Measuring Problem Solving Skills in Plants vs. Zombies 2

Valerie J. Shute
Florida State University
1114 West Call Street
Tallahassee, FL 32306
vshute@fsu.edu

Gregory R. Moore
Florida State University
1114 West Call Street
Tallahassee, FL 32306
grm13@my.fsu.edu

Lubin Wang
Florida State University
1114 West Call Street
Tallahassee, FL 32306
lw10e@fsu.edu

ABSTRACT
We are using stealth assessment, embedded in *Plants vs. Zombies 2*, to measure middle-school students' problem solving skills. This project started by developing a problem solving competency model based on a thorough review of the literature. Next, we identified relevant in-game indicators that would provide evidence about students' levels on the various problem-solving facets. Our problem solving model was implemented in the game via Bayesian networks. To validate the stealth assessment, we ran a small pilot study to collect data from students who played our game-based assessment and completed an external problem solving measure (*MicroDYN*). Preliminary results indicate that problem solving estimates derived from the game significantly correlate with the external measure, suggesting that our stealth assessment is valid. Our next steps include running a larger validation study (in progress) and developing tools to help educators interpret the results of the assessment.

Keywords
Stealth Assessment, Problem Solving, Game-Based Learning, Bayesian Networks

1. INTRODUCTION
In this paper, we describe the design, development, and preliminary validation of an assessment embedded in a video game to measure the problem solving skills of middle school students. After providing a brief background on stealth assessment and problem solving skills, we describe the game (*Plants vs. Zombies 2*) used to implement our stealth assessment, and discuss why it is good vehicle for assessing problem solving skills. Afterwards, we present the in-game indicators (i.e., gameplay evidence) of problem solving, describing how we decided on these indicators and how the indicators are used to collect data about the in-game actions of players. While discussing the indicators, we show how the evidence is used in a Bayesian network to produce an overall estimate for students' problem solving skills. We then discuss the results of a pilot validation study, which show that our stealth assessment estimate of problem solving significantly correlates with an external measure of problem solving (*MicroDYN*). We conclude with the next steps in developing the assessment and practical applications of this work.

2. BACKGROUND
2.1 Stealth Assessment
Good games are engaging, and engagement is important for learning. The challenge is validly and reliably measuring learning in games without disrupting engagement, and then leveraging that information to bolster learning. For the past 6-7 years, we have been researching various ways to embed valid assessments directly into games with a technology called *stealth assessment* (e.g., [15, 16, 20]). Stealth assessment is grounded in an assessment design framework called evidence-centered design (ECD) [10]. In general, the main purpose of any assessment is to collect information that will allow the assessor to make valid inferences about what people know, can do, and to what degree (collectively referred to as "competencies" in this paper). ECD defines a framework that consists of several conceptual and computational models that work in concert. The framework requires an assessor to: (a) define the claims to be made about learners' competencies, (b) establish what constitutes valid evidence of a claim, and (c) determine the nature and form of tasks or situations that will elicit that evidence.

Stealth assessment complements ECD by determining specific gameplay behaviors (specified in the evidence model and referred to as indicators) and linking them to the competency model [19]. As students interact with tasks/problems in a game during the solution process (see Figure 1), they are providing a continuous stream of data (captured in a log file, arrow 1) that is analyzed by the evidence model (arrow 2). The results of this analysis are data (e.g., scores) that are passed to the competency model, which statistically updates the claims about relevant competencies in the student model (arrow 3).

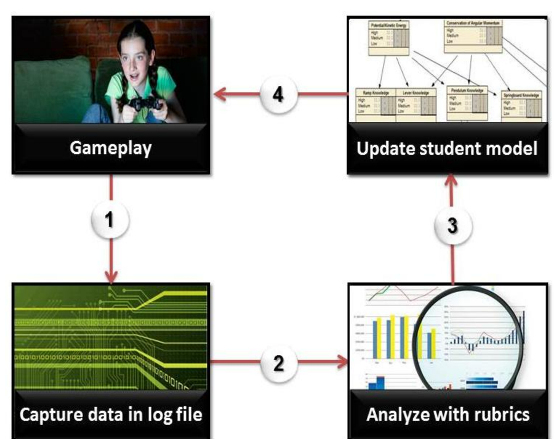

Figure 1. Stealth assessment cycle.

The ECD approach, combined with stealth assessment, provides a framework for developing assessment tasks that are explicitly linked to claims about personal competencies via an evidentiary chain (i.e., valid arguments that connect task performance to competency estimates), and are thus valid for their intended purposes. The estimates of competency levels can also be used diagnostically and formatively to provide adaptively selected levels, feedback, and other forms of learning support to students as they continue to engage in gameplay (arrow 4). Given the dynamic nature of stealth assessment, it is not surprising that it promises advantages, such as measuring learner competencies continually, adjusting task difficulty or challenge in light of learner performance, and providing ongoing feedback.

Examples of stealth assessment prototypes, designed to measure a range of knowledge and skills—from systems thinking to creative problem solving to causal reasoning—can be found in relation to the following games: *Taiga Park* [18], *Oblivion* [20], and *World of Goo* [17], respectively. For the game *Physics Playground* (formerly Newton's Playground, see [19]), three stealth assessments were created and evaluated in relation to the validity and reliability of the assessments, student learning, and student enjoyment (see [21]). The stealth assessments correlated with associated external validated measures for construct validity and demonstrated reliabilities around .85 (i.e., using intraclass correlations among the in-game measures such as number of gold trophies received for various objects created). Furthermore, students (167 middle school students) significantly improved on an external physics test (administered before and after gameplay) despite no instruction in the game. Students also enjoyed playing the game (reporting a mean of 4 on a 5-point scale in where 1 = strongly dislike and 5 = strongly like).

Next, we briefly review our focal competency for this project—problem solving skills—and discuss the natural fit between this construct and particular video games (i.e., action, puzzle solving, simulation, and strategy games).

2.2 Problem Solving Skills

Problem solving has been studied by researchers for many decades (e.g., [3, 7, 11]). It is generally defined as any goal-directed sequence of cognitive operations [1] and is seen as one of the most important cognitive skills in any profession, as well as in everyday life [7]. Mayer and Wittrock [9] identified several characteristics of problem solving: (a) it is a cognitive process; (b) it is goal directed; and (c) the complexity (and hence difficulty) of the problem depends on one's current knowledge and skills.

In 1984, Bransford and Stein [2] integrated the collection of research at that time and came up with the IDEAL problem solving model. Each letter of IDEAL stands for an important part of the problem solving process: *Identify* problems and opportunities; *define* alternative goals; *explore* possible strategies; *anticipate* outcomes and act on the strategies; and *look* back and *learn.* Gick [4] presented a simplified model of the problem-solving process, which included constructing a representation, searching for a solution, implementing the solution, and monitoring the solution. Recent research suggests that there are two main facets of problem-solving skills: rule identification and rule application [14, 23]. "Rules" are the principles that govern the procedures, conduct, or actions in a problem-solving context. Rule identification involves acquiring knowledge of the problem-solving environment, while rule application involves controlling the environment by applying that knowledge.

Can problem solving skills be improved with practice? Polya [12] argued that people are not born with problem-solving skills. Rather, people cultivate these skills when they have opportunities to solve problems. Researchers have long argued that a central point of education should be to teach people to become better problem solvers [1, 13]. However, there is a gap between problems in formal education and those that exist in real life. Jonassen [6] noted that the problems students encounter in school are mostly well-defined, which contrasts with real-world problems that tend to be messy, with multiple possible solutions. Moreover, many problem-solving strategies that are taught in school entail a "cookbook" type of memorization and result in functional fixedness, which can obstruct students' ability to solve problems for which they have not been specifically trained. Additionally, this pedagogy can stunt students' epistemological development, preventing them from developing their own knowledge-seeking skills [8]. This is where good digital games—which have a set of goals and complicated scenarios that require the player to generate new knowledge—come in. Researchers (e.g., [22]) have argued that playing well-designed video games can promote problem-solving skills because games require constant interaction between the player and the game, usually in the context of solving many interesting and progressively more difficult problems. However, empirical research examining the effects of video games on problem-solving skills is still sparse. Our research begins to fill this gap.

3. PRESENT WORK

3.1 The Game

We are using a slightly modified version of the game *Plants vs. Zombies 2* (Popcap Games and Electronic Arts) as the vehicle for our problem solving assessment. In *Plants vs. Zombies 2* (*PvZ2*), players must plant a variety of special plants on their lawn to prevent zombies from reaching their house. Each of these plants has different attributes. For example, some plants (offensive ones) attack zombies directly, while other plants (defensive ones) slow down zombies to give the player more time to attack the zombies. A few plants generate "sun," an in-game resource needed to purchase more plants. The challenge of the game comes from determining which plants to use and where to place them in order to defeat all zombies in each level of the game.

We chose *PvZ2* as our assessment environment for two main reasons. First, we are able to alter the game because of our association with the Glasslab. Glasslab has access to the source code for *PvZ2,* so we can make direct changes to the game as needed (e.g., the particular information to be collected in the log files). This is important because it allows us to build stealth assessments directly into the game itself and to make alterations to the design of the game if needed. Second, *PvZ2* requires players to apply problem solving skills. Thus, our stealth assessment will be able to collect data relevant to problem solving and estimate learners' levels (e.g., low, medium, high) on the facets and problem solving as a whole. However, because problem solving is not easily measured, we cannot assess it directly. We instead need to define directly observable, in-game indicators of problem solving and its associated facets.

3.2 Problem Solving Model

Based on a review of the literature, we built a problem solving competency model. We divided problem solving into four facets: (a) analyzing givens and constraints, (b) planning a solution pathway, (c) using tools and resources effectively, and (d) monitoring and evaluating progress. We then identified relevant in-game indicators of the four facets (see Section 3.3 for details). The rubrics for scoring each indicator and the statistical links between the indicators and the competency model variables comprise the evidence model. The competency and evidence models are implemented together in Bayesian networks. We created a unique Bayes net for each game level (42 total) because many indicators do not apply in every level and simple networks make computations more efficient. In the Bayes nets, the overall problem solving variable, each facet, and the associated indicators are nodes that influence each other. Each of the nodes has multiple potential states and a probability distribution that defines the likely true state of the variable. The Bayes nets accumulate data from the indicators and propagate this data throughout the network by updating the probability distributions. In this way, the indicators influence our estimates of the student's problem solving competency and its associated facets dynamically.

3.3 Indicators of Problem Solving

In line with the stealth assessment process, we defined indicators for each of the four facets of problem solving by identifying observable actions that would provide evidence per facet. This was an iterative process which began by brainstorming a large list of potential indicators. After listing all potential indicators, we evaluated each one for (a) *relevance* to their associated facets and (b) the *feasibility* of being implemented in the game. We then removed indicators that were not closely related to the facets or were too difficult or vague to implement. We repeated this process of adding, evaluating, and deleting indicators until we were satisfied with the list of indicators.

In total, there are 32 indicators for our game-based assessment: 7 for analyzing givens and constraints, 7 for planning a solution pathway, 14 for using tools and resources effectively, and 4 for monitoring and evaluating progress. Examples of indicators for each facet are shown in Table 1.

Table 1. Examples of indicators for each problem solving facet

Facet	Examples of Indicators
Analyzing Givens & Constraints	• Plants > 3 Sunflowers before the second wave of zombies arrives • Selects plants off the conveyor belt before it becomes full
Planning a Solution Pathway	• Places sun producers in the back, offensive plants in the middle, and defensive plants up front • Plants Twin Sunflowers or uses plant food on (Twin) Sunflowers in levels that require the production of X sun
Using Tools and Resources Effectively	• Uses plant food when there are > 5 zombies in the yard *or* zombies are getting close to the house (within 2 squares) • Damages > 3 zombies when firing a Coconut Cannon
Monitoring and Evaluating Progress	• Shovels Sunflowers in the back and replaces them with offensive plants when the ratio of zombies to plants exceeds 2:1

3.4 Preliminary Findings

To test the validity of the stealth assessment of problem solving skills, we recruited ten undergraduate students to play *PvZ2* for 90 minutes, as well as complete an external measure of problem solving — MicroDYN [5], a computer-based test in which participants analyzed the relationships between variables in a system and manipulated those variables to achieve a desired state. This comprised our pilot validation study. We correlated the MicroDYN scores with our stealth assessment estimates of problem solving skill to test for construct validity. The results suggest that our game-based assessment is significantly correlated with MicroDYN (r = .74, p = .03). These preliminary findings suggest that our problem solving stealth assessment is valid, but needs to be further tested with a larger sample size. We are currently running a larger validation study with 200 middle-school students and will have the results from that study in time for the EDM conference.

3.5 Limitations

There are several methodological issues with this pilot validation study. First, the sample of students was very small. Second, the participants were not from the target population of our assessment. This pilot was done with undergraduate students, but our target audience is middle school students. It is unclear if similar results will be seen with our target audience. However, middle school students do enjoy playing PvZ2 and our external measure (MicroDYN) has been successfully tested with that age group. Finally, the participants had a very limited amount of time to play the game in the small pilot study. Ninety minutes is only enough time to play about 15-20 of the game's levels. To improve the validity and reliability of the stealth assessment, players need to engage in gameplay for a longer period of time and over multiple sessions.

4. NEXT STEPS

This work is still in its early stages and we have a lot to do before it can have a meaningful impact on education. We are currently running a validation study with 200 middle school students. These students are playing *PvZ2* over three days, one hour per day. On the fourth day, the students complete MicroDYN [5] and a demographic questionnaire. For every 30 students who complete the study, we are examining the results to see if adjustments need to be made to our Bayes nets. This provides us with multiple opportunities to adjust our Bayes nets throughout the course of the validation study. Thus, this larger, ongoing study will help us to create a more valid and reliable assessment.

Our long term goal is to implement the *PvZ2* game-based assessment in middle school classrooms to help educators improve students' problem solving abilities. As part of this effort, we are teaming with Glasslab to create a dashboard that allows educators to easily interpret the results of the assessment — overall and at the individual facet level. The development of this dashboard and other tools to aid the game's implementation will occur alongside our ongoing validation study.

This focus on the validity and practicality of our game-based problem solving assessment makes it much more likely that the assessment will be both accurate and useful in classroom settings. Students can be assessed on problem solving, a key cognitive skill, in an engaging environment that presents rich problem solving situations and can parse complex patterns of students' actions. Teachers get a valuable tool that will allow them to pinpoint students' abilities in various aspects of problem solving and, in turn, help each student improve their problem solving skills. These benefits stem from our use of evidence-centered design, which gives a framework for creating valid assessments, and stealth assessment, which gives us the ability to invisibly embed such assessments into complex learning environments such as games. By embracing evidence-centered design and stealth assessment, other researchers can also create complex and engaging assessments that meet their specific needs.

5. ACKNOWLEDGMENTS

We would like to thank our colleagues at GlassLab for supporting our work assessing problem solving in Plants vs. Zombies 2—specifically Jessica Lindl, Liz Kline, Michelle Riconscente, Ben Dapkiewicz, and Michael John. We also thank Weinan Zhao for his great programming assistance, as well as Sam Greiff and Katarina Krkovic for letting us use *MicroDYN.*

6. REFERENCES

[1] Anderson, J. R. 1980. *Cognitive psychology and its implications.* Freeman, New York, NY.

[2] Bransford, J. and Stein, B.S. 1984. *The IDEAL problem solver: A guide for improving thinking, learning, and creativity.* W. H. Freeman, New York, NY.

[3] Gagné, R. M. 1959. Problem solving and thinking. *Annual Review of Psychology*. 10, 147-172.

[4] Gick, M. L. 1986. Problem-solving strategies. *Educational Psychologist,* 21, 99-120.

[5] Greiff, S. and Funke, J. 2009. Measuring complex problem solving: The MicroDYN approach. In *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*, F. Scheuermann and J. Björnsson, Eds. Office for Official Publications of the European Communities, Luxembourg, Luxembourg, 157-163.

[6] Jonassen, D. H. 2000. Toward a design theory of problem solving. *Educational Technology Research and Development*. 48, 4, 63-85.

[7] Jonassen, D. 2003. Using cognitive tools to represent problems. *Journal of Research on Technology in Education.* 35, 3, 362-381.

[8] Jonassen, D. H., Marra, R., and Palmer, B. 2004. Epistemological development: An implicit entailment of constructivist learning environments. In *Curriculum, plans, and processes of instructional design: International perspectives,* N. M. Seel and S. Dikjstra, Eds. Lawrence Erlbaum Associates, Mahwah, NJ, 75-88.

[9] Mayer, R. E. and Wittrock, M. C. 1996. Problem-solving transfer. *Handbook of educational psychology,* D. C. Berliner and R. C. Calfee, Eds. Macmillan Library Reference, New York, NY, 47-62.

[10] Mislevy, R. J., Steinberg, L. S., and Almond, R. G. 2003. On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives.* 1, 1, 3-62.

[11] Newell, A. and Shaw, J. C. 1958. Elements of a theory of human problem solving. *Psychological Review.* 65, 3, 151-166.

[12] Polya, G. 1945. *How to solve it: A new aspect of mathematical method.* Princeton University Press, Princeton, NJ.

[13] Ruscio, A. M. and Amabile, T. M. 1999. Effects of instructional style on problem-solving creativity. *Creativity Research Journal.* 12, 251-266.

[14] Schweizer, F., Wüstenberg, S., and Greiff, S. 2013. Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences.* 24, 42-52.

[15] Shute, V. J. 2011. Stealth assessment in computer-based games to support learning. In *Computer games and instruction,* S. Tobias and J. D. Fletcher, Eds. Information Age Publishers, Charlotte, NC, 503-524.

[16] Shute, V. J. and Ke, F. 2012. Games, learning, and assessment. In *Assessment in game-based learning: Foundations, innovations, and perspectives*, D. Ifenthaler, D. Eseryel, and X. Ge, Eds. Springer, New York, NY, 43-58.

[17] Shute, V. J. and Kim, Y. J. 2011. Does playing the World of Goo facilitate learning? In *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning,* D. Y. Dai, Ed. Routledge Books, New York, NY, 359-387.

[18] Shute, V. J., Masduki, I., and Donmez, O. 2010. Conceptual framework for modeling, assessing, and supporting competencies within game environments. *Technology, Instruction, Cognition, and Learning.* 8, 2, 137-161.

[19] Shute, V. J. and Ventura, M. 2013. *Measuring and supporting learning in games: Stealth assessment.* The MIT Press, Cambridge, MA.

[20] Shute, V. J., Ventura, M., Bauer, M. I., and Zapata-Rivera, D. 2009. Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In *Serious games: Mechanisms and effects*, U. Ritterfeld, M. Cody, and P. Vorderer, Eds. Routledge, Taylor and Francis, Mahwah, NJ, 295-321.

[21] Shute, V. J., Ventura, M., and Kim, Y. J. 2013. Assessment and learning of qualitative physics in Newton's Playground. *The Journal of Educational Research*. 106, 423-430.

[22] Van Eck, R. 2006. Building intelligent learning games. In *Games and simulations in online learning: Research & development frameworks,* D. Gibson, C. Aldrich, and M. Prensky, Eds. Idea Group, Hershey, PA.

[23] Wüstenberg, S., Greiff, S., and Funke, J. 2012. Complex problem solving — more than reasoning? *Intelligence.* 40, 1-14.