Computers in Human Behavior 63 (2016) 106–117

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Measuring problem solving skills via stealth assessment in an engaging video game

Valerie J. Shute [a, *], Lubin Wang [a], Samuel Greiff [b], Weinan Zhao [a], Gregory Moore [a]

[a] Florida State University, Tallahassee, FL, USA
[b] University of Luxembourg, Walferdange, Luxembourg

ARTICLE INFO

Article history:
Received 27 August 2015
Received in revised form
16 April 2016
Accepted 13 May 2016

Keywords:
Problem solving skills
Stealth assessment
Validity

ABSTRACT

We used stealth assessment, embedded in a game called *Use Your Brainz* (a slightly modified version of *Plants vs. Zombies 2*), to measure middle-school students' problem solving skills. We began by developing a problem solving competency model based on a review of the relevant literature. We then identified in-game indicators that would provide evidence about students' levels on various problem-solving facets. Our problem solving model was then implemented in the game via Bayesian networks. To validate the stealth assessment, we collected data from students who played the game-based assessment for three hours and completed two external problem solving measures (i.e., Raven's Progressive Matrices and MicroDYN). Results indicated that the problem solving estimates derived from the game significantly correlated with the external measures, which suggests that our stealth assessment is valid. Our next steps include running a larger validation study and developing tools to help educators interpret the results of the assessment, which will subsequently support the development of problem solving skills.

1. Introduction

Problem-solving skill is generally defined as a person's ability to engage in cognitive processing to understand and resolve problem situations where a method to solve the problem is not immediately available. According to the Organisation for Economic Cooperation and Development (OECD), problem-solving skill also includes the motivation to engage with such situations in order to "achieve one's potential as a constructive and reflective citizen" (OECD, 2014). This important competency is one that we believe should be fully embraced in our education systems. However, according to the recent OECD Report, students in the U.S. rank 15th out of 44 participating countries on the Programme for International Student Assessment (PISA) Problem Solving test.

The Director of Education and Skills at the OECD recently noted that today's 15-year-old students with poor problem-solving skills will develop into tomorrow's adults attempting to find or keep a good job. He recommended a shift towards supporting problem solving skills in school curricula (OECD, 2014). However, one issue with teaching problem solving skills in a classroom context is that the problems presented in formal education tend to be qualitatively different from those encountered in the real world. That is, problems presented in assessment situations in schools are typically clearly defined and structured, whereas problems in real life are often ill-structured. Well-designed digital games offer a viable alternative to assessing and developing complex problem solving skills that are needed to succeed in the real world (Greiff & Funke, 2009; Greiff et al., 2014; Shute & Wang, in press; Shute, Ventura, & Ke, 2015).

U.S. students' mediocre development of problem solving skills is also of concern to American business leaders, who are dissatisfied with college graduates' lack of problem solving skills. A recent survey of business leaders conducted by the Association of American Colleges and Universities indicates that problem solving skills are increasingly desired by American employers, but only 24% of employers report that recently hired American college graduates are able to analyze and solve complex problems at work (Hart Research Associates, 2015). Therefore, developing good problem solving skills is very important to successfully navigating through school, career, and life in general (Bransford & Stein, 1984; Jonassen, 1997).

In this paper, we describe the design, development, and validation of an assessment embedded in a video game to measure the problem solving skills of middle school students. After providing a

* Corresponding author. Florida State University, 1114 West Call Street, Tallahassee, FL, 32306-4453, USA.
E-mail address: vshute@fsu.edu (V.J. Shute).

http://dx.doi.org/10.1016/j.chb.2016.05.047
0747-5632/

brief background on stealth assessment and problem solving skills, we describe the game (*Use Your Brainz*) used to implement our stealth assessment, and discuss why it is a good vehicle for assessing problem solving skills. Afterwards, we present our competency model and 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 inserted into a Bayesian network to produce overall and facet-level estimates of students' problem solving skills (using an example reported in Wang, Shute, & Moore, 2015). We then discuss the results of a validation study, which suggest that our stealth assessment estimates of problem solving skill correlate significantly with external measures of problem solving (i.e., Raven's Progressive Matrices and 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 (e.g., Arum & Roksa, 2011; Dede, 2009; Taylor & Parsons, 2011). One of the challenges of harnessing the engagement that games can produce for learning is validly and reliably measuring learning in games without disrupting engagement, and then leveraging that information to bolster learning. Over the past eight years, we have examining various ways to embed valid assessments directly into games with a technology called *stealth assessment* (e.g., Shute & Ke, 2012; Shute, 2011; Shute, Leighton, Jang, & Chu, 2016; Shute, Ventura, Bauer, & Zapata-Rivera, 2009). Stealth assessment is grounded in an assessment design framework called evidence-centered design (ECD; Mislevy, Steinberg, & Almond, 2003). The main purpose of any assessment is to collect information that will allow the assessor to make valid inferences about what people know, what they can do, and to what degree they know or are able to do something (collectively referred to as "competencies" in this paper). ECD is a framework that consists of conceptual and computational models that work together harmoniously. The framework requires one to: (a) define the claims concerning learners' competencies, (b) establish what represents valid evidence of a claim, and (c) determine the kind of tasks or situations that will elicit that evidence.

Stealth assessment complements ECD by determining specific gameplay behaviors that can act as evidence of a claim (specified in the evidence model and referred to as indicators) and linking them to the competency model (Shute & Ventura, 2013). As students interact with each problem (or level) in a game during the solution process, they provide an ongoing stream of performance data, captured in a log file. The performance data is automatically analyzed and scored by the evidence model, then inserted into to the competency model, which statistically updates the claims about relevant competencies in the student model (i.e., the instantiated competency model for each individual). The ECD approach, combined with stealth assessment, provides a framework for developing assessment tasks that are clearly linked to claims about personal competencies via an evidentiary chain (i.e., valid arguments that connect task performance to competency estimates), and thus are valid for their intended purposes. The estimates of competency levels can be used diagnostically and formatively to provide adaptively selected game levels, targeted feedback, and other forms of learning support to students as they continue to engage in gameplay. Given the dynamic nature of stealth assessment, it promises advantages such as measuring learner competencies continually, adjusting task difficulty or challenge in light of learner performance, and providing ongoing feedback.

Some examples of stealth assessment prototypes have been described elsewhere (e.g., Shute et al., 2016) — from systems thinking to creative problem solving to causal reasoning relative to the following games: *Taiga Park* (Shute, Masduki, & Donmez, 2010), *Oblivion* (Shute et al., 2009), and *World of Goo* (Shute & Kim, 2011), respectively. For the game *Physics Playground* (see Shute & Ventura, 2013), three stealth assessments—measuring persistence, creativity, and conceptual physics understanding—were created and evaluated for validity and reliability, student learning, and enjoyment (see Shute, Ventura, & Kim, 2013). The stealth assessments correlated with associated externally validated measures for construct validity, and demonstrated reliabilities around 0.85 (i.e., using intraclass correlations among the in-game measures such as the number of gold trophies received). Moreover, 167 middle school students significantly improved on an external physics test (administered before and after gameplay) despite no instructional support relative to the physics content in the game. Students also enjoyed playing the game (reporting a mean of 4 on a 5-point scale, where 1 = strongly dislike and 5 = strongly like).

In summary, some of the benefits of employing stealth assessment include: providing assessments in engaging and authentic environments, reducing or eliminating test anxiety (which can hamper validity), measuring competencies continuously over time, providing ongoing feedback to support learning, and adjusting the difficulty of the learning/gaming environment in response to a person's current level of understanding or skill at various grain sizes (i.e., overall and at the sub-skill level).

Next, we review our focal competency—problem solving skill—in terms of its underlying conceptualization, and discuss the natural fit between this construct and particular video games (i.e., action, puzzle solving, and strategy games).

2.2. Problem solving skills

Problem solving has been studied by researchers for many decades (e.g., Anderson, 1980; Gagné, 1959; Jonassen, 2003; Mayer & Wittrock, 2006; Newell & Shaw, 1958) and is seen as one of the most important cognitive skills in any profession, as well as in everyday life (Jonassen, 2003). Mayer and Wittrock (1996, 2006) 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 integrated the collection of problem-solving 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 (1986) 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 (also see the PISA conceptualization of individual and interactive problem solving, OECD, 2014). More recent research suggests that there are two overarching facets of problem-solving skills that can be empirically distinguished and that usually collate several of the more narrow processes mentioned above: rule (or knowledge) acquisition, and rule (or knowledge) application (Schweizer, Wüstenberg, & Greiff, 2013; Wüstenberg, Greiff, & Funke, 2012). "Rules" are the principles that govern the procedures, conduct, or actions in a problem-solving context. Rule acquisition (or identification) involves acquiring knowledge of the