Utilizing Cognitive Load Theory and Evidence-Centered Design to Inform the Design of Game-Based Learning Environments

Dolly Bounajim[1], Arif Rachmatullah[1], Danielle Boulden[1], Bradford Mott[1], James Lester[1], Trudi Lord[2], Frieda Reichsman[2], Paul Horwitz[2], Chad Dorsey[2], Eric Wiebe[1]
North Carolina State University, Raleigh, North Carolina[1]
Concord Consortium, Concord, Massachusetts[2]

Digital game-based learning (DGBL) environments are increasingly utilized to facilitate classroom instruction. For the game in our study, a formative stealth assessment tool, in the form of an intelligent tutoring system (ITS) is guided by evidence-centered assessment design (ECD). Cognitive Load Theory and ECD are utilized as diagnostic tools to analyze upsurges in hints delivered by the ITS and inform game design revisions that will promote improved learner support and learning outcomes.

INTRODUCTION

Since technology has become ubiquitous in educational settings, it has been utilized as a tool to provide students with meaningful learning experiences. This new avenue for education has been observed in the progression of gamification of learning objectives (Shute, 2011). With the intention of providing a motivating and engaging framework for learning, the use of game design in the development of educational pedagogy has proved effective (Hosseini, Hartt, & Mostafapour, 2019). The current evolution of gaming technologies is able to harvest real-time user data to create adaptive learning environments. For example, gaming environments embedded with intelligent tutoring systems (ITS) assess student knowledge while providing active guidance during gameplay. Embedded within the framework of the ITS, evidence-centered assessment design (ECD) is utilized as a formative stealth assessment tool (Shute, 2011) to inform learner support tools such as an ITS. A current challenge from a UX perspective is how active guidance works in concert with the overall learning context, including the individual's cognitive state, to promote learning. Therefore, it is important to understand the limitations of our human cognitive architecture to inform both the game design and its underlying intelligent analytic tools. We used cognitive load theory (Sweller, 2011) in addition to evidence-centered assessment design (ECD) (Mislevy & Hartel, 2006) to explore new methods for better understanding how game design decisions affect student learning opportunities.

Cognitive Load Theory

Cognitive load theory (CLT) is an instructional theory based on aspects of human cognitive architecture as it is integrated with instructional design and learning procedures (Sweller, 2011). As it applies to this theory, cognitive architecture has two essential facets, both of which are interrelated. First recognizing working memory, our conscious processor of information, has limited capacity and duration when dealing with novel, unorganized information. Second being long term memory (LTM) as the unlimited storage of varying cognitive schemas (knowledge structures) that categorize information for intended use (Kalyuga & Liu, 2015). From this, cognitive load could be defined as a multidimensional construct representing working memory resources required for a given learning task and is influenced by a students' level of prior knowledge towards elements related to the task. Most crucially is how students' cognitive knowledge structure and the structure of the learning task affect the working memory during the time of the task at hand (Windell, Wiebe, Converse-Lane, & Beith, 2006).

Cognitive load theory distinguishes two main types of cognitive load: intrinsic and extraneous (Kalyuga & Liu, 2015). Intrinsic cognitive load accommodates the relevant information components needed to achieve learning goals for the task at hand and is dependent on the level of learner prior knowledge. Learning tasks can be defined by the number of informational (instructional) elements which the learner needs to manage in working memory. The integration and processing of these instructional elements in working memory is referred to as the level of element interactivity in working memory. A higher level of expertise provides schemas that can chunk these instructional elements, lowering the element interactivity and thus the load on working memory. In contrast, extraneous cognitive load exists in cases of sub-optimal instructional design. This can occur in instances where instruction design requires learners to engage in cognitive processes that are not required (irrelevant) to learning. It can also occur when the learning task is not appropriately matched to the ability level of the learner.

Sources of cognitive load have various implications within instructional environments. When applied, cognitive load theory provides a lens to examine effects of learner prior knowledge, information representations (e.g., graphic versus text), task conditions with the represented information, as well as how information is arranged over space and time in the learning environment (Kalyuga, 2007). Within this scope, the variability effect acts as an implication of cognitive load when scaffolding instruction to provide students with target information in multiple representations and learning task types. The variability effect suggests that high variability better facilitates learning in comparison to tasks with less variability (Paas and van Merriënboer 1994). However, interaction with degree of expertise suggests that levels of variability should be tailored to learner expertise (Likourezos, Kalyuga, & Sweller, 2019). That is, more knowledgeable learners benefit from high variability problems because their level of element interactivity of the conceptual components of the task is relatively low. Thus, there is availability in working memory capacity to handle the raised element interactivity related to the variability of task and representational form of the instructional materials. Conversely, novice learners contend with high load from both the conceptual components of the task and the task variability, leading to an overload of working memory (Likourezos, Kalyuga, & Sweller, 2019). This mismatch of learner to learning task can be considered an extraneous load.

When applied correctly, the variability effect should consider the student's level of expertise–thus level of intrinsic load–in the problem-solving tasks. Sound game design will leverage students' current knowledge by utilizing representations and tasks that use this knowledge as a foundation to build and reinforce it (Hosseini, Hartt, & Mostafapour, 2019). Following cognitive load theory, when learners are provided with external guidance, such as from an ITS, if this information aligns well with the task at hand, it can provide both an immediate learning opportunity and scaffolding that lowers intrinsic load and brings the task within range of working memory capacity. However, misdirected ITS support presents representations of information that do not align with existing knowledge structures, requiring the learner to mentally coordinate instructional elements not germane to the learning task, thus a form of extraneous load (Kalyuga, 2007).

Additionally, physical design elements should be designed and displayed in space and time in ways that minimize extraneous load. Separation of related elements may hinder performance as it poses a source of extraneous load (Kalyuga, 2007). As elements of information need to be processed simultaneously to complete the task, separation in display leaves learners to hold units of information in working memory while other units must be searched for, attended to, and processed. Suggesting that close proximity of the display elements would alleviate extraneous cognitive load, allowing learners to attend to relevant information simultaneously.

These instructional design principles were explored within each of their examples in the results of this study, as their violations result in extraneous load for the student. Additionally, sources of extraneous load are assessed based on whether instances occur momentarily or continuously throughout the game. Alongside these examinations, considerations of the scaffolded support provided by an ITS and guided by the ECD frameworks involved are valuable in making conclusions regarding seemingly problematic areas, as they may work with or go beyond the confines of cognitive load theory.

Evidence-centered assessment design (ECD)

Analysis of trace data of student behaviors are guided by an evidence-centered design (ECD) framework for assessing student knowledge (Mislevy & Hartel, 2006). At the heart of this framework is the idea that assessment is an evidentiary argument of student learning from imperfect evidence. ECD identifies different components, or models, used in the development of assessments. Most germane to this study is 1) the task model, which defines learning tasks designed to allow students to demonstrate their knowledge of key skills or concepts; and 2) the evidence model, which uses student behaviors in the learning environment to provide evidence of their mastery (or lack thereof) of these skills and knowledge components. Evidence is accumulated over multiple tasks as to whether a student is likely to have demonstrable mastery. This process is probabilistic and acknowledges that students may or may not be able to demonstrate mastery in differing contexts. This premise is of particular interest to our study. Embedded within the design framework of an ITS, ECD is utilized as a formative stealth assessment tool (Shute, 2011). Stealth assessment is a process embedded in instructional environments in which learner behavioral data is gathered continuously during gameplay. In efforts to further utilize this data and create robust game-based learning systems, ECD guides the mapping of learning goals to game activity, identifying evidence of learning mastery within the gameplay. This incorporation of ECD elicits the idea that student knowledge can be assessed through synthesis of multiple forms of imperfect evidence across elements of play within the game. An ITS can, in turn, utilize ECD-guided evidence data to drive its tutorial engine. For example, evidence that students have not mastered a concept could trigger delivery of hints to the student to help scaffold them through the mastery of a new concept or skill.

When a gameplay element is mapped via ECD to evidences of mastery there is an implied relationship with the cognitive load a student would experience. Interpreted through cognitive load theory, a student who is able to demonstrate mastery of a concept would experience low intrinsic load, assuming that high extraneous load has not interfered with their ability to utilize existing schemas of related knowledge. Thus, lack of demonstration of mastery could be due to high extraneous load from poor design of the game element rather than evidence of lack of knowledge/skill needed to demonstrate mastery. In such a case, an ITS could incorrectly interpret evidence and deliver inappropriate hints to a student already subjected to high extraneous load.

In an effort to inform the design of DGBLs, these aspects of cognitive load theory, coupled with the ECD framework, were used as a diagnostic tool in an examination of game tasks (challenges) that evoke abnormal upsurges of hint deliveries within Geniventure. These challenges were inspected with regards to the target concepts for learning and whether high intrinsic or extraneous load is the likely cause of this spike. For the purpose of this paper, we focused on sources of extraneous cognitive load, as it is counterproductive to learning and caused by poor instructional design. In addition, alignment of the ECD-derived evidences of mastery to actual behaviors were also examined since incorrect evidence definition could be a source of erroneous hint delivery and thus its own source of extraneous load. Incorrect alignment could indicate there is, in fact, no cognitive load issues. Thus, applying both a cognitive load perspective and inspection of the ECD mapping of evidences can help improve game design and ITS performance and, ultimately, learning outcomes.

RESEARCH QUESTION

For this study, informed by cognitive load theory and ECD, we investigated the following research question: How do trajectories of student hint data and subsequent analysis of game design elements inform the design of a digital game-based learning environment?

METHODS

Participants

A total of 394 secondary level students participated in this study. The students were recruited through their science or biology teachers who agreed to use the Geniventure game as part of their curricular materials to teach genetics concepts. Demographically, the participants varied in terms of gender, grade level, English language learner (ELL) status, and ethnicity. Regarding grade level, half of the participants were in their tenth grade (50%), 19% in ninth grade and the remaining 31% were seventh, eleventh, and twelfth graders. The participants were equal in terms of gender distribution, with 47% identified as female and 46% as male, with the remaining 7% not providing gender information. Of the total participants, 47% were Non-Hispanic White, 17% Black/African-American, 8% Latinx, and the remaining 28% gave Other as their response. Last, 11% indicated they were ELL.

Materials

Geniventure is the product of a NSF-funded project to support student learning in genetics using epistemic oriented scientific practices. The curriculum was developed on the foundation that genetics can be taught as an active, experimental science. Geniventure is a DGBL that engages students in a greater story narrative as they solve challenges around breeding baby dragons (drakes), who can inherit multiple traits (e.g. wings, horns) through various allele combinations. The game features both genotypic (display of the chromosome) and phenotypic (observable physical characteristics of genes) representations of the drakes. While the drakes are fictional, these breeding practices are designed to reflect accurate, real-world genetics. Students are tasked with problem-solving challenges as their learning is supported through immediate feedback based on their actions within the game in the form of ITS-driven hints and remedial challenges.

Data analysis

Analysis of trace data of student behaviors against an inventory of genetics concepts was guided by an ECD framework for assessing student knowledge through a series of student actions. This analysis provides the evidence used by the ITS to determine when hints are delivered to students. In each of the challenges throughout the game, learner behavior is associated with evidences of mastery of specific concepts. Thus, a series of actions in a challenge that provides negative evidence of mastery prompts delivery of a text-based hint relative to that concept. Challenges could contain from one to three levels of hints. Therefore, to analyze the number of hints delivered by challenge, normalized hint scores were computed to make the hint data comparable. Normalized hints scores were calculated by dividing the average hint for each particular challenge to the maximum number of hints in that particular challenge. This resulted in normalized hint scores ranging from 0 to 1, indicating a range from lack of hint delivery to frequent hint delivery. For the purposes of this analysis, concepts tracked by the ITS were selected for examination based on their prevalence throughout the game (Table 1). The computed normalized hint scores were then grouped based on the concept they measured, and graphed to visualize the trend of hints delivered over the course of challenges associated with a particular concept. The longitudinal alignment of hint score per challenge for the given concept was arranged by challenge along the x-axis, and normalized hint scores on the y-axis. Strong upsurges in hint delivery were visually identified in the graphs. The strongest of these upsurges were selected for further analysis.

Table 1. *Concepts examined*

Concept	In Game Applications
Sex Determination (LG1.A3)	Females have two X chromosomes. Males have one X and one Y.
Simple Dominance (LG1.C2a)	Only one dominant allele is needed to produce the dominant trait.
Recessive Traits (LG1.C2b)	Two recessive alleles are needed to produce a recessive trait.
Incomplete Dominance (LG1.C3)	For some traits, both alleles will have some effect, with neither being completely dominant.

RESULTS AND DISCUSSION

Variability Effect

An analysis of game-based data identified three upsurges in hint trajectories that provide an opportunity to discuss explanations, guided by ECD and cognitive load analysis, that inform the design of DGBL environments.

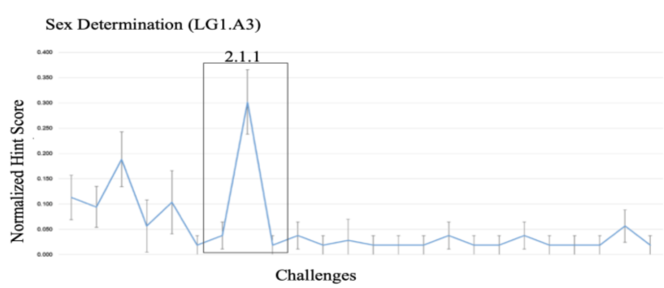

Figure 1. *Hint Trajectories for Sex Determination Concept*

An abnormal upsurge of hints delivered within Challenge 2.1.1 (see Figure 1), calls for an examination of this area of the game, referred to as the Hatchery. Prior to this challenge, students were provided with multiple avenues to determine sex: examining genotypical evidence (abstract visual representation of a gene display), a female to male control switch, or by the visual attribute (phenotypical) of a neck flap that only appears in male drakes (see Figure 2). Following this challenge, differing chromosomes within the genotypic display presented in the Hatchery is the only mode provided to determine sex (pictured in the right display in Figure 3). Thus, Challenge 2.1.1 was the first challenge that learners had to depend on only a genotypic representation for sex determination.

This challenge is in the initial stages of the game since it is important that students learn the genetic component of sex and how it is represented genotypically. In order not to overwhelm novice learners, the initial tasks presented should have low task variability, with variability increasing as they progress and gain experience, encouraging acquisition and transfer between various representations. Data seems to indicate that the variability created by the shift in available genetic information in Challenge 2.1.1 created a short-term rise in cognitive load–possibly representing extraneous load for some learners. However, it can be observed within Figure 1 that students effectively adjusted to this more limited information set in the following challenges. Alongside support from the ITS, the eventual acquisition of the varied genetic information representations is supported by the lessening intrinsic load of the learners. Evidence if this conclusion is supported by a lack of hint upsurges in later challenges presenting evidences for this concept.

Based on an analysis of hint score data and a review of mapped evidences in these challenges, the ECD framework for this concept was mapped correctly.

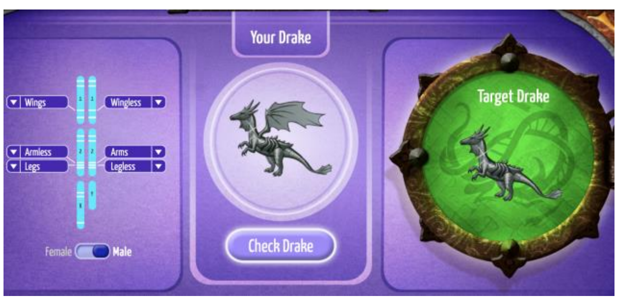

Figure 2. *Challenge 1.2.1*

Figure 3. *Challenge 2.1.1*

External Instructional Guidance

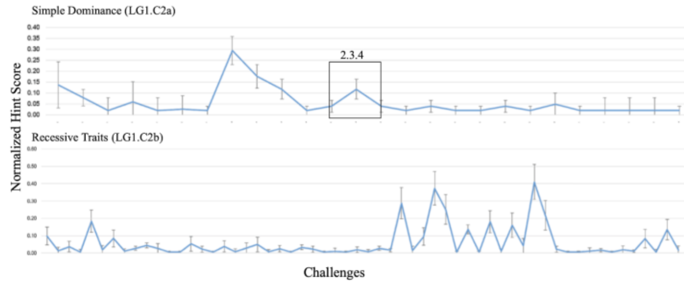

Figure 4. *Hint Trajectories for Simple Dominance (upper) and Recessive Traits.*

Learners historically have struggled with the interrelated concepts of dominant and recessive traits, and how they are expressed phenotypically and genotypically. While the interdependence of these concepts raises the intrinsic load, extraneous load is raised in the game due to a strong shift in representational form from the upper and lowercase letter forms (e.g., WW, Ww, ww) used in most textbooks and learning materials to the underlying genotypic form and resulting phenotypic expression, as seen in Figure 2. For dominant traits with drakes (e.g., wings), the fact that multiple combinations of alleles will result in a correct solution may hide a lack of full understanding of dominant versus recessive traits. The right-hand side of the Recessive Traits (LG1.C2b) graph in Figure 4 shows where challenges requiring knowledge of recessive traits (e.g., horns) begins to be introduced along with the resulting spikes in hint scores. The continued spiking seems to indicate that the game experience has not resolved this conceptual struggle for some students.

A review of the ECD mapping of evidences for these two concepts points to the challenges of separating the intertwined concepts of dominant and recessive. For example, the spike in hint score for Challenge 2.3.4 (Figure 4) occurs for a task focused on a recessive trait (horns) and should not be triggering hints related to dominance, pointing to the difficulty of defining evidence at the trace data level for these two concepts. It further calls into question whether we can assume that the dominant trait concept has been mastered by a majority of students even though the Simple Dominance (LG1.C2a) graph has flattened on the right side.

In summary, DGBL environments should reflect prior understanding of concepts as a means to anchor in existing knowledge structures in order to promote both the acquisition and transfer of knowledge. In this case, a possible solution is that the traditional upper and lowercase letter forms of dominance and recessive could be integrated into the representational forms used in the game. Similarly, the ITS hints should also integrate this form and to treat the composition of hints holistically across these two concepts to more fully address student confusion.

Spatial Separation of Related Elements

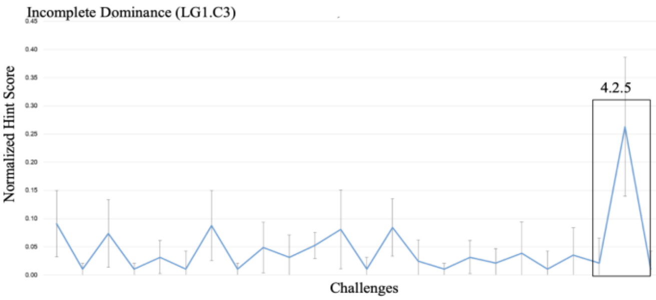

Figure 5. *Hint Trajectories for Incomplete Dominance*

For the concept of incomplete dominance, hint score data (Figure 5) indicates that students are able to correctly solve challenges within an acceptable range of intrinsic load. However, the isolated upsurge of hints is seen when the game introduces the Breeding Room (Figure 6). As pictured in the game display, manipulations to one parent's alleles are made in congruence with those of the other parent. Thus, the introduction of the Breeding Room is an example of a representational shift that creates variability-related load for some students. In addition, this representational form may create additional extraneous load due to its spatial layout. That is, students must evaluate the genotypic attributes of one parent (e.g., Female) to inform decisions of manipulations of the other parent's (e.g., Male) genotype in an attempt to breed the target drake. Parents are visually on the far right and left of the screen, separated by the area where the clutch (i.e., a batch of drakes) is to be presented. Since the information from both sides of the display must be processed simultaneously to complete the task, students suffer from extraneous load due to the inability to visually process both the Male and Female representations simultaneously, adding to the working memory load. Based on an analysis of hint score data and a review of mapped evidences in these challenges, the ECD framework for this concept was mapped correctly.

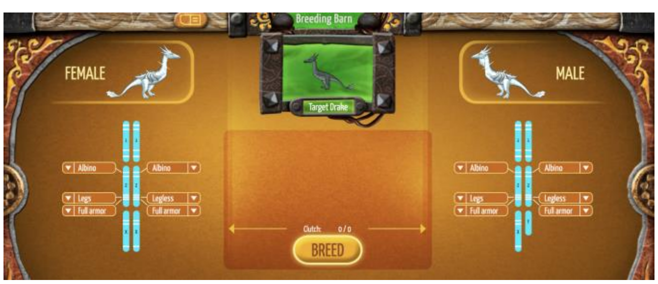

Figure 6. *Challenge 4.2.5, Breeding Room*

CONCLUSION

This investigation utilized cognitive load theory and ECD to analyze hint trajectories of select learning concepts within Geniventure to inform potential game design revisions intended to improve student learning opportunities. In the context of the game, cognitive load theory and ECD frameworks consider how effectively: 1) the designed display of needed information for the learning task responds to known limitations of human cognitive architecture, and 2) the ITS identifies and responds to evidences of student mastery of key genetics concepts. Through these avenues, we can attribute explanations to fine-grained data that go beyond post-hoc, summative assessment of student conceptual understanding.

The results of this study point to the utility of hint generation patterns to identify specific game challenges that should be scrutinized more carefully for potential game design revisions. Considerations of task variability (and its associated information representations), prior knowledge, and working memory load, work to highlight ramifications of game design on student learning outcomes. Further, it provides guidance to the design of ECD frameworks which drive hint generation in the ITS.

This study needs to acknowledge the limitations of our approach. Identification of hint upsurges was done qualitatively and could, in fact, be created by factors not considered by this analysis. Perhaps more importantly, the standardized hint scores represent the aggregation of all students and do not consider individual student differences. Finally, while our analysis points to potential design revisions to the game, these revisions have not been implemented and assessed for efficacy.

This work is supported in part by the National Science Foundation through Grant DRL-1503311. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation. We would also like to thank our participating teachers and students.

REFERENCES

Hosseini, H., Hartt, M., & Mostafapour, M. (2019). Learning Is Child's Play. *ACM Transactions on Computing Education*, *19*(3), 1–18.

Kalyuga, S. (2007). Enhancing instructional efficiency of interactive E-learning environments: A cognitive load perspective. *Educational Psychology Review, 19*(3), 387-399.

Kalyuga, S., & Liu, T. (2015). Guest editorial: Managing cognitive load in technology-based learning environments. *Journal of Educational Technology & Society*, *18*(4), 1-8.

Likourezos, V., Kalyuga, S., & Sweller, J. (2019). The variability effect: When instructional variability is advantageous. *Educational Psychology Review, 31*(2), 479-497.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice, 25*(4), 6-20.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. Computer games and instruction, 55(2), 503e524.

Sweller, J. (2011). Chapter Two - cognitive load theory. *Psychology of Learning and Motivation, 55*, 37-76.

Windell, D., Wiebe, E., Converse-Lane, S., & Beith, B. (2006). A Comparison of Two Mental Workload Instruments in Multimedia Instruction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(17), 1764–1766.