# Validity of a Content Agnostic Game Based Stealth Assessment

Vipin Verma[1(✉)], Ashish Amresh[2], Scotty D. Craig[1], and Ajay Bansal[1]

[1] Arizona State University, Mesa, AZ, USA
{vverma9,scotty.craig,ajay.bansal}@asu.edu
[2] Arizona State University, Tempe, AZ, USA
amresh@asu.edu

**Abstract.** In an attempt to predict the learning of a player during a content agnostic educational video game session, this study used a dynamic bayesian network in which participants' game play interactions were continuously recorded. Their actions were captured and used to make real-time inferences of the learning performance using a dynamic bayesian network. The predicted learning was then correlated with the post-test scores to establish the validity of assessment. The assessment was moderately positively correlated with the post-test scores demonstrating support for its validity.

**Keywords:** Game based assessment · Sensor-free · Stealth assessment · Dynamic bayesian network · Educational games

## 1 Introduction

Serious games are prodding the education domain to delineate the meaning of learning and assessment which are both challenging in nature [1,9]. While developing a game itself takes a lot of time, adding assessment adds another time-consuming process to it. To ease this process, our study used the content agnostic game engineering (CAGE) architecture [2] embedded with stealth assessment [15,16,18,19].

Baron [2] created CAGE architecture to facilitate the development of multiple educational games at once. It develops game mechanics that are independent of the learning content taught in the game and is called content agnostic mechanics. This allows using a single game to teach multiple learning contents, saving time and money which would otherwise be used to create multiple games to teach all of them. Further, Shute and colleagues [16] developed stealth assessment strategies for learning games that use player interactions as evidence that can aid in inferring learning while the player is actively playing the game. Our study used stealth assessment within CAGE to create a content agnostic game based assessment strategy. We present our results with validating the assessment so that it can be adopted as an effective strategy for educational games [8].


F. de Rosa et al. (Eds.): GALA 2021, LNCS 13134, pp. 121–130, 2021.
https://doi.org/10.1007/978-3-030-92182-8_12

## 2 Content Agnostic Game Engineering

Usually an educational game is designed to teach a specific learning content and to teach another content it is required to develop another game specific to that content. However, designing and developing an educational video games is a time and cost intensive process [2]. Therefore, teaching multiple content would mean creating multiple different games, leading to more effort, more time and money requirements. CAGE can alleviate this problem by using content agnostic mechanics that separates the game mechanics from the learning content. This is not only beneficial for researchers, but also benefits industry due to savings in terms of time and money.

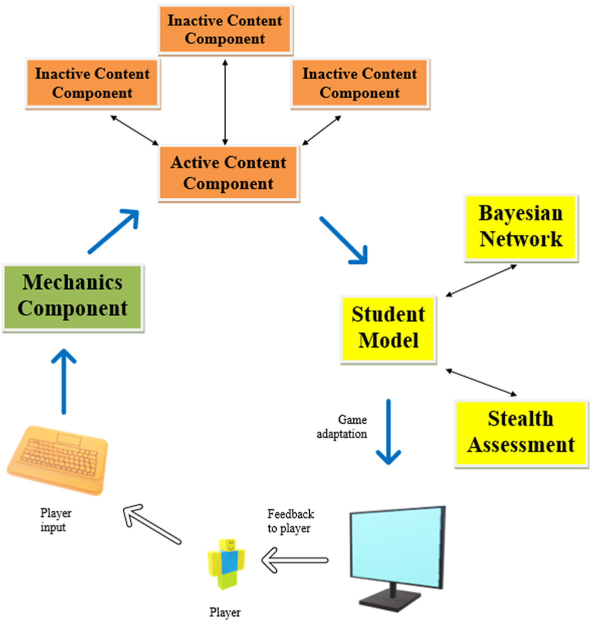


**Fig. 1.** CAGE model for educational game development.

CAGE follows a component-based architecture consisting of three components; mechanics component, content component, and student model as depicted in Fig. 1. The mechanics component is designed to be content agnostic. CAGE consists of multiple content components which can be switched at anytime. Student model creates a model of the student learning using an embedded stealth assessment by utilizing the player interactions within the game. Multiple stealth assessment strategies can be combined to create a student model [18]. These three components are held together with the help of the underlying framework that connects the player input with the game mechanics.

## 3 Dynamic Bayesian Network

A Bayesian Network (BN) uses probabilistic graph to model several variables which are conditionally dependent on each other [7]. A Dynamic Bayesian Network (DBN) is a type of BN that allows modeling time-series and sequences using variables that have probabilistic dependence over a period called lag or time-steps [13,14]. A simple DBN called as knowledge tracing [6] was used as a basis for the current study, shown in Fig. 2. The Figure shows a 2-quiz series that consists of three nodes; a participant node (S), a question node (Q), and a knowledge node (K). There are four performance parameters which are prior knowledge, slip rate, learn rate, and guess rate, that model the network.

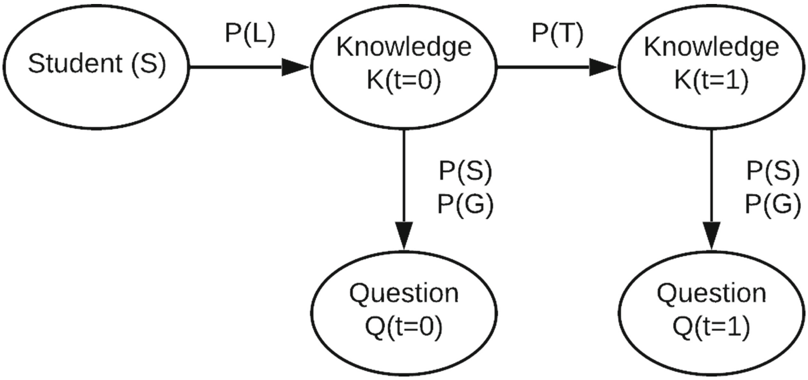


**Fig. 2.** Two time slices of the knowledge tracing model [6].

An individual learner (S) is represented using the student node and regulates the prior knowledge parameter P(L) which depicts the prior knowledge level of a learner before they start playing the game [12]. A diagnostic pre-test can be used to obtain the parameter value. State of learner's knowledge at any point in time is modeled using the latent knowledge node (K). This node is replicated across each time-step and is termed as a temporal node. At any time-step, the knowledge node is conditionally dependent on the knowledge node from the previous time-step. This dependence of knowledge node with itself across time-steps is depicted using learn rate which is also called as transition rate P(T). It determines the probability that the learner will transition from an unlearned state to a learned state when moving to the next time-step. The question node (Q) depicts the question which was asked to measure learner's knowledge. It consists of two states, true and false, which represents learner's answer being correct or incorrect. It is conditionally dependent on the time-specific knowledge level of the participant. The question node is replicated across all the time-steps and is therefore temporal. It consists of two parameters, slip rate P(S) and the guess rate P(G). Slip rate models the probability of answering incorrectly when a learner possess the knowledge, while guess rate accounts for guessing correctly despite not having the required knowledge.

# 4 Method

## 4.1 Participants

The current study was conducted online due to the Covid-19 pandemic. It recruited 172 undergraduate students of which 61 students did not complete the study due to potential bugs in the game or the issues with the game interface when playing on devices with different resolutions. A total of 111 participants completed the study (91 male, 20 female, average age 21.6 years, standard deviation of 6.17). Sixty-one participants reported having played games with an average play time of sixteen hours per week and a standard deviation of fifteen hours. Their participation lasted up to 2 h (mean playtime of 95 min, standard deviation of 29.5 min).

## 4.2 Background

CAGE architecture [2,3] was used to create a 2D platformer game titled “Chemo-o-crypt” in Unity3D (v2018.1.9f2). In CAGE, the game play mechanics were designed in such a manner that they were not tightly connected to the educational content. This allowed the same mechanics to be used to teach multiple subjects without significantly modifying the game play mechanics. Educational games that follow this architecture are getting popular as content dependent mechanics make the game development costly and time-consuming [4].

**Game Mechanics.** In Chemo-o-crypt, the game play mechanics allowed left and right player movement, jumping, and ladder climbing. Three different types of enemies patrolled certain areas within the game environment. It reduced a portion of the player health when they collided with an enemy. The game environment also consisted of two different types of static hazards, spikes and water. On falling into these hazards, the player health was immediately reduced to zero.

The game consisted of two learning contents, chemistry and cryptography. Each learning content had 4 distinct levels representing a learning problem for that given level. During each level, the player was tasked with collecting the required number of elements corresponding to the goal of that level. The game map had coins and heart-shaped items (1-up) which were scattered at various locations. A player initially had three lives which could be increased by collecting one hundred coins or a 1-up.

**Game Content.** For the cryptography content, the players were required to collect the alphabetical letters corresponding to the encryption/decryption of a given text using the encryption key provided to them. However, there were either excess or different letters present in the game map which would act as a distractor. This was deliberately done to see if a player is collecting everything instead of collecting only required letters. It was also expected to make the game more stimulating for learners. All the collectible letters were displayed in white color. However, the distractor letters became red when picked while the rest were shown in glowing green ink signaling that they were not distractors. Operant conditioning was implemented to prevent the players from collecting distractors [17]. On picking up a distractor, players received a kickback and some health loss which was governed by the content level. For example, during the first content level, there was no health loss, during the second content level there was some health loss, but during the last content level the player instantly died on collecting a distractor. A collectible element randomly became either a required letter or a distractor when the player came into its proximity. On collecting all the required letters, the "GO" (completion text) appeared. But if there were distractors which were not yet collected, there was a 50% chance that a collectible would become a distractor instead of becoming a completion text. On collecting the completion text, the same encryption/decryption problem appeared (as a quiz) which they solved with the help of game play mechanics. The next content level was loaded on answering the quiz, irrespective of the answer being right or wrong. The game had four content levels, each having their own background music which got more intense as the player moved to higher and more complex content levels. The task along with solutions for each level are enumerated 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"

For the chemistry content, players were tasked with collecting the exact number of molecules that partake in a given chemical reaction to balance it. The chemical equation for the first content level require 3 Oxygen ($O_2$) and 2 Ozone ($O_3$) molecules to balance. The excess of these molecules would act as a distractor. The balanced equation for each content level is listed 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. $Al_2O_3 + 6\,HCl \longrightarrow 2\,AlCl_3 + 3\,H_2O$

**Stealth Assessment.** The DBN used for assessment in the game is shown in Fig. 3. To implement the network in the game, Bayes Server 8.17 [5] was used. The network had eleven nodes, of which five were temporal (Knowledge1, Distractor10, Distractor11, Distractor12, and Question1). In addition to the Prior Knowledge, and Question node from the knowledge tracing model [12], the network consisted of several distractor nodes. Each of them represented the event of picking up a distractor. Distractor0x nodes denoted the evidence that player has collected this distractor at time-step $t = 0$ (i.e. content level 1). However, Distractor1x nodes designated the evidence for time-steps $t = 1, 2, 3$ (i.e. content level 2, 3, 4 respectively). To assign the value to Prior node, a pre-test containing twenty questions was used which would serve as evidence for the Knowledge0 node which is conditionally dependent on it. Knowledge0 node has two states, true and false, indicating the presence or absence of knowledge required to achieve the level aim. It expresses the probability that the player possess the required knowledge at the time-step $t = 0$ while Knowledge1 if for the time-steps $t = 1, 2, 3$. The probabilities for the first content level were intentionally made different from the rest of the content levels with the help of the complex DBN structure. This allowed for learning the game mechanics on the first content level, since players were not initially aware that they were not allowed to collect distractors. While learning the game mechanics during first content level, the probabilities of collecting the distractor were expected to be higher when compared to the rest of the content levels. The conditional probabilities of the network were estimated using a combination of parameter learning [11] and expert raters advice. An earlier experiment consisting of 107 participants was used to collect the data for parameter learning. The estimated probabilities are depicted in Table 1 and 2.


**Fig. 3.** DBN used for assessment in the Chem-o-crypt.

**Table 1.** Conditional probabilities used for Prior and Knowledge0 node.

| Prior score states (pre-test score) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Knowledge0 | | .01 | .01 | .01 | .03 | .04 | .05 | .05 | .08 | .30 | .37 | .05 |
| | True | .40 | .42 | .44 | .46 | .48 | .50 | .52 | .54 | .56 | .58 | .60 |
| | False | .60 | .58 | .56 | .54 | .52 | .50 | .48 | .46 | .44 | .42 | .40 |

**Table 2.** Conditional probabilities used for Distractor, Knowledge1, and Question nodes.

| Knowledge0 | Distractor00 | | Distractor01 | | Distractor02 | | Question0 | | Knowledge1 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 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 |
| Knowledge1 | Distractor10 | | Distractor11 | | Distractor12 | | Question1 | | 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 |

**External Assessment.** The experiment consisted of pre and post-test which were isomorphic in nature. They consisted of 20 multiple choice questions and randomized the order or the questions as well the available choices within each question. Pre-test was used as a diagnostic assessment for evaluating the state of the prior node while post-test was used as a summative assessment to be compared with the inference of knowledge from the DBN at the end of the game play.

### 4.3 Procedure

Participants downloaded the game and instructions upon consenting to partake in the study. Figure 4 shows a typical participant workflow during the game. On starting the game, player walked through a tutorial level where the basic game play mechanics to navigate the gaming environment were introduced. After the tutorial level, they went through the reading material for the first learning content, followed by the pre-test and the actual game play. On finishing the four content levels for the first learning content, they completed the post-test and survey and went to the reading material for the second learning content. Each learning content consisted of four content levels. The game ended when player finished all the $4 \times 2$ levels. After finishing the workflow they were rewarded course-completion credits.

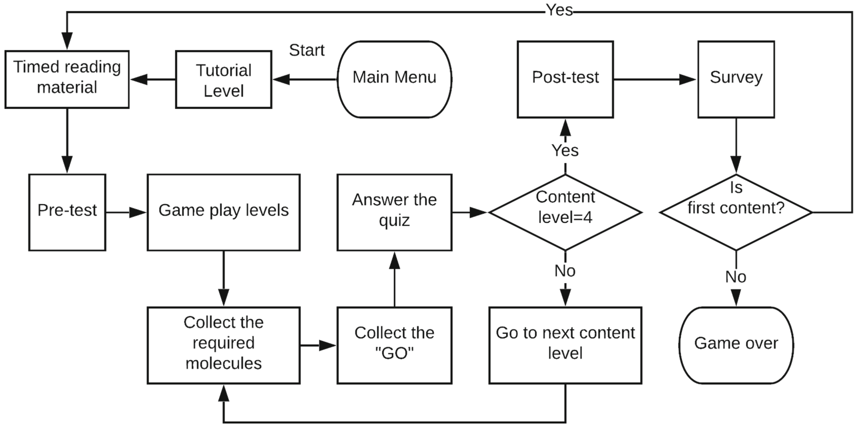


**Fig. 4.** Typical participant workflow for the experiment.

## 5 Results

To check the relationship between the post-test scores and the knowledge inferred from the DBN, a correlation analysis was conducted. The probability value of the Knowledge1 node at the end of game play was treated as the inferred knowledge for the purpose of comparison with the post-test scores. Cryptography post-test score was found to be positively correlated to the knowledge inferred from the DBN, $r_s(111) = .46$, $p < .001$. Chemistry post-test score was positively correlated as well, $r_s(111) = .36$, $p < .001$. The overall post-test scores for the two contents combined were also positively correlated to the knowledge inferred from DBN, $r_s(222) = .31$, $p < .001$. Overall, DBN showed a small but significant correlation of knowledge with the post-test scores for both the contents.

## 6 Discussion

Correlations observed between the post-test score and the knowledge level inferred from the DBN were significant and small to moderately positive. This shows validity for using the BNs to model learner beliefs in the CAGE based games. Previously BNs have been used to predict students' final grades in university education [10]. The current experiment demonstrated that the game based stealth assessment based on DBN can be applied in a content agnostic way as the same network was used for both the learning contents in the game. Thus, DBN acts as a valid instrument to create a content agnostic game based stealth assessment.

There are possible avenues for improving the DBN constructed for use in this study. For example, the current game used only three distractors to keep it simple. Instead of a fixed number, a variable number of distractors could be employed in the game and the network structure be modified to reflect that.

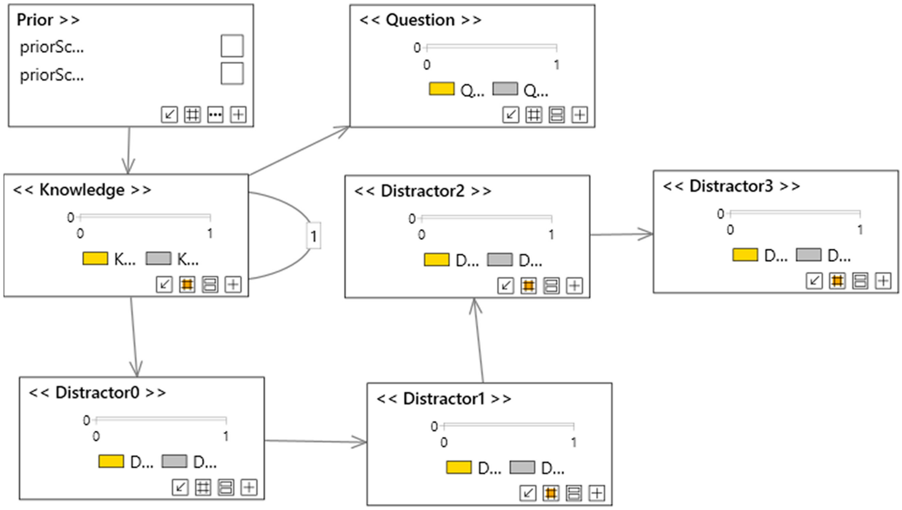


**Fig. 5.** A possible DBN with a different structure.

Apart from this, there are alternate network structures which could be employed to model the current scenario in the game such as one depicted in Fig. 5. This network consists of four distractors. Only the first distractor in this structure has conditional dependence on knowledge, while rest of the distractors are conditionally dependant on the previous distractor that was collected.

## 7 Conclusion

A significant small to moderately positive correlation between the post-test scores and knowledge inferred from the DBN demonstrated support for the validity of the content agnostic game based assessment using DBN. CAGE helps in reducing the time and cost requirement for creating an educational video game. The current evidence in support of the validity of the DBN further adds to the aim of the CAGE games by embedding a content agnostic assessment within the game. Using the current findings, educational games can support embedding the assessment within CAGE games to provide for a holistic educational environment that not only supports learning but also assess it.

## References

1. Amresh, A., Verma, V., Baron, T., Salla, R., Clarke, D., Beckwith, D.: Evaluating gamescapes and simapps as effective classroom teaching tools. In: European Conference on Games Based Learning, p. 22-XII. Academic Conferences International Limited (2019)
2. Baron, T.: An architecture for designing content agnostic game mechanics for educational burst games. Ph.D. thesis, Arizona State University (2017)

3. Baron, T., Amresh, A.: Word towers: assessing domain knowledge with non-traditional genres. In: European Conference on Games Based Learning, p. 638. Academic Conferences International Limited (2015)
4. Baron, T., Heath, C., Amresh, A.: Towards a context agnostic platform for design and assessment of educational games. In: European Conference on Games Based Learning, p. 34. Academic Conferences International Limited (2016)
5. BayesServer: Dynamic Bayesian networks - an introduction (2020). https://www.bayesserver.com/docs/introduction/dynamic-bayesian-networks
6. Corbett, A.T., Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. User Model. User-Adap. Inter. **4**(4), 253–278 (1994). https://doi.org/10.1007/BF01099821
7. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Mach. Learn. **29**(2–3), 131–163 (1997). https://doi.org/10.1023/A:1007465528199
8. Ifenthaler, D., Eseryel, D., Ge, X.: Assessment for game-based learning. In: Ifenthaler, D., Eseryel, D., Ge, X. (eds.) Assessment in Game-Based Learning, pp. 1–8. Springer, New York (2012). https://doi.org/10.1007/978-1-4614-3546-4_1
9. Kim, Y.J., Ifenthaler, D.: Game-based assessment: the past ten years and moving forward. In: Ifenthaler, D., Kim, Y.J. (eds.) Game-Based Assessment Revisited. AGL, pp. 3–11. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15569-8_1
10. Mouri, K., Okubo, F., Shimada, A., Ogata, H.: Bayesian network for predicting students' final grade using e-book logs in university education. In: 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT), pp. 85–89. IEEE (2016)
11. Neapolitan, R.E.: Learning Bayesian Networks, vol. 38. Pearson Prentice Hall, Upper Saddle River (2004)
12. Pardos, Z.A., Baker, R.S., San Pedro, M.O., Gowda, S.M., Gowda, S.M.: Affective states and state tests: investigating how affect and engagement during the school year predict end-of-year learning outcomes. J. Learn. Anal. **1**(1), 107–128 (2014)
13. Reichenberg, R.: Dynamic Bayesian networks in educational measurement: reviewing and advancing the state of the field. Appl. Measur. Educ. **31**(4), 335–350 (2018)
14. Reye, J.: Student modelling based on belief networks. Int. J. Artif. Intell. Educ. **14**(1), 63–96 (2004)
15. Rheem, H., Verma, V., Becker, D.V.: 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. SAGE Publications Sage, Los Angeles (2018)
16. Shute, V.J., Ventura, M., Bauer, M., et al.: Melding the power of serious games and embedded assessment to monitor and foster learning: flow and grow. In: Serious Games, pp. 317–343. Routledge (2009)
17. Skinner, B.F.: The Behavior of Organisms: An Experimental Analysis. BF Skinner Foundation, Cambridge (2019)
18. Verma, V., Baron, T., Bansal, A., Amresh, A.: Emerging practices in game-based assessment. In: Ifenthaler, D., Kim, Y.J. (eds.) Game-Based Assessment Revisited. AGL, pp. 327–346. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15569-8_16
19. Verma, V., Rheem, H., Amresh, A., Craig, S.D., Bansal, A.: 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.) GALA 2020. LNCS, vol. 12517, pp. 447–452. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63464-3_45