Content Agnostic Game Engineering: Impact of Stealth assessment and content order on player engagement

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

Arizona State University, Mesa AZ 85212, USA,
vverma9@asu.edu

Abstract. Content Agnostic Game Engineering (CAGE) architecture utilizes content agnostic mechanics to create educational games that can teach multiple contents. However, the player engagement goes down when second content is played using the same game mechanics. A content agnostic stealth assessment can aid a CAGE game in sustaining the engagement level of its players. A potentially generalizable method for this was tested using Chem-o-crypt, a CAGE game that can teach chemistry and cryptography contents. The game automatically detects frustration, flow, and boredom using the Affdex SDK from Affectiva. A randomized controlled experiment incorporating real-time game adaptation revealed that using stealth assessment can help sustain engagement in a CAGE game when playing multiple contents.

Keywords: content agnostic, stealth assessment, affective game, adaptive serious game

1 Introduction

Development of an educational game and assessment takes plenty of time, and once the development is complete, the developers may have to start over to create another game [1]. Baron [2] designed a content agnostic architecture called Content Agnostic Game Engineering (CAGE) for creating multiple educational games that rely on the same game mechanics, leading to lower time and cost requirements for building several games at once. However, the architecture did not implement a content agnostic student model of assessment built into it, and the study employed survey questionnaires to assess the engagement, which Baron [2] noted are interruptive in nature and leads to a reduction in the motivation level of players.

1.1 Game Mechanics and Content Domain

A game mechanic is a control mechanism, a rule of game play used by a player for interactions within the game world to achieve the goals of the game [3]. In Angry Birds, the player can fire a bird into the sky by dragging them off a catapult using touch and drag on screen, and then release to launch, a mechanism called slingshotting. The content domain of a game is the subject knowledge that the game is intended to impart [4]. Consider a game designed to teach chemical equation balancing skills to its players. Chemistry would be the content domain for such a game. While game mechanics are important for any video game, the content domain is considered only with regards to an educational game. Commercial games do not usually define a content domain as they are not trying to teach anything specific. Educational games, however, need to define a content domain to make sure that the game is designed to impart skills in that domain.

1.2 Game-Based Assessment

Chin et al. [5] described the assessment as the procedure used to decide if the learning goals are met or not, with the help of data. Consider a game designed to teach cryptographic encryptions to its player. Then the role of the assessment would be to identify if the player has gained the knowledge of how to use the encryption methods like Caesar cipher.

Assessment of the students' knowledge is as important as setting up the content and mechanics of the game. In a level-based game, the student will be allowed to progress to the next level only if they demonstrate through their game play that they have learned the knowledge required to progress to the following stage. If there is no assessment, then the level of progress will not be an indicator of the skill level or knowledge gained by the player, and they will be stuck on the current level forever and get frustrated.

Plass et al. [6] have identified three variables of interest during an educational assessment: general trait variables, general state variables, and situation-specific variables. Trait variables such as executive functions and spatial abilities of players are more or less stable but are not typically targeted in educational video games, although they can be impacted by game play. State variables such as knowledge in an area are the ones that are targeted in serious games. Engagement, cognitive load, affective state, are the situational variables and are there because of the player's interaction within the gaming environment. A typical game would thus be governed by a player's trait variables and should be designed to level up their state variables while keeping their situation variables in an optimum range for best results.

1.3 Game Mechanics and Assessment dependent on Content

Previously, commercial games have been adapted for educational purposes, but they pose several challenges [7]. While, some of the problems arise because of the inability of the educators to make required modifications to the game [8], many issues occur because the content being taught is not tied to the mechanics of the game. This suggests the ideal solution is to link the mechanics with the game content [7]. However, it causes other problems as explained below.

However, linking mechanics and game content could cause other problems. Consider an educational video game that is designed to teach cipher-text to its players. A development studio makes a successful game that teaches cipher-text and embeds an assessment into it to evaluate the learning as the game progresses. Over time, as user needs change, the studio may decide to make a new game for teaching chemistry. The problem that they will come across is that how can the example game which is used to teach cipher-text can also be used to teach chemical equation balancing while having a valid assessment at the same time?

It would be rather difficult to efficaciously teach chemical equation balancing using the mechanics of the cipher-text game. It would be equally difficult to assess the learning of chemical equation balancing with the assessment that was developed for a cipher-text game. Developers may need to make a lot of adjustments to the game mechanics and assessment, spend significant time in coding the game or start an entirely new project from scratch.

1.4 Disconnecting the Three

As mentioned previously, the mechanics and assessment are not transferable across various content domains if they are heavily tied to them. But if they are transferable then it may pose two problems. The first one is that it can lead to inaccuracy in the content and assessment and thus pose difficulty using it as a good educational tool, the same problem which is encountered when using commercial games for educational purposes [7]. However, CAGE architecture can be used to palliate this, as the game design will incorporate learning and assessment strategies from the inception of the game [9].

The second problem is the over-generalization that this may cause. Mechanics that are omnipresent are hard to enjoy and could be detrimental to further learning [2]. It would become boring to play many games all of which employ the same game mechanics while teaching different contents. Thus, it can have serious ramifications for learning, as repetitiveness predicts boredom [10], and boredom is linked to decreased learning in interactive learning environments [10]. There exist many specialized skills, for example, operating a nuclear power plant, that requires focused training. It will be extremely hard to build a universal set of mechanics and assessment which can be used to teach and assess any type of content. However, keeping this in mind from the beginning while developing a game and trying to accommodate it using stealth assessment for dynamic game adaptation and feedback will help alleviate this problem to a considerable extent. Further, mechanics and assessment that can work across several domains would be better over the current state where a dedicated game is required for each type of content and assessment.

1.5 Stealth Assessment

Stealth assessment is an unobtrusive Evidence-Centered Design (ECD) [11] based assessment technique embedded deeply within the game, utilizing the enormous data generated during the game play for inferring player performance at several grain sizes [12, 13, 14, 15]. It has been used for the assessment of creativity [16], persistence [13], physics knowledge [17], problem-solving skills [18], systems thinking [19], causal reasoning [20], team performance [21], and cognitive load [22] in the past. ECD supports embedding the assessments holistically into the game play, with the main disadvantage being the associated high cost for enforcing a full-scale model [23]. A more befitting approach is to accommodate just the design framework of ECD instead of carrying out the full implementation, with key elements being the student, task, and evidence models.

1.6 The CAGE model

Usually, the game mechanics are tightly tied to the educational content being taught by the game, which renders the programming code of the game unusable for further development [2]. Therefore, it requires a major overhaul of the game program for future projects, and often the code is discarded as starting over is more cost and time efficient. CAGE is a model for designing educational games which alleviates this problem by separating the game mechanics from the educational content of the game. This is beneficial for both industry and academia as it will help in the rapid creation of educational games and savings in terms of time and money. Only the first game project will require full-scale expenses, all the subsequent games can be rapidly developed by re-using the code of the first game.

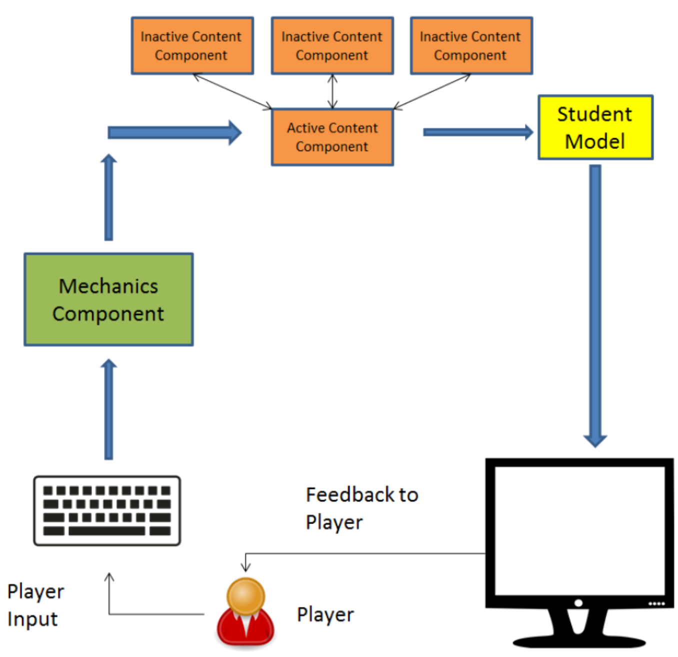

Fig. 1. CAGE Model for educational game development adopted from Baron [2].

The CAGE model, depicted in Figure 1, is composed of the following four components: the framework, the mechanics component, the content component, and the student model. The framework is a static part responsible for gluing the components together [2]. It connects player input with game mechanics, which is linked to the content component. Evaluation from these components updates the student model, which passes the feedback to the player through the framework. The mechanics component upon receiving the input from the player interprets it into a corresponding action in the game. In CAGE, this component is designed to be content agnostic, independent of the content being taught by the game, giving CAGE its name. The content component evaluates the action to update the student model and pass the corresponding feedback to the player. This component is dynamic and can be switched for teaching different content using the same game mechanics. The student model is the one that corresponds to the state of knowledge of a student at any point in time during the game play. It should be pliable enough to be able to assess any domain.

In a study conducted by Baron [2] regarding the effectiveness of CAGE using the eleven games created using the CAGE framework, participants were required to play two versions of the game chosen randomly. They were surveyed using a questionnaire at the end of each game to assess their cognitive load and engagement after the completion of the game. Results indicated a decrease in cognitive load from the first game to the second one, irrespective of the order in which the two versions were played, which is desirable. This can be attributed to the same mechanics being used for both games, eliminating the need to orient a student multiple times towards the game play. Further, it was found that engagement levels decreased from the first game to the second one, regardless of their play order, owing to the same game mechanics being employed in the second version of the game [2]. This is something that needs to be worked upon if the CAGE games are to be employed in a regular classroom. To deal with this issue, stealth assessment should be embedded into the game play to maintain the learner's flow and sustain their engagement in the learning process.

2 Current Study

The current study uses facial emotion tracking with the help of Affdex Software Development Kit (SDK) from Affectiva. An easily accessible webcam is utilized to seize the facial features using the SDK which is then used to observe the emotions. Provided that the lighting is set up correctly and use is front-facing the camera, the SDK yields a high detection rate for facial features [24]. By providing the potential to adapt dynamically, emotion detection can positively affect the fast-paced immersive environments.

The current study was devised to measure the effect of dynamic adaptation on the engagement of players when they play CAGE games. The dynamic adaptation was conducted using stealth assessment, which included facial emotion tracking [15]. It was anticipated that this real-time adaptation would help sustain the engagement of players when they play multiple contents in a CAGE game.

3 Methods

3.1 Design

The experiment implemented a randomized 2x2 factorial design with order (chemistry first or second) and adaptivity (On or Off) as factors to determine the impact on participants' engagement with the gaming system. The Order factor consisted of two levels which determined the order in which the contents were played. An order of chemistry second meant that the player played the cryptography content first, followed by the chemistry content. While chemistry first meant chemistry was played first. The adaptivity factor had two levels as well (on or off), which were used to denote if the stealth assessment was used to adapt the game. Adaptivity being on would mean that the game was adapted using affect and player interactions. Participants' perceptions were measured using the revised User Engagement Scale [25].

3.2 Chem-o-crypt

Content Agnostic Game Engineering (CAGE) architecture [2, 9] was used to create a 2D platformer game called "Chemo-o-crypt" in Unity3D (v2018.1.9f2). In Chemo-o-crypt, the game mechanics allowed left and right player movement, ladder climbing, and jumping. There were three different types of patrolling enemies which reduced a partial portion of the player's health on collision. There were also two types of environmental hazards, which were spikes and water. It would reduce the available player health to zero when they fell into these hazards. Also, these penalties were determined based on the game difficulty that ranged from one to four. For example, full life was reduced if a player collided with an enemy when the game difficulty was set at five, but only 25% of health was reduced if the difficulty level was set at one. The game could be played either for chemistry or cryptography content learning. Each content had four levels, which were distinct from the game difficulty levels. Later levels featured moving platforms, which were either moving by default or started moving when a player jumped onto them. The moving direction could be horizontal or vertical. There were coins and heart-shaped items (1-up) scattered throughout the game map. A player initially had three lives which could be increased by collecting one hundred coins or a 1-up.

Each game level in Chemo-o-crypt was divided into 4 navigable chunks that lied next to each other in a sequence. Governed by the game difficulty, every chunk held a game scene in it. Consequently, each chunk could have four possible scenes that it could be populated with. Therefore, there were 4x4, i.e. 16 maximum possible layouts for the game level environment at any point in time which was dependant on the game difficulty. A player could easily move between the chunks as they were continuous, but only if the player avatar was on the ground level. During the first content level, players spawned in the first chunk, they spawned in the second chunk for the second content level, and so on.

For the adaptive version of the game, the layout of a given chunk only changed when the player crossed a chunk boundary that was not adjacent to that chunk. For example, if the player was moving from chunk 2 to chunk 3, then the environment layout for chunk 1 and chunk 4 may change but not for chunk 2 and chunk 3. Similarly, when they were moving from chunk 3 to chunk 4, the layout may change for chunk 1 and chunk 2 since they were not located next to this boundary. This was done to avoid the distortion of the gaming world in front of the learner's eyes. However, this layout change depended on the game difficulty only. If the adaptive algorithm determined that the game difficulty should increase when a player moved from chunk 1 to chunk 2, then the layout for chunk 3 and chunk 4 will change corresponding to that difficulty level.

Fig. 2. Screen capture of the goal of the game Chem-o-crypt.

Game Content Chem-o-crypt implemented two learning contents using the content agnostic mechanics: Chemistry and Cryptography. For the chemistry version of the game, players were required to collect the correct number of elements and molecules that take part in the chemical reaction to balance it. Consider the chemical equation represented in Figure 2, it required 3 Oxygen (O_2) and 2 Ozone (O_3) molecules to balance this equation. However, there were be 3 distractors present in the game environment, which were the excess of these molecules. For example, for the equation shown in Figure 2, more quantity of

Oxygen or Ozone than needed would act as a distractor. This was done to make the game more challenging and to keep in check if the players were collecting everything instead of collecting only the required quantities. All the collectible elements were initially displayed in a static white color whether they were distractors or not. However, distractor elements become red when picked up and the rest were displayed in a glowing green color indicating that they were not distractors. The player received a kickback and possible health loss depending on the content level they were playing on picking up a distractor. When the player came in the proximity of a collectible, it randomly became either a required molecule or a collectible with an equal probability. The "GO" (completion text) text appeared once all the required molecules were collected. However, if there were some distractors that were not yet collected, then there was a 50% chance that the completion text would show up and a 50% chance that the distractor would be displayed. When the player collected the completion text the same equation appeared (as a quiz) which they had balanced with the help of game play mechanics (Figure 3). On hitting the submit button, the next content level was loaded irrespective of the wrong or right answer. However, they were given 1 more attempt before submitting if the answer was wrong. The game consisted of four content levels, each having its own background music that gets more intense as the player moved to higher content levels. The balanced equation for each content level is enumerated below:

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

For the cryptography version of the game, each content level aimed to encode or decode a piece of text using the encryption key provided to the player. Similar to the chemistry version, there were either different or excess letters present that would act as a distractor. The task and its corresponding solution for each content level are listed below:

1. Encrypt the Plain Text: "ATTACK AT DAWN" using the Key: 2
   Resulting encryption = "CVVCEMCVFCYP"
2. Decrypt the Cipher Text: "EFGFOE UIF DBTUMF" using the Key: 1
   Resulting decryption = "DEFENDTHECASTLE"
3. Encrypt the Plain Text: "PURA VIDA" using the Key: 13
   Resulting decryption = "CHENIVQN"
4. Decrypt the Cipher Text: "URON RB KNJDCRODU" using the Key: 9
   Resulting decryption = "LIFEISBEAUTIFUL"

Affdex Software Development Kit Affdex Software Development Kit (SDK) from Affectiva [24] was integrated into the Chemo-o-crypt game. SDK tracked the facial features of the players to output the probabilities for their emotions with a sampling rate of 20 Hz. A template size of 640 by 480px (height by width)

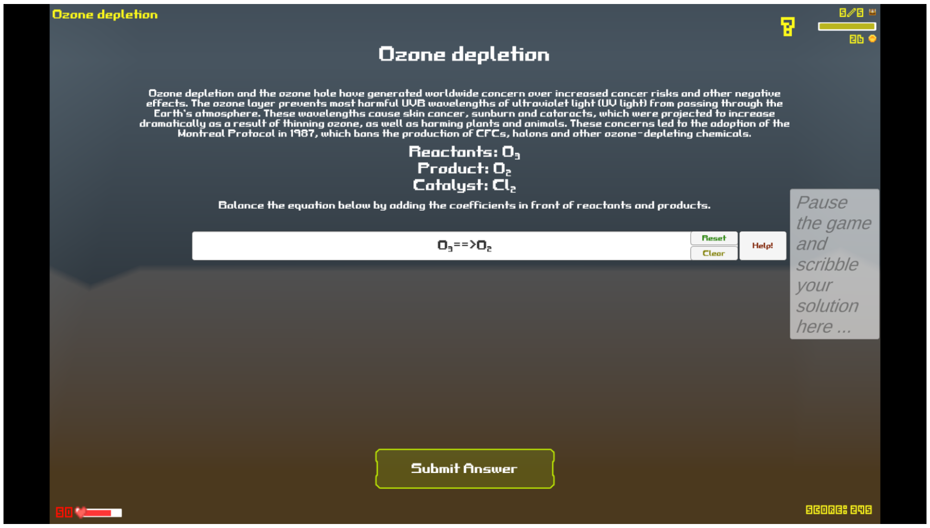

Fig. 3. Screen capture of the level-end task or quiz which appeared on collecting the completion text.

was used to capture their face. When the player moved out of the field of view of the camera, then the game paused itself, asking the player to re-orient themselves so that the camera could detect their face. SDK traced the seven basic Ekman emotions of Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Contempt, in real-time. The output probability ranged from 0 (emotion absent) to 100 (emotion fully present). The SDK also tracked the physical properties of 15 different facial features (facial expressions) which included Attention, BrowFurrow, BrowRaise, ChinRaise, EyeClosure, InnerBrowRaise, LipCornerDepressor, LipPress, LipPucker, LipSuck, MouthOpen, NoseWrinkle, Smile, Smirk, UpperLipRaise. These expressions correspond to the action units from Ekman and Friesen's Facial Action Coding System [26]. A description of these expressions is available on the iMotions website [27].

Affect detection This experiment utilized the model equations obtained from Verma et al. [28]. Expressions were used to predict the probability value of each affective state for a given time-frame. The one which had the highest value was assigned as the affective state for that time-frame. For example, if the predicted values obtained for boredom, flow, and frustration were .23, .53, and .64 respectively, then the affective state of frustration was assigned to that particular time-frame. All this data was then aggregated for the entire time-period during which they stayed in the chunk, and then the affect with the most frequent occurrence during that time-period was assigned to the event of chunk crossing.

Difficulty increased when the aggregate state detected during the chunk boundary-crossing event was boredom and the player score was at least 40%

of the maximum possible score at that instant. Conversely, it decreased when the aggregated state was frustration and the proportional score was less than 20%. The maximum possible score kept updating itself as the chunk layouts changed, as it was governed by the number of coins and lives that were present in the game environment. The score gain for collecting a coin was fixed at 10 points, while the score for collecting lives (or 1-ups shaped like hearts) depended on its location in the environment. It ranged from 100 to 1000 depending on the ease with which it can be collected. Hard to collect 1-ups gave more points than the easier ones.

Game adaptability Chem-o-crypt game with both chemistry and cryptography content was used for this experiment. The game had a tutorial level which was designed to gauge the game play skills of participants while simultaneously walking them through the game mechanics. The score on this tutorial level was used to assign the initial difficulty as well as the maximum difficulty for the game. If the participants completed the tutorial level without using all the available lives, then the difficulty cap was set to 4, in all other cases it was assigned using the formula, $\mathit{diffcap} = 4 \times \mathit{score} \div \mathit{maxscore}$, where $\mathit{maxscore}$ is the maximum possible score possible during the tutorial level. A player could earn the score by collecting coins and lives and may lose it when they collide with the enemy. For example, if the player achieved a score of 75% of the maxscore, then the maximum value for difficulty would be set to 3. If the $\mathit{diffcap} > 2$, then the initial difficulty was set to two for the participant. Therefore the player performance during the tutorial level was taken into account to set the game difficulty's initial and maximum value irrespective of the condition they were assigned to. However, for the participants that belonged to the adaptivity off condition, their difficulty remained at the initial level throughout the game play. For the participants who were in the group corresponding to adaptivity on, the difficulty may have increased or decreased when they crossed the chunk boundary.

User Engagement Scale This study used the revised User Engagement Scale (UESz) as a psychometric tool adopted from Wiebe et al. [25] for measuring the player engagement in the game Chem-o-crypt. The UESz composed of 28 items was measured as a 5-point Likert scale. It was administered at the end of the game play for each content that was played in Chem-o-crypt.

3.3 Participants

A total of 172 undergraduate students were recruited to take part in this online experiment. This experiment was conducted online due to the pandemic situation which did not allow in-person studies. Thirty-five students did not complete the game without completing a single game. This could be attributed to potential bugs in the game that were not discovered during the game testing or issues with the game user interface (UI) on different screen resolutions since the experiment could not control the game environment due to the online nature of the study.

The game was tested on a computer that had a resolution of 1920 x 1080. It was not possible to test it on other resolutions and therefore the game UI might have appeared differently on different resolutions causing some UI elements to go off-screen or scale abruptly. Further, twenty-six participants completed only the first game but dropped out before they could complete the second game. Consequently, 111 people completed the entire study without dropping out. Table 1 indicates the number of participants in each group (within parentheses) who completed both the contents, completed only one, and dropped out without completing, respectively. Of these 111 participants who completed the study, 91 were male, and the rest female ($M_{age} = 21.6years, SD = 6.17years$). Their participation lasted up to 2 hours ($M_{playtime} = 95minutes, SD = 29.5minutes$) and they were given 2-course credit. Seventy-six participants reported having played games with an average game play time of sixteen hours per week and a standard deviation of fifteen hours.

Table 1. Experiment's factorial design, 2×2, with the number of participants who completed both, one, none of the contents, respectively.

Stealth Assessment \ Play Order	Chemistry 2nd (Crypto 1st)	Chemistry 1st (Crypto 2nd)
On	Chem 2nd, On (19,6,5)	Chem 1st, On (21,4,11)
Off	Chem 1st, Off (35,7,7)	Chem 2nd, Off (36,9,12)

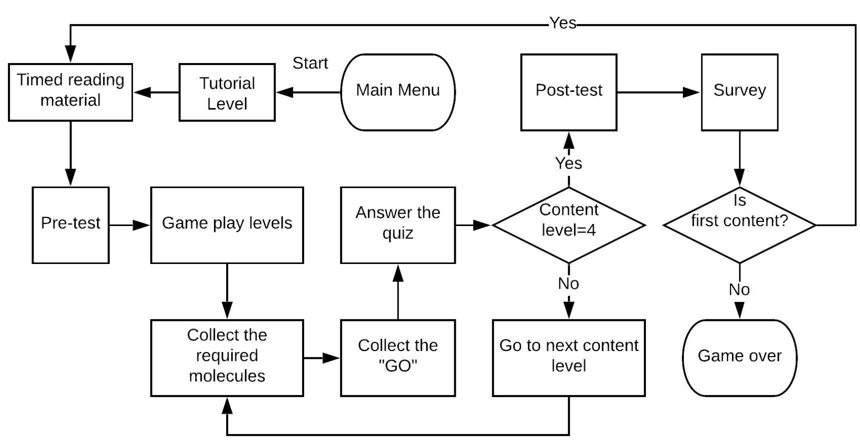

Fig. 4. Flowchart depicting a typical participant workflow for the Experiment.

3.4 Procedure

The experiment took place in an online environment. Upon consenting to partake, participants downloaded the game and instructions from the researcher's google. They were asked to calibrate their webcam before starting the game. 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. As the game started, it assigned the participants into one of the four groups randomly. Then they played the game as per the workflow depicted in Figure 4 until they finished it and were rewarded course-credits upon game completion. There were four content levels for each content and the experiment ended when the player cleared all the 4×2 levels.

4 Results

User engagement score from the UESz scale was analyzed as the dependent variable of interest with order and adaptivity as independent variables. A $2 \times 2 \times 2$ mixed ANOVA was performed with participants' UESz scores for chemistry and cryptography as a within-subject factor and content order and adaptivity as between-subject factors.

Table 2. Mean and SD for UESz by condition.

	Adaptivity	Order	Mean	SD	N
Chemistry Engagement	OFF	Chem 2nd	81.03	23.614	35
		Chem 1st	80.11	24.03	36
		Total	80.56	23.66	71
	ON	Chem 2nd	78.05	23.33	19
		Chem 1st	92.95	15.96	21
		Total	85.87	20.94	40
	Total	Chem 2nd	79.98	23.34	54
		Chem 1st	84.84	22.16	57
		Total	82.48	22.77	111
Cryptography Engagement	OFF	Chem 2nd	85.94	21.46	35
		Chem 1st	76.06	27.70	36
		Total	80.93	25.14	71
	ON	Chem 2nd	84.42	23.26	19
		Chem 1st	79.05	26.20	21
		Total	81.60	24.68	40
	Total	Chem 2nd	85.41	21.90	54
		Chem 1st	77.16	26.96	57
		Total	81.17	24.87	111

The $2 \times 2 \times 2$ mixed ANOVA did not indicate any significant three-way interactions between the variables. There was no main effect of Adaptivity. However, there was a significant main effect observed for the content order (See Table 3 for ANOVA results and Table 2 for means, standard deviations by the group).

Table 3. Results from the $2 \times 2 \times 2$ mixed ANOVA for UESz.

	$F(1, 107)$	p	η^2
Engagement	.777	.380	.007
Adaptivity	.437	.510	.004
Order	.006	.941	.000
Engagement * Adaptivity	1.227	.270	.011
Engagement * Order	14.893	$<.001$	.122
Adaptivity * Order	1.406	.238	.013
Engagement * Adaptivity * Order	2.225	.139	.020

When played as first content, there was no significant difference between the chemistry and cryptography UESz scores, $t(109) = .135$, $p = .893$; $Cohen's\ d = .026$. When played as second content, there was no significant difference either, $t(103) = .425$, $p = .672$; $Cohen's\ d = .083$.

4.1 Chemistry

When adaptation was off, the mean for the chemistry UESz score was higher (81.03) when played as second content as compared to when played as first content (80.11). However, when adaptation was on, the mean was lower when chemistry was played as second content. Overall, the adaptivity condition had a higher average (85.87) as compared to no adaptivity (80.56); and playing chemistry as the first content led to a higher mean (84.84) in comparison to playing it second (79.98). The chemistry UESz score when played as first content was not significantly different from the score when it was played as second content, $t(109) = 1.12$, $p = .263$; $Cohen's\ d = .214$.

4.2 Cryptography

When adaptation was off, the mean for the cryptography UESz score was higher (85.94) when played as first content as compared to when played as second content (76.06). However, when adaptation was on, the mean was lower when cryptography was played as second content, a pattern similar to that observed for chemistry content. Overall, the adaptivity condition had a higher average (81.60) as compared to no adaptivity (80.93); and playing cryptography as the first content led to a higher mean (85.41) in comparison to playing it second (77.16). The cryptography UESz score when played as first content was not significantly different from the score when it was played as second content, $t(109) = 1.764$, $p = .081$; $Cohen's\ d = .335$.

5 Discussion

This analysis examined adaptivity using affect assessment in a CAGE game to determine its effectiveness for sustaining engagement when playing multiple games that use content agnostic mechanics. There was no significant interaction between the adaptation and content order, suggesting that the player engagement was maintained when playing multiple contents within a CAGE game. Although the UESz score differed significantly depending on the content order, it was not significantly different when adaptation came into play. The mean UESz score was better in adaptive condition for both the game contents irrespective of the order but when the adaptation was on, the UESz mean was lower when played as second content.

A previous study by Baron [2] found that engagement was reduced when playing second content in a CAGE game, probably due to fatigue effect or boredom. The current study replicates this finding as the UESz score significantly dropped when second content was played. However, the current study included game adaptation supported by stealth assessment, which might have helped in preventing this decrease in engagement when playing second content. Therefore, current results may suggest that using adaptation can help sustain motivation, but not increase it when playing multiple contents within a CAGE game.

A study conducted by Sharek and Wiebe [29] found that the adaptation in a puzzle-based game led to similar engagement as compared to linear game play or a game play driven by player choices. They used the past and current performance of a player, along with the secondary task and in-game behaviour to select the next game level for the player. The current study adapted the game play differently but obtained the same results. The overall engagement observed was not significantly different in the adaptive game as compared to the non-adaptive game.

However, the current results should be interpreted with caution as the estimated effect sizes for the analysis were rather low. The ANOVA showed that the means were not significantly different due to adaptation but the effect size was small. The partial eta squared was .011, which means that the adaptation by itself accounted for only 1.1% of the overall variance in the scores. Similarly, order explained 12.2% but together order and adaptation accounted for only 2% of the observed variance in the score.

A major limitation of the experiment was the online nature of the study. It had to be conducted online due to the pandemic situation. As a result, there was no control over the system in which the game was being run and therefore many participants dropped out of the study. A total of 35% of the participants did not complete the study which could be attributed to potential bugs or issues with the game user interface (UI) as indicated previously. However, the study did not appear to have a problem of attrition in relation to any specific condition as the dropouts appeared to be random irrespective of the condition.

6 Conclusion

Currently, there is very limited research that provides a content agnostic way to create an educational video game. This results in more time and cost requirements to build any game for research or commercial purposes. Although a content agnostic game can help tackle this issue, it does not entice a player when they repeatedly play the game with different learning content. The proposed research ameliorates this by building an adaptive stealth assessment into a CAGE game even though the adaptation in the current game environment did not work as hypothesized.

Bibliography

[1] Pablo Moreno-Ger, Ivan Martinez-Ortiz, Manuel Freire, Borja Manero, and Baltasar Fernandez-Manjon. Serious games: A journey from research to application. In *2014 IEEE Frontiers in education conference (FIE) proceedings*, pages 1–4. IEEE, 2014.

[2] Tyler Baron. *An Architecture for Designing Content Agnostic Game Mechanics for Educational Burst Games*. PhD thesis, Arizona State University, 2017.

[3] Miguel Sicart. Defining game mechanics. *Game Studies*, 8(2), 2008.

[4] Tyler Baron and Ashish Amresh. Word towers: Assessing domain knowledge with non-traditional genres. In *European Conference on Games Based Learning*, page 638. Academic Conferences International Limited, 2015.

[5] Jeffrey Chin, Richard Dukes, and William Gamson. Assessment in simulation and gaming: A review of the last 40 years. *Simulation & Gaming*, 40 (4):553–568, 2009.

[6] Jan L Plass, Bruce D Homer, Charles K Kinzer, Yoo Kyung Chang, Jonathan Frye, Walter Kaczetow, Katherine Isbister, and Ken Perlin. Metrics in simulations and games for learning. In *Game analytics*, pages 697–729. Springer, 2013.

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

[8] Stephen Tang, Martin Hanneghan, and Abdennour El Rhalibi. Introduction to games-based learning. In *Games-based learning advancements for multi-sensory human computer interfaces: Techniques and effective practices*, pages 1–17. IGI Global, 2009.

[9] Tyler Baron, Corey Heath, and Ashish Amresh. Towards a context agnostic platform for design and assessment of educational games. In *European Conference on Games Based Learning*, page 34. Academic Conferences International Limited, 2016.

[10] Scotty Craig, Sidney D'Mello, Barry Gholson, Amy Witherspoon, Jeremiah Sullins, and Arthur Graesser. Emotions during learning: The first steps toward an affect sensitive intelligent tutoring system. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pages 264–268. Association for the Advancement of Computing in Education (AACE), 2004.

[11] Robert J Mislevy, Russell G Almond, and Janice F Lukas. A brief introduction to evidence-centered design. *ETS Research Report Series*, 2003(1): i–29, 2003.

[12] Valerie Shute, Matthew Ventura, Matthew Small, and Benjamin Goldberg. Modeling student competencies in video games using stealth assessment. *Design recommendations for intelligent tutoring systems*, 1:141–152, 2013.

[13] Matthew Ventura, Valerie Shute, and Matthew Small. Assessing persistence in educational games. *Design recommendations for adaptive intelligent tutoring systems: Learner modeling*, 2(2014):93–101, 2014.

[14] Valerie Shute, Iskandaria Masduki, Oktay Donmez, Vanessa P Dennen, Yoon-Jeon Kim, Allan C Jeong, and Chen-Yen Wang. Modeling, assessing, and supporting key competencies within game environments. In *Computer-based diagnostics and systematic analysis of knowledge*, pages 281–309. Springer, 2010.

[15] Vipin Verma, Tyler Baron, Ajay Bansal, and Ashish Amresh. Emerging practices in game-based assessment. In *Game-Based Assessment Revisited*, pages 327–346. Springer, 2019.

[16] Yoon J Kim and Valerie Shute. Opportunities and challenges in assessing and supporting creativity in video games. In *Video games and creativity*, pages 99–117. Elsevier, 2015.

[17] Valerie Shute, Matthew Ventura, and Yoon Jeon Kim. Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research*, 106(6):423–430, 2013.

[18] Valerie Shute and Lubin Wang. Measuring problem solving skills in portal 2. In *E-Learning Systems, Environments and Approaches*, pages 11–24. Springer, 2015.

[19] Valerie Shute. Stealth assessment in computer-based games to support learning. *Computer games and instruction*, 55(2):503–524, 2011.

[20] Valerie Shute and Yoon Jeon Kim. Does playing the world of goo facilitate learning. *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning*, pages 359–387, 2011.

[21] Igor Mayer, Dirk van Dierendonck, Theodore Van Ruijven, and Ivo Wenzler. Stealth assessment of teams in a digital game environment. In *International Conference on Games and Learning Alliance*, pages 224–235. Springer, 2013.

[22] Hansol Rheem, Vipin Verma, and D. Vaughn Becker. Use of mouse-tracking method to measure cognitive load. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1):1982–1986, 2018. doi: 10.1177/1541931218621449.

[23] Geoffrey T Crisp. Assessment in next generation learning spaces. In *The future of learning and teaching in next generation learning spaces*. Emerald Group Publishing Limited, 2014.

[24] Martin Magdin and F Prikler. Real time facial expression recognition using webcam and sdk affectiva. *IJIMAI*, 5(1):7–15, 2018.

[25] Eric N Wiebe, Allison Lamb, Megan Hardy, and David Sharek. Measuring engagement in video game-based environments: Investigation of the user engagement scale. *Computers in Human Behavior*, 32:123–132, 2014.

[26] Paul Ekman and WV Friesen. Facial action coding system: A technique for the measurement of facial movement, consulting psychologists press. *Palo Alto*, 1978.

[27] iMotions Inc. Affectiva channel explained. https://help.imotions.com/hc/en-us/articles/360011728719-Affectiva-channel-explained?flash_digest=bc976eecb2594b9e177c096f4bf19208e25d613e&flash_digest=e0c132629b4dd862a0d36813c8dfdfffe15bfff3, 2018. Accessed: 2020-04-17.

[28] Vipin Verma, Hansol Rheem, Ashish Amresh, Scotty D Craig, and Ajay Bansal. Predicting real-time affective states by modeling facial emotions captured during educational video game play. In *International Conference on Games and Learning Alliance*, pages 447–452. Springer, 2020.

[29] David Sharek and Eric Wiebe. Investigating real-time predictors of engagement: Implications for adaptive videogames and online training. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 7(1):20–37, 2015.