Computers & Education 91 (2015) 92–105

Contents lists available at ScienceDirect

Computers & Education

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

Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving

Samuel Greiff [a, *], Sascha Wüstenberg [a], Francesco Avvisati [b]

[a] University of Luxembourg, Luxembourg
[b] Organisation for Economic Co-operation and Development, France

ARTICLE INFO

Article history:
Received 8 April 2015
Received in revised form 14 July 2015
Accepted 20 October 2015
Available online 23 October 2015

Keywords:
Country-specific developments
Cross-cultural projects
Interactive learning environments
Simulations
Interdisciplinary projects

ABSTRACT

This paper aims at showcasing the potential of log-file analyses by capitalizing on the computer-based assessment of complex problem solving (CPS) in the 2012 cycle of the Programme for International Student Assessment (PISA). We analyzed log-file data from the CPS item *Climate Control* encompassing $N = 16{,}219$ students from 44 countries and economies. In Research Question 1, we related the application of an optimal exploration strategy (i.e., vary-one-thing-at-at-time, VOTAT) to performance on this specific problem and to overall level of CPS proficiency, both on an individual and on the country level. In Research Question 2, we identified several groups of students with different levels of non-mastery on a fine-grained level. Results indicated that (1) the VOTAT strategy was strongly related to performance on the *Climate Control* item as well as to overall problem solving proficiency, and (2) that there were different levels of non-mastery that ranged from applying no systematic strategic behavior to actually applying VOTAT but still failing to solve the item. On the backdrop of these results, we discuss implications and future potentials of log-file analyses in educational large-scale assessments for researchers, teachers, and policy makers.

1. Introduction

Ever since the advent of computers to psychological and educational assessment, the possibility of analyzing behavioral processes and sequences of actions through information captured in computer-generated log-files has been praised (Bunderson, Inouye, & Olsen, 1989). An asset of log-files is that they are a straightforward source of information containing data on every single behavioral action of each student (Bunderson et al., 1989; Greiff et al., 2014). As often as this potential has been praised, it has just as infrequently been exploited. In fact, almost 20 years after the initial excitement, Williamson, Bejar, and Mislevy (2006) state that surprisingly little use has been made of computer-generated log-files and that the exploitation of information contained in them has been lagging behind initial expectations.

The question of how log-files can be exploited to better understand students' levels of proficiency gained once again momentum when in the 2012 cycle of the Programme for International Student Assessment (PISA), the computer-based

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

http://dx.doi.org/10.1016/j.compedu.2015.10.018
0360-1315/

assessment of complex problem solving (CPS; labeled creative problem solving in PISA; OECD, 2014a) was included, PISA is among the internationally most widely recognized educational large-scale assessments (OECD, 2009b). Importantly, through its 2012 cycle it provides for the first time ever log-files containing the data of problem solving behavior from a representative sample of 15-year-old students in over 40 countries and economies that have now become available for further analyses.

The rationale behind including CPS in the PISA cycle in the first place is that CPS is considered a central 21st century skill of high importance for several outcomes including academic achievement (Sonnleitner, Keller, Martin, & Brunner, 2013; Wüstenberg, Greiff, & Funke, 2012; cf. also Wirth & Klieme (2003)). It encompasses a set of higher-order thinking skills that require students to engage in strategic planning, to think ahead, to carry out multi-step sequences of actions, to react to a dynamically changing system, to test hypotheses, and, if necessary, to adequately correct these hypotheses (Raven, 2000; Wüstenberg et al., 2012). In other words, it depends substantially on complex behavioral patterns and sequences of actions (Funke, 2010; Greiff et al., 2014).

Thus, CPS is an obvious candidate for in-depth analyses of the specific behaviors that underlie successful and unsuccessful performance. Information on exactly where and how students succeed or fail in their problem solving efforts is expected to equip researchers, teachers, and policy makers alike with important information about students' proficiency and about how to support them in optimizing their cognitive potential (cf. OECD, 2012). And yet, the exploitation of this rich resource through dedicated log-file analyses is still in its infancy. The study at hand was conducted to remedy this shortcoming by showcasing the potential of theoretically motivated log-file analyses by capitalizing on the PISA 2012 assessment of CPS. This assessment was comprised of several items in which students worked on novel and dynamic problem situations in computer-based assessment environments (OECD, 2014a). For our initial efforts, we chose one specific item from the dynamic CPS unit *Climate Control*. *Climate Control* offered a particularly good reflection of the underlying concept of CPS according to the CPS framework (OECD, 2014a) and was the focus of our initial efforts to understand students' behaviors when working on CPS tasks. A specific and detailed description of *Climate Control* is provided in section 2.2.

We targeted two research questions with regard to CPS. In the first research question, we investigated whether the specific strategic approach that students employed while solving the *Climate Control* item was related to their performance on this specific problem and to their overall level of CPS proficiency. In the second research question, we further exploited log-file data on a very fine-grained level in the attempt to identify several groups of students with different levels of mastery and non-mastery on the backdrop of their specific behaviors and their exploration strategies. To address these two research questions, we investigated individual log-files of the PISA 2012 cycle and used them for individual-level and country-level analyses.

1.1. Behavioral patterns and computer-generated log-files: an often-noted but seldom used potential

Nearly three decades ago, Bunderson et al. (1989) were among the first to highlight the potential of computer-based assessments that came along with a so-far unprecedented amount of data on students' test-taking behaviors. Indeed, by analyzing test-taking behaviors stored in log-files, not only can the final outcome be measured (i.e., correct or incorrect) but also the preceding steps and actions that resulted in the specific outcome. That is, instead of just answering *What* has been achieved, log-files provide direct insights into *How* the results were produced. More specifically, important information can be gained about how a student interacted with the problem and how this resulted in specific mastery or non-mastery of the task.

As they had this potential in mind, it is not surprising that Bunderson et al. (1989) asked scholars to put all hands on deck to fully exploit the almost infinite amount of data as a door opener to “intelligent measurement” in order to offer advice to learners and teachers on the basis of students' profiles and as a window into the black box of cognitive processing. Clearly, it is impossible to identify what happens in students' brains while they work on problems, but researchers can make theoretically motivated inferences about the underlying cognitive mechanisms from overt behavior. Being able to track the behavioral sequences of actions that lead to a specific performance outcome constitutes an important step toward understanding what happened during the test. For instance, log-files provide information about mistakes that were made while working on a problem, and such mistakes may be associated with an incorrect understanding of the problem situation (Ifenthaler, Eseryel, & Ge, 2012). Thus, in the long run, the identification of specific test-taking behavior might be a gold mine for any kind of teaching or intervention as it enables differentiated instruction that can provide students with individually tailored avenues for learning.

A closer look at the current practice of how log-files are exploited and how meaningful information is derived from them shows a remarkable discrepancy to the potential that is assigned to log-files: Theory-driven research on log-files has been stagnating for years, and Bunderson et al. (1989) vision of intelligent measurement has not come to fruition (Williamson, Mislevy, & Bejar, 2006). The lack of studies investigating process data is – at least to a certain extent – due to the challenges and obstacles associated with analyzing the often-fuzzy relations found in large amounts of log-file data. As noted by Csapó, Ainley, Bennett, Latour, and Law (2012), it is very difficult to “make sense of the hundreds of pieces of information students may produce when engaging in a complex assessment” (p. 216). That is, even though powerful techniques for analyzing log-file data such as educational data mining have recently emerged (Romero, Ventura, Pechenizkiy, & Baker, 2011), it is often difficult to give conceptual meaning to and to derive specific implications from the behavioral patterns found in log-files. Further, the technical expertise required for extracting meaningful pieces of information out of log-files (cf. Dumais, Jeffries, Russell, Tang, & Teevan, 2014) might further slow the rate of or even hinder widespread investigations. This is