Computers in Human Behavior 61 (2016) 36–46

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files

Samuel Greiff [a,*], Christoph Niepel [a], Ronny Scherer [b], Romain Martin [a]

[a] University of Luxembourg, Luxembourg
[b] Centre for Educational Measurement at the University of Oslo, Norway

ARTICLE INFO

Article history:
Received 24 August 2015
Received in revised form
23 February 2016
Accepted 24 February 2016
Available online xxx

Keywords:
Complex problem solving
Computer-based assessment
Nonlinear relations
Log file data
Time on task
VOTAT

ABSTRACT

Computer-based assessments of complex problem solving (CPS) that have been used in international large-scale surveys require students to engage in an in-depth interaction with the problem environment. In this, they evoke manifest sequences of overt behavior that are stored in computer-generated log files. In the present study, we explored the relation between several overt behaviors, which $N = 1476$ Finnish ninth-grade students (mean age $= 15.23$, $SD = .47$ years) exhibited when exploring a CPS environment, and their CPS performance. We used the MicroDYN approach to measure CPS and inspected students' behaviors through log-file analyses. Results indicated that students who occasionally observed the problem environment in a noninterfering way in addition to actively exploring it (*noninterfering observation*) showed better CPS performance, whereas students who showed a high frequency of (potentially unplanned) interventions (*intervention frequency*) exhibited worse CPS performance. Additionally, both too much and too little time spent on a CPS task (*time on task*) was associated with poor CPS performance. The observed effects held after controlling for students' use of an exploration strategy that required a sequence of multiple interventions (*VOTAT strategy*) indicating that these behaviors exhibited incremental effects on CPS performance beyond the use of VOTAT.

1. Introduction

The advent of computers in educational assessment has led to a comprehensive shift away from paper-pencil-based assessments to assessments that are administered on computers. This development is exemplified by the way in which the arguably most noted educational large-scale assessment, the Programme for International Student Assessment (PISA; OECD, 2009, 2014), has changed its mode of administration from a paper-pencil-based to a computer-based test administration: In 2003 and 2006, during the first and second PISA cycles, the administration was mostly paper-pencil; in 2009 and 2012, there was a partial shift to computer-based test administration; in the current and the upcoming cycles in 2015 and 2018, there is an even stronger shift toward computer-based assessment, and, in the future, new items might be developed solely for computer-based testing.

One advantage that comes along with this broad shift is that all of the students' observable test-taking behaviors are stored in computer-generated log files and can be accessed to provide additional information beyond students' overall performance. Ever since computers have been available for assessment, the potential of this almost infinite amount of information has been extensively praised. Indeed, the potential applications of information gathered in log files are manifold. For example, they might open up new avenues for understanding how test performance evolves from a research perspective, they might explain how differences between countries are grounded in behavioral differences from an educational policy perspective, and they might integrate assessment and learning through direct feedback and dedicated interventions from an instructional science perspective. Along these lines, the vision that Bunderson, Inouye, and Olsen (1989) had for the advent of different generations of computerized tests has finally led to what they named *intelligent measurement* as the description of a comprehensive integration of behavioral processes for the assessment of students' skills while students learn on the computer.

This vision corresponds, in a sense, to the description of a log-

* Corresponding author. ECCS unit, University of Luxembourg, 11, Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg, Luxembourg.
E-mail address: samuel.greiff@uni.lu (S. Greiff).

http://dx.doi.org/10.1016/j.chb.2016.02.095
0747-5632/

file-driven integration of computerized assessment and learning. But despite the great amount of optimism placed on the potential of log files and despite the high expectations that they can be used to generate a better understanding of student performance, research that provides insights into relevant aspects of high-level cognitive behavior by drawing on log-file processing is relatively scarce (see Thillmann, Gößling, Marschner, Wirth, & Leutner, 2013; for one example). The present work aims to further close this gap.

The more actions and exploratory behavior that a specific skill and the assessment instruments that are targeted toward measuring it require, the longer the behavioral sequences and the richer and the more valuable the information found in log files. Thus, one skill that might potentially produce particularly comprehensive log files is complex problem solving (CPS; Funke, 2001; Kluge, 2008; Wüstenberg, Greiff, & Funke, 2012). Unlike more fundamental skills such as numerical reasoning, CPS relies on complex behavioral patterns, strategic exploratory behavior, and multistep solution processes that are expressed through overt behaviors when students demonstrate their proficiency levels. Computer-based CPS assessments – as employed in the present study – use log files to capture these behaviors. Because of its importance as an educational outcome, CPS was measured by means of computer-based testing as a transversal skill in the PISA 2012 cycle in over 40 countries worldwide (OECD, 2014). However, the focus of the PISA 2012 assessment was on overall CPS performance, and it focused only a little on the unique information found in CPS-based log files.

In an attempt to advance our understanding of the information found in log files, this study focused on CPS as a skill that requires students to engage in in-depth interactions with the tasks. Hence, CPS assessments produce particularly rich log files. Specifically, we investigated how three specific behavioral indicators (i.e., time on task, noninterfering observation, and high intervention frequency) affected the overall proficiency level on the two overarching CPS dimensions, knowledge acquisition and knowledge application. Additionally, we examined whether the resulting effects of these three specific behavioral indicators held after controlling for a more complex behavioral pattern that has already received considerable attention in prior CPS research; namely, students' use of VOTAT as a multistep exploration strategy (to be described below). We comprehensively investigated these questions in a large sample ($N = 1476$) of Finnish ninth-grade students.

1.1. Complex problem solving and its two dimensions

CPS has recently attracted a great deal of attention as an important marker of educational achievement when it was included in the 2012 cycle of the PISA survey (which was labeled *creative problem solving* in the PISA 2012 survey; OECD, 2014; see also Greiff, Holt, & Funke, 2013; Scherer & Gustafsson, 2015; for problem solving in PISA). In PISA 2012, 15-year-old students in over 40 countries across the globe had to demonstrate their levels of problem-solving proficiency when interacting with dynamically changing and intransparent problem environments that could not be solved with prior knowledge and that were contextually embedded in diverse contents from different domains (Greiff, Wüstenberg et al., 2014; OECD, 2014).

The necessity of engaging in an interaction with the problem situation through a sequence of actions and a targeted pattern of behaviors in order to gather information on the problem was stressed by the OECD's (2014) definition of problem-solving competence in general, which was the starting point for the conceptualization of CPS in PISA 2012. Accordingly, problem-solving competence can be described as "an individual's capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious" (p. 30). Buchner (in Frensch & Funke, 1995) provided a definition of CPS in particular, which underlay the operationalization of the employed CPS items in PISA 2012. Specifically, Buchner defined CPS as "the successful interaction with task environments that are dynamic (i.e., change as a function of user's intervention and/or as a function of time) and in which some, if not all, of the environment's regularities can only be revealed by successful exploration and integration of the information gained in that process" (p. 14).

Both the OECD's and Buchner's definitions stress the relevance of how students interact with the problem environment for their final CPS performance. CPS performance is usually reported separately for the two overarching CPS dimensions, knowledge acquisition and knowledge application. Knowledge acquisition describes how a mental representation of a problem's structure is established (Jonassen, 2011; Mayer & Wittrock, 2006; Wüstenberg et al., 2012), whereas knowledge application describes the process of actually using this knowledge in order to find a solution to a given problem (Fischer, Greiff, & Funke, 2012; Novick & Bassok, 2005). Even though highly correlated (i.e., a student who understands the structure behind a problem usually has a better chance of solving it correctly), the two dimensions are empirically separable (Schweizer, Wüstenberg, & Greiff, 2013; Wüstenberg et al., 2012). It is, however, unclear how a set of different behaviors precedes performance in these two dimensions and whether different behaviors impact the two dimensions in different ways.

Finding the underlying structure of a problem situation (i.e., knowledge acquisition) and engaging in a solution process (i.e., knowledge application) require a number of specific behaviors that subsequently cumulate in a more complex behavioral pattern. This, in turn, will lead to either a successful or an unsuccessful solution to the problem. In our Research Questions 1 and 2, we relate specific behaviors (i.e., time on task, noninterfering observation, and intervention frequency) and a well-established complex behavioral pattern (i.e., the use of the VOTAT strategy), respectively, to the two CPS dimensions of knowledge acquisition and knowledge application.

1.2. Research question 1: specific behaviors and their relations to overall performance in CPS

Among the few studies that have related specific behaviors to overall performance in cognitive assessments, most have focused on time on task as relevant behavioral indicator (e.g., Dodonova & Dodonov, 2012; Goldhammer et al., 2014; Kupiainen, Vainikainen, Marjanen, & Hautamäki, 2014). In this context, time on task refers to the time students spent in order to solve a task in a given cognitive assessment. Building upon this and other research, in the current study, we investigated how three different behavioral indicators that students might show to varying extents during CPS assessment are related to overall performance in knowledge acquisition and knowledge application, respectively: time on task (Research Question 1a), noninterfering observation (Research Question 1b), and intervention frequency (Research Question 1c).

1.2.1. Research question 1a: time on task and CPS performance

RQ1a *How is time on task related to the CPS dimensions of knowledge acquisition and knowledge application?*

Students differ substantially in the amount of time they spend on a task. Goldhammer et al. (2014) proposed a dual process theory framework that assumed that the relation between time on task and overall test performance was moderated by the type of