

# MEASURING PROBLEM SOLVING SKILLS IN *PORTAL 2*

Valerie J. Shute<sup>1</sup> and Lubin Wang<sup>2</sup>

*Florida State University*

<sup>1</sup>3205G Stone Building, 1114 W Call Street, Tallahassee, FL32306

<sup>2</sup>3212 Stone Building, 1114 W Call Street, Tallahassee, FL32306

## ABSTRACT

This paper examines possible improvement to problem solving skills as a function of playing the video game *Portal 2*. Stealth assessment is used in the game to evaluate students' problem solving abilities--specifically basic and flexible rule application. The stealth assessment measures will be validated against commonly accepted external measures. For instance, basic rule application will be correlated with Raven's Progressive Matrices, and flexible rule application will be correlated with the Uses test, insight problems, and the compound remote association test. Improved problem solving outcomes will support our claim that *Portal 2* can be an effective method to enhance problem solving skills--one of the most important cognitive skills in the 21<sup>st</sup> century.

## KEYWORDS

Problem solving; game-based learning; stealth assessment

## 1. INTRODUCTION

This paper describes our current research investigating the use of the video game *Portal 2* (Valve Corporation) as a vehicle to assess and ultimately support problem solving skills in high school students. The reason this research is important is because in today's interconnected world, being able to solve complex problems is, and will continue to be of great importance. However, students today are not receiving adequate practice solving such problems. Instead, they are exposed to problems that tend to be sterile and flat in classrooms and experimental settings (e.g., math word problems, Tower of Hanoi). Also, learning and succeeding in a complex and dynamic world is not easily or optimally measured by traditional types of assessment (e.g., multiple-choice responses, self-report surveys). Instead, we need to re-think assessment, identifying skills relevant for the 21st century—such as complex problem solving—and then figuring out how best to assess students' acquisition of the skills.

The organization of our short paper is as follows. We begin with a brief review of problem solving skills. Next, we discuss assessment in games focusing on a systematic approach for designing valid performance-based assessments (i.e., evidence-centered design). This is followed by our study design and outcome measures. We conclude with ideas for future research in the area.

### 1.1 Literature Review

#### 1.1.1 Problem Solving Ability

Problem solving has been studied extensively by researchers for decades (e.g., Gagné 1959; Jonassen 2003; Newell & Shaw 1958). It is generally defined as “any goal-directed sequence of cognitive operations” (Anderson 1980, p. 257) and is regarded as one of the most important cognitive skills in any profession as well as in everyday life (Jonassen 2003). There are several characteristics of problem solving as identified by Mayer and Wittrock (1996): (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.

Researchers have long argued that a central point of education should be to teach people to become better problem solvers (Anderson 1980). And the development of problem-solving ability has often been regarded as a primary goal of the education process (Ruscio & Amabile 1999). But there is a gap between problems in formal education versus those that exist in real life. Jonassen (2000) 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 solutions possible. Moreover, many problem-solving strategies that are taught in school entail a “cookbook” type of memorization, resulting in functional fixedness which can obstruct students’ ability to solve problems for which they have not been specifically trained. Additionally, this pedagogy also stunts students’ epistemological development, preventing them from developing their own knowledge-seeking skills (Jonassen et al. 2004). This is where good digital games (e.g., *Portal 2*) come in--which have a set of goals and unknowns requiring the learner to generate new knowledge.

Recent research suggests that problem solving skills involve three facets: rule identification, rule knowledge, and rule application (Westenberg et al. 2012; Schweizer et al. 2013). Rules in problem solving refer to the pattern of a problem and criteria of how to reach the goal. Because we will be using “stealth assessment” in the game (i.e., an unobtrusive and ubiquitous assessment embedded deeply within the game, e.g., Shute 2011), we will not be able to directly collect data on students’ rule identification and rule knowledge. However, since rule application is the outward expression of one’s rule identification and rule knowledge, the measurement of rule application will reflect students’ ability to identify rules and their rule knowledge.

Any given problem in *Portal 2* requires the application of either basic rules or rules that require cognitive flexibility--i.e., the ability to adjust prior thoughts or beliefs in response to a change in goals (Miyake et al. 2000). Cognitive flexibility is the opposite of functional fixedness, defined as the difficulty that a person experiences when attempting to think about and use objects (or strategies) in unconventional ways (Duncker 1945). Such cognitive rigidity causes people to view a particular type of problem as having one specific kind of solution without allowing for alternative strategies and explanations (Anderson 1983). Research has shown improved cognitive flexibility after playing first-person shooter (FPS) games (Colzato et al. 2010). The authors suggested that video games could be used to train older people to compensate for the decline in cognitive functions such as the ability to adapt to changes in the environment. However, empirical research examining the effects of video games on cognitive flexibility is still sparse. Our research is intended to begin to fill this gap. Below is the internal structure of problem solving skills that guided our research (discussed more in the next section on assessment):

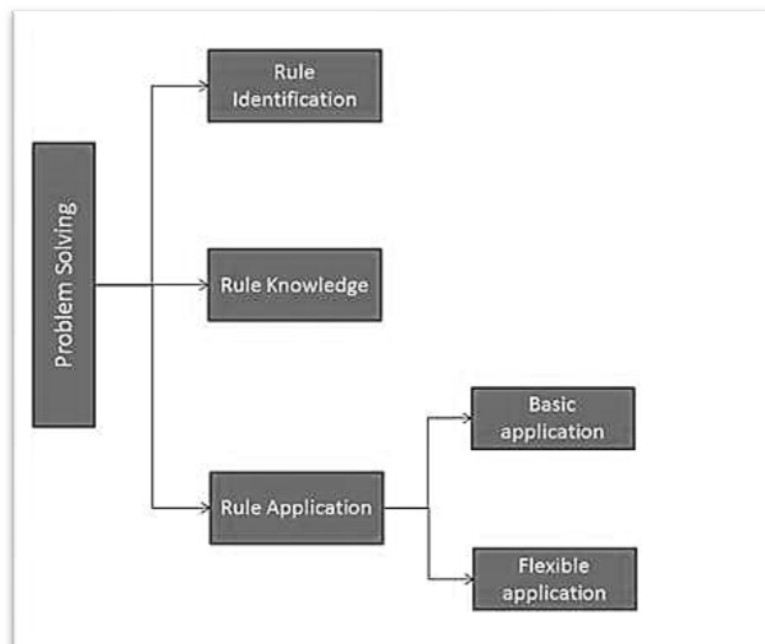


Figure 1. Internal Structure of Problem Solving Skill

### 1.1.2 Assessment with Evidence-Centered Design

Assessments can be deficient or invalid if the tasks or problems are not engaging, meaningful, or contextualized. This calls for more authentic and engaging assessments, which has motivated our recent research efforts in relation to weaving assessments directly and invisibly within good games. In contrast, the amount of engagement in traditional (e.g., paper and pencil, multiple-choice) assessment is negligible. Another downside of traditional assessments is that they often invoke test anxiety, which can be a major source of construct-irrelevant variance. When these problems associated with traditional assessment—inauthentic and decontextualized items, and provoking anxiety—are removed (e.g., by using a game as the assessment vehicle), then the assessment should be more engaging. Additionally, if the assessment is designed properly, such as by using an evidence-centered design approach (Mislevy et al. 2003) then it should also be as, or even more valid than a traditional assessment.

Evidence-centered design provides a way to assess students' levels on target competencies by analyzing evidence extracted from the students' interactions with the game or other complex learning environments (Shute 2011; Shute et al. 2009). Evidence-centered design includes three main models that work together: (a) competency model—which outlines the knowledge, skills or other attributes that are to be assessed; (b) evidence model—which establishes the statistical links between the competencies and their associated metrics; and (c) task model—which identifies the features of tasks that can elicit the necessary evidence. Examples of indicators of the competency in the game will be provided in the external outcome measures section.

For example, in *Portal 2*, if a player follows basic rules such as avoiding harmful objects (e.g., turrets and acid river), or making use of the tools and other objects in the environment (e.g., refraction cubes and light bridges), this provides evidence that the player is competent at basic rule application. Particular facets will be modeled by different game levels selected by us with variation in difficulty. The students' competency will be measured by the time it takes them to finish the tasks.

## 2. METHOD

### 2.1 Participants

Around 220 11-12<sup>th</sup> graders from a high school located in the northern Florida will be recruited for the study. We will try to have equal number of girls and boys to control for any gender effects related to gaming. Demographic data will be collected from all participants. Students who are expert gamers or have beat *Portal 2* before will be excluded from the study. Half of the students will be randomly assigned to the control group, which will play a web-based “brain trainer” game called *Lumosity* (which claims to improve problem solving skills). The other half will be assigned to our experimental condition, playing *Portal 2* (which is a commercial game making no claims relative to learning). Students will be compensated with \$100 for full participation (i.e., 14 hours of gameplay and 2-3 hours of pretests and posttests—our external measures).

### 2.2 Design

Students will play their assigned game for 14 hours. The entire study will be completed at the students' home at their own pace--within a 3-week window. After game play for both groups, students will be instructed on how to extract the gameplay log files from their computer and upload it for our analysis. Students will also take online-based problem solving tests both before and after the intervention. This is intended to (a) validate our stealth assessment measures of problem solving skill, and (b) provide data relative to learning from the game. We hypothesize that students' problem solving skills will improve after the game intervention, possibly more in the experimental than control condition despite the control condition's game being explicitly touted as support for such skills.

## 2.3 Materials

*Portal 2* is a first-person, puzzle-platform, “shooter” game, consisting of a series of provocative puzzles. These puzzles must be solved by teleporting your player’s character and various objects using the “portal gun,” a device that can create inter-spatial portals between two flat planes. To solve the progressively more difficult challenges, players must figure out how to locate, obtain, and then combine various objects effectively to open doors and navigate through the environment, with dangers abounding. *Portal 2* provides a unique environment that can promote problem solving skills through providing players extensive practice figuring out solutions to complex problems on their own. We posit that problem solving skills learned in the game can be transferred beyond the immediate game environment.

*Lumosity*, the intervention for the control group, is a web-based platform that hosts several small-scale games. Advertisements for Lumosity claim the games were designed by neuroscientists, and improve brain health and performance. Some of the games especially feature problem solving and cognitive flexibility. They also claim that the games provide personalized training to different users and that 10 hours of Lumosity training creates drastic improvements.

## 2.4 Assessment in *Portal 2*

Stealth assessment of problem solving will be embedded in the game. We will use original levels in the game as well as several customized levels from the community that help elicit specific evidence for our competency—problem solving ability. A total of 75 levels were collected for the research. Basic and flexible rule application load on different levels with varying weights. A level may be easy on basic rule application but difficult on flexible rule application. Below are examples of how we designed/selected levels to elicit evidence for the two facets of rule application:

- Basic rule application:

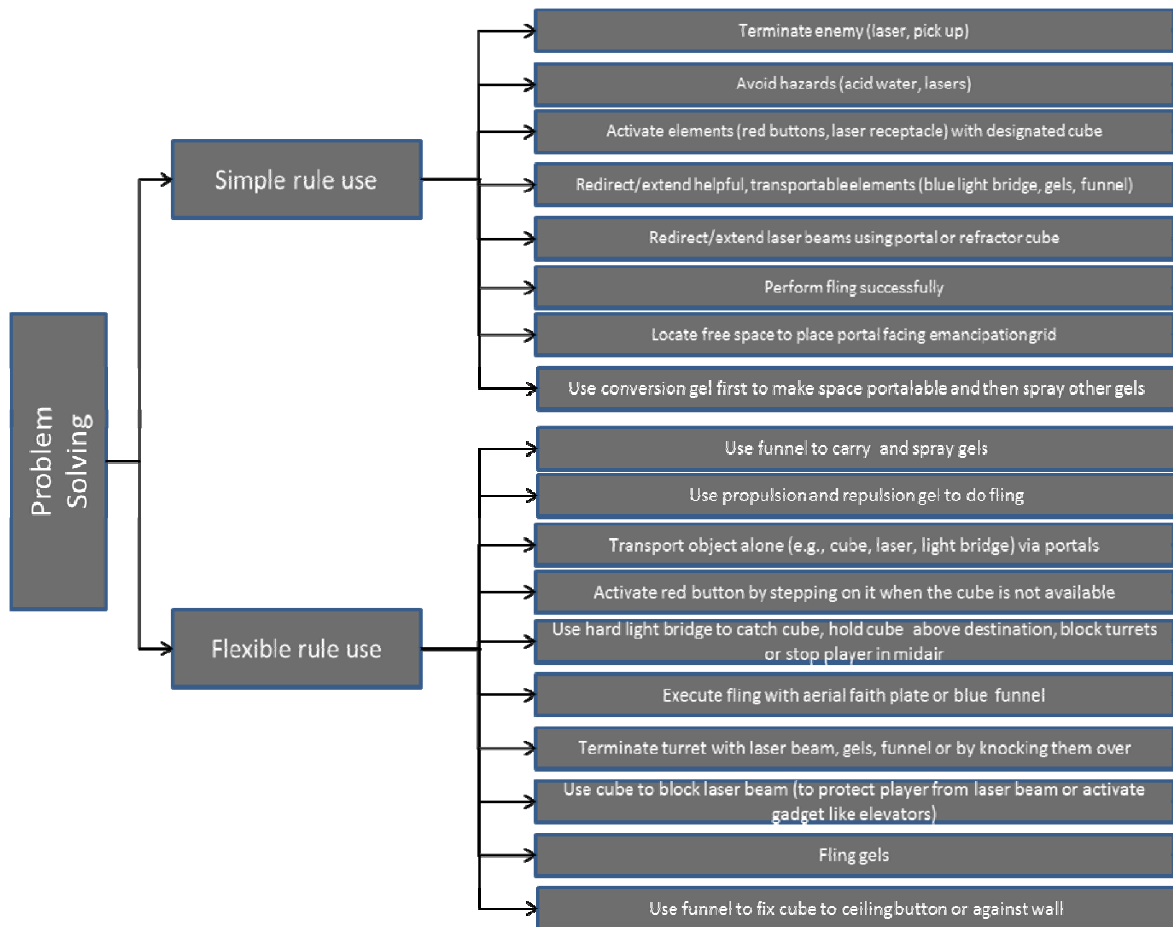
Basic rules in *Portal 2* are rules directly instructed or can be picked up easily. For example, players should be able to learn that the river is hazardous from the cueing picture on the floor near the river. Or if a player fails to notice it and falls into the river, he will die and resurrect from the last automatic saving point. Afterwards, he should be aware of the rule. Other basic rules relate to avoiding laser beams, knocking over turrets to terminate them, putting a cube on the weighted button to activate any device connected to it, etc. Task modeling of basic rule application involves the manipulation of the number of rules present in a level and the difficulty of the rules to be included.

- Flexible rule application:

Flexible rules in *Portal 2* refer to rules that can only be inferred from the basic rules. For example, the basic rule is that the weighted button can be activated by the weight of a cube. Cognitive flexibility requires players to realize that the body weight of the player may be a replacement when a cube is not readily available. Other flexible rules in the game include the use of the hard light bridge to catch a falling cube or to hold it above a destination (e.g., a weighted button to be pressed) and release it after a sequence of actions to be performed before the release. Similar to the task modeling of basic rules, the modeling of flexible rule application involves the number of rules present in a level, the difficulty of a single rule, and the combined use of different rules.

Log files that record students’ performance in gameplay are extracted by enabling the developer console of the game. Students’ problem solving performance can be assessed by information in the log files such as the time it takes to solve a level, number of steps to reach the exit door of each level, the total number of solved puzzles, and the number of attempts before successfully complete specific steps.

Examples of specific indicators per facet of problem solving from the game can be seen in the following figure:

Figure 2. Competency Model of Problem Solving Skills with Indicators from *Portal 2*

## 2.5 External Outcome Measures

The stealth assessment of students' problem solving skills will be validated against external measures of problem solving. Two sub-facets of rule application (i.e., basic rule application and cognitive flexibility) will be measured. Basic rule application will be measured by Raven's Standard Progressive Matrices. The test requires participants to infer the pattern of the missing piece from the other patterns given. Although the test is widely used as an intelligence test (e.g., Prince et al. 1996; Rushton & Jensen 2005), as Raven (2000) pointed out, the Raven's Progressive Matrices focus on two components of general cognitive ability—eductive and reproductive ability. Eductive ability involves making meaning out of confusion and generating high-level schema to handle complexity. Reproductive ability is the ability to recall and reproduce information. In *Portal 2*, for example, players are instructed that the laser beam is deadly. If the player knows this rule, she should realize that the turret is also harmful since it emits a laser beam. We have selected 12 items from the Raven's Progressive Matrices test for the pretest and 12 items for the posttest.

Cognitive flexibility will be measured by three tests: alternative uses, insight problems, and remote association.

The *Alternative Uses test*, developed by Guilford in 1967, requires respondents to find unusual uses for commonly seen objects. For example, the common use for a newspaper is to read it. The test requires examinees to "think of other uses the object or part of the object could serve." Answers may be something like to start a fire, to wrap objects, or to make up a kidnap note (Wilson et al. 1953). We have created four items in the pretest and four items in the posttest and students are allowed four minutes to work on each item.

We posit that playing *Portal 2* can improve performance on the test because many tasks in the game require the players to come up with different uses of a given tool or to deal with a problem via different

methods. For example, the original function of the hard light bridge is to serve as a bridge for the player to walk on. Later, the players need to come up with new uses of the hard light bridge, such as serving as barricade to block turrets.

*Insight problems* are intended to yield an “Aha” moment for problem solvers when the solution occurs after a short or long moment of bafflement (Chu & MacGregor, 2011). Insight problems require problem solvers to shift their perspective and look at obscure features of the available resources or to think of different ways to make use of an object. We have selected five insight problems for the pretest and five for the posttest. For instance: *Marsha and Marjorie were born on the same day of the same month of the same year to the same mother and the same father yet they are not twins. How is that possible?* The answer is that they are triplets. The question is not particularly hard, but it requires problem solvers to break from routine thinking and think beyond the immediate context. The posttest will be an alternative form of the pretest and reliability of the tests will be examined during the study.

*The Remote Association test* was originally developed by Mednick (1962) to test creative thought without any demand on prior knowledge. Each item consists of three words and problem solvers are required to find the solution word associated with all words that appear to be unrelated. The fourth word can be associated with each of the three words in multiple forms, such as synonymy, formation of a compound word, or semantic association (Chermahini, Hickendorff, & Hommel, 2012). For example, the answer to the triad night/wrist/stop is “watch.” Schooler and Melcher (1995) reported that problem solvers’ success on this test correlates with their success on classic insight problems. We have selected five items for the pretest and five for the posttest.

### 3. DISCUSSION AND IMPLICATIONS

We are currently recruiting participants for this study which is scheduled to run during the summer of 2013. In addition to the full set of Portal 2 levels, we have collected levels from the community and also created additional levels (with the Portal 2 “modding” tool) that will allow us to array levels by difficulty to provide for adaptive challenges to support the “zone of proximal development” (Vygotsky, 1987) and “flow” state (Csikszentmihalyi, 1990). By the time of the CELDA 2013 conference, we will have collected all of the data and we will present our findings.

As mentioned earlier, problem solving is one of the most important cognitive skills in any profession and in everyday life (Jonassen, 2003). Positive findings from this study will support the effectiveness of video games in developing this important skill. Using games as assessment vehicles sounds like a great idea, but there are some issues that need to be addressed. For instance, there are potential sources of error variance in videogame assessments such as the level of interest in the target game. However, we believe this will not be a problem with *Portal 2* given its broad appeal (e.g., over 3 million copies have been sold since it came out last year, according to GameFront). In short, we believe that *Portal 2* can be used to assess problem solving better than traditional assessments by virtue of having more authentic, contextualized, and very engaging tasks.

### REFERENCES

- Anderson, J. R., 1980. *Cognitive psychology and its implications*. New York: Freeman.
- Anderson, J. R., 1983. *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Chermahini, S. A. et al. 2012. Development and validity of a Dutch version of the Remote Associates Task: An item-response theory approach. *Thinking Skills and Creativity*, Vol. 7, pp. 177-186.
- Chu, Y. & MacGregor, J. N. 2011. Human performance on insight problem solving: A review. *The Journal of Problem Solving*, Vol. 3, No.2, pp. 119-150.
- Colzato, L. S. et al. 2010. DOOM'd to switch: Superior cognitive flexibility in players of first person shooter games. *Frontiers in Psychology*, Vol.1, pp. 1-5.
- Csikszentmihalyi, M. 1990. *Flow: The psychology of optimal experience*. New York: Harper and Row.
- Duncker, K. 1945. On problem-solving. *Psychological Monographs*, Vol.58, No.5, pp. 1-113.
- Gagné, R. M. 1959. Problem solving and thinking. *Annual Review of Psychology*, Vol.10, pp. 147-172.

- Jonassen, D. H. 2000. Toward a design theory of problem solving. *Educational Technology Research and Development*, Vol.48, No. 4, pp. 63-85.
- Jonassen, D. H. 2003. Using cognitive tools to represent problems. *Journal of Research on Technology in Education*, Vol. 35, No.3, pp. 362-381.
- Jonassen, D. H. et al. 2004. Epistemological development: An implicit entailment of constructivist learning environments. In N. M. Seel & S. Dijkstra (Eds.) *Curriculum, plans and processes of instructional design: International perspectives* (pp. 75-88). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mayer, R. E., & Wittrock, M. C. 1996. Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp.47-62). New York: Macmillan Library Reference.
- Mednick, S. A. 1962. The associate basis of the creative process. *Psychological Review*, Vol.69, pp. 220-232.
- Mislevy, R. J. et al. 2003. On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective*, Vol. 1, No.1, pp. 3-62.
- Miyake, A. et al. 2000. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, Vol. 41, pp. 49-100.
- Newell, A., & Shaw, J. C. 1958. Element of a theory of human problem solving. *Psychological Review*, Vol. 65, No.3, pp. 151-166.
- Prince, M. J. et al. 1996. Is the cognitive function of older patients affected by antihypertensive treatment? Results from 54 months of the Medical Research Council’s treatment trial of hypertension in older adults. *British Medical Journal*, Vol. 312, No. 30, pp. 801-805.
- Raven, J. 2000. The Raven’s progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, Vol.41, pp. 1-48.
- Ruscio, A., M., Amabile, T. M., 1999. Effects of instructional style on problem-solving creativity, *Creativity Research Journal*, Vol.12, pp. 251-266.
- Rushton, J. P., & Jensen, A. R. 2005. Thirty years of research on race differences in cognitive ability. *Psychology, Public Law and Policy*, Vol. 11, No. 2, pp. 235-294.
- Schooler, J. W., & Melcher, J. 1995. The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 97-133). Cambridge, MA: MIT Press.
- Schweizer, F. 2013. Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, Vol. 24, pp. 42-52.
- Shute, V. J. 2011. Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503-524). Charlotte, NC: Information Age Publishers.
- Shute, V. J. et al. 2009. Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295-321). Mahwah, NJ: Routledge, Taylor and Francis.
- Vygotsky, L. S. 1987. *The collected works of L. S. Vygotsky*. New York: Plenum.
- Westenberg, S. 2012. Complex problem solving-more than reasoning? *Intelligence*, Vol. 40, pp. 1-14.
- Wilson, R.C. et al. 1953. The measurement of individual differences in originality. *Psychological Bulletin*, Vol. 50, No. 5, pp. 362-370.