAGE this component is designed to be dynamic and easily replaceable with another content, being independent of the game mechanics (Baron, 2017). It evaluates the player action for their knowledge and skill level in that domain and passes the evaluation results to the student model to update the state of the student model.

16.7.4 Student Model

Student model represents the knowledge state of a given player at any point in time. It processes and accumulates the results from the content component. It is also used to dynamically provide appropriate feedback and remediation to the players, to aid their learning process. The student model has three-fold benefits associated with it. Firstly, it provides a dynamic assessment of the student knowledge state. Secondly, it can be used for dynamic feedback, remediation, and as a deterrent to behaviours that are not favourable to learning. Thirdly, it provides dynamic game adaptation capabilities to adjust the game or content difficulty on the basis of the skill level of the players and thus keep them in the zone of proximal development.

16.8 Conclusions

The growing volume of literature on game-based assessment suggest a bright future ahead. Games are intrinsically motivating and have the potential to promote sustained learning during the game-play session. The learning can be scaffolded into the gaming environment such that the mastery of learning is attained during the process of mastering the game environment. As opposed to traditional forms of assessment which allows measurement of state variables only, game-based assessment enables quantification of trait variables, state variables as well as situation specific-variables. It enables measuring skills such as persistence and systems thinking that are hard to measure using pen-and-paper tests while keeping the test anxiety at bay. It can be used for all sorts of assessment, diagnostic, formative, as well as summative. There is a wide range of assessment techniques available at our disposal. Emerging practices for game-based assessment involve tackling multiple content assessments using a single game without making the assessment obvious to the learner while building and adapting the learning strategy as the learner progresses through the game. Dynamically personalizing the game in accordance with the skill level of a player not only helps in keeping the player in flow but also helps in improving their learning.

References

Baron, T. (2017). *An Architecture for Designing Content Agnostic Game Mechanics for Educational Burst Games* (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). Academic Conferences International Limited.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction, 2013*, 1.

Birk, M. V., Mandryk, R. L., & Atkins, C. (2016, October). The motivational push of games: The interplay of intrinsic motivation and external rewards in games for training. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 291–303). New York, NY: ACM.

Brawl Stars. (2017). *Supercell*.

Chatzisarantis, N. L., Biddle, S. J., & Meek, G. A. (1997). A self-determination theory approach to the study of intentions and the intention–behaviour relationship in children's physical activity. *British Journal of Health Psychology, 2*(4), 343–360.

Cheng, M.-T., Rosenheck, L., Lin, C.-Y., & Klopfer, E. (2017). Analyzing gameplay data to inform feedback loops in the radix endeavor. *Computers & Education, 111*, 60–73.

Chin, J., Dukes, R., & Gamson, W. (2009). Assessment in simulation and gaming: A review of the last 40 years. *Simulation & Gaming, 40*(4), 553–568.

Conrad, S., Clarke-Midura, J., & Klopfer, E. (2014). A framework for structuring learning assessment in a massively multiplayer online educational game: Experiment centered design. *International Journal of Game-Based Learning (IJGBL), 4*(1), 37–59.

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

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of Humanistic Psychology, 75*, 41.

Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., ... & Rossi, L. (2012). Towards sensor-free affect detection in cognitive tutor algebra. *International Educational Data Mining Society*.

D'Mello, S., & Graesser, A. (2010, June). Mining bodily patterns of affective experience during learning. In *Educational data mining 2010*.

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie Canadienne, 49*(3), 182.

Deci, E. L., & Vansteenkiste, M. (2004). Self-determination theory and basic need satisfaction: Understanding human development in positive psychology. *Ricerche di psicologia*.

Dragon Age: Origins. (2009). *BioWare*.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Faulkenberry, T. J. (2016). Testing a direct mapping versus competition account of response dynamics in number comparison. *Journal of Cognitive Psychology, 28*(7), 825–842.

Fletcher, J., & Tobias, S. (2011). *Computer games and instruction*. Charlotte, NC: Information Age Publishing.

Freeman, B., & Higgins, K. (2016). A randomised controlled trial of a digital learning game in the context of a design-based research project. *International Journal of Technology Enhanced Learning, 8*(3–4), 297–317.

Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science, 20*(10), 1183–1188.

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

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

García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education, 49*(3), 794–808.

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

Järvinen, A. (2008). *Games without frontiers: Theories and methods for game studies and design*. Tampere, Finland: Tampere University Press.

Kibler, J. (2011). Cognitive disequilibrium. *Encyclopedia of child behavior and development*, 380–380.

League of Legends. (2009). *Riot Games*.
Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: How action influences perceptual decision making. *PLoS Computational Biology, 11*(4), e1004110.
Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*(4), 333–369.
Metrics. (2019). Retrieved from https://developer.affectiva.com/metrics/
Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series, 2003*(1), i–29.
Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining, 4*(1), 11–48.
Moreno-Ger, P., Martinez-Ortiz, I., Freire, M., Manero, B., & Fernandez-Manjon, B. (2014). Serious games: A journey from research to application. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings* (pp. 1–4). IEEE.
Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review, 19*(5), 906–913.
Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a Bayesian networks implementation of knowledge tracing. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 255–266). Berlin, Germany: Springer.
Plass, J. L., Homer, B. D., Kinzer, C. K., Chang, Y. K., Frye, J., Kaczetow, W., … Perlin, K. (2013). Metrics in simulations and games for learning. In *Game analytics* (pp. 697–729). London, UK: Springer.
Renard, L. (2016a). *The ABC of a school without grades. An interview with Christel moors*. Retrieved from https://www.bookwidgets.com/blog/2016/11/the-abc-of-a-school-without-grades-an-interview-with-christel-moors
Renard, L. (2016b). *What is formative assessment? The 6 building blocks.* Retrieved from https://www.bookwidgets.com/blog/2016/11/what-is-formative-assessment-the-6-building-blocks
Rheem, H., Verma, V., & Becker, V. (2017). *Give and take in a moustracking choice paradigm*. Poster presented at Proceedings of the 58th Annual Meeting of Psychonomic Society, Vancouver, BC.
Rheem, H., Verma, V., & Becker, V. (2018). Use of mouse-tracking method to measure cognitive load. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, pp. 1982–1986). Los Angeles, CA: SAGE Publications.
Ritterfeld, U., Cody, M., & Vorderer, P. (2009). *Serious games: Mechanisms and effects*. New York, NY: Routledge.
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.
Ryan, R. M., LaGuardia, J. G., & Rawsthorne, L. J. (2005). Self-complexity and the authenticity of self-aspects: Effects on well being and resilience to stressful events. *North American Journal of Psychology, 7*(3), 431.
Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 344–360.
Scheuer, O., & McLaren, B. M. (2012). Educational data mining. In *Encyclopedia of the sciences of learning* (pp. 1075–1079). Berlin, Germany: Springer.
Second Life. (2003). *Linden Labs*.
Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer Games and Instruction, 55*(2), 503–524.
Shute, V. J., & Spector, J. M. (2008). Scorm 2.0 white paper: Stealth assessment in virtual worlds. *Unpublished manuscript*. Tallahassee, FL: Florida State University
Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. Cambridge, MA: MIT Press.
Sicart, M. (2008). Defining game mechanics. *Game Studies, 8*(2). Retrieved from: http://gamestudies.org/0802/articles/sicart.

Snow, E., Jacovina, M., Varner, L., Dai, J., & McNamara, D. (2014, July). Entropy: A stealth measure of agency in learning environments. In *Educational Data Mining 2014*.

Sørebø, Ø., & Hæhre, R. (2012). Investigating students' perceived discipline relevance subsequent to playing educational computer games: A personal interest and self-determination theory approach. *Scandinavian Journal of Educational Research, 56*(4), 345–362.

Team Ico. (2005). *Shadow of the Colossus* (PlayStation 2).

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

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children, 23*(3), 34–41.

Wang, L., Shute, V. J., & Moore, G. (2015). *Best practices and lessons learned of stealth assessment*. Retrieved December, 9, 2015.

West, D. M., & Bleiberg, J. (2013). Education technology success stories. *Issues in Governance Studies.*

Yamauchi, T., & Xiao, K. (2018). Reading emotion from mouse cursor motions: Affective computing approach. *Cognitive Science, 42*(3), 771–819.