Computers in Human Behavior 68 (2017) 513–519

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Home alone: Complex problem solving performance benefits from individual online assessment

Johannes Schult [a, *], Matthias Stadler [b, c], Nicolas Becker [a], Samuel Greiff [b], Jörn R. Sparfeldt [a]

[a] Saarland University, Germany
[b] University of Luxembourg, Luxembourg
[c] Universität Regensburg, Germany

ARTICLE INFO

Article history:
Received 21 December 2015
Received in revised form
23 November 2016
Accepted 28 November 2016

Keywords:
Complex problem solving
Computer-based assessment
Individual administration
Group testing
Online testing

ABSTRACT

Computer-based assessments of complex problem solving performance often take place in group settings like classrooms and computer laboratories. Such computer-based procedures provide an excellent opportunity to examine setting effects that might occur while participants are tested in a non-group session online at a time and place of their own choosing. For this purpose, $N = 273$ teacher students were randomly assigned to one of two settings: the individual online condition ($n = 216$) or the computer laboratory group condition ($n = 57$). Strong factorial measurement invariance was evidenced. Participants performed significantly better in the individual online condition than in the group condition (knowledge acquisition: $d = 0.38$; knowledge application: $d = 0.39$). The worse performance in the group setting compared to the individual setting could neither be explained by exploration time, nor by time on task. The internal experimental design validity strengthens the conclusion that setting-related differences in cognitive ability testing are not negligible but noteworthy.

1. Introduction

Computer-based assessments of cognitive abilities can take place in various settings ranging from university laboratories to classrooms to private homes. Research findings indicate that such assessments are not immune to effects of the testing environment (Birnbaum, 2004). Relevant setting characteristics include, among others: *location*, *presence of others*, and *schedule*. The choice of a specific location is based on its adequacy, efficiency, and convenience. For instance, psychological laboratories provide optimal control over environmental factors, but online testing can lead to higher response rates and lower costs (Birnbaum, 2004). Compared to individual testing, group administrations are useful in terms of economic efficiency and standardized procedures. However, the presence of others is associated with distractions that interfere with the task at hand (Becker, Koch, Schult, & Spinath, in press; Milgram & Milgram, 1976). Finally, online testing can be administered on demand around the clock (Dandurand, Shultz, & Onishi, 2008), whereas lab visits and group sessions have to be scheduled. At the time of testing, participants might be preoccupied with other things like subsequent appointments, and their commitment might be affected, as well. The present study investigates the influence of the testing environment (i.e., scheduled group administration in a computer laboratory vs. time-flexible individual online administration) for a fully computer-administered assessment of complex problem solving (CPS). Therefore, this experiment contributes to the understanding of how setting characteristics influence test performance.

1.1. Complex problem solving

An excellent framework for the study of setting effects can be found in CPS tasks. They offer an assessment of cognitive performance with a clear theoretical and psychometrical foundation. CPS can be defined as “the successful interaction with task environments that are dynamic (i.e., change as a function of the user's interventions 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” (Buchner in Frensch & Funke,

* Corresponding author. Bildungswissenschaften, Campus A5 4, Saarland University, 66123 Saarbrücken, Germany.
E-mail address: jutze@jutze.com (J. Schult).

http://dx.doi.org/10.1016/j.chb.2016.11.054
0747-5632/

1995, p. 14). Keeping in line with this definition, CPS tasks differ from static tasks, for example as used in intelligence tests, in that CPS tasks demand the performance of more complex mental processes than do static problem-solving tasks (Rigas & Brehmer, 1999). These include the active interaction with the problem to acquire knowledge in the problem environment, as well as the application of this knowledge to solve the problem (Novick & Bassok, 2005). To allow for active interaction between the problem solver and the problem, CPS tasks are by necessity computer-based (Greiff et al., 2013). This resulted in a vast array of research on computer-based assessments in the realm of CPS, such as the use of standardized instructions and the analysis of process data log files (Greiff, Wüstenberg, & Avvisati, 2015; Scherer, Greiff, & Hautamäki, 2015). The computer-based assessment can be administered in various settings without any adaptations. Therefore, the surface properties of the CPS tasks are identical in different settings.

When researching setting effects on CPS assessments, the process data availability makes it especially possible to go beyond the mere discovery of performance differences towards corresponding differences in specific behaviors. Such behaviors may include the exploration time used to gather knowledge about a complex problem or the total time spent on each task. These indicators are strongly related to CPS performance (Scherer et al., 2015) and thus, are viable prospects for explaining setting differences in computer-based test performance.

1.2. Setting-related differences in cognitive test performances

One important aspect characterizing different settings is the presence or absence of other people. Social facilitation theory (Bond & Titus, 1983; Strauss, 2002) suggests a stimulating rivalry between people who are carrying out an identical task. This should lead to increased activation and better performances in group settings than in individual settings. Still, social facilitation might only work with simple tasks. If the presence of others leads to an increased activation, dominant reactions should occur more frequently (Zajonc, 1965). For simple tasks, the dominant reaction should be the correct solution. For complex tasks, though, such higher activation should thereby increase the chance for selecting wrong solutions. Consequently, while working on more complex tasks as in CPS a higher CPS performance should occur in individual settings.

Thus far, computer-based CPS assessments were almost exclusively administered in group settings (e.g., classrooms). We know of only one study, in which CPS performance in a laboratory group setting was compared with the CPS performance in an individual online setting (Dandurand et al., 2008). Participants in the laboratory group were university students ($n = 63$), whereas the online subsample comprised 30 university students and 33 unspecified web users. All participants worked on a problem-solving task that involved estimating the weight of various animals by interactively experimenting with a pair of scales and several balancing weights. The interpretation of the results ($d = -0.44$ in favor of the laboratory group) can be questioned due to substantial dropout rates (79% for online participants) and the lack of a random assignment to the two conditions.

Given the scarcity of CPS-related setting comparisons mentioned above, intelligence research might offer further insights regarding setting effects in computer-based assessments. CPS and general intelligence reveal a substantial conceptual overlap and are empirically related (e.g., $r \approx .7$ in Lotz, Greiff, & Sparfeldt, 2016, mean weighted Hedges' $g = 0.43$ reported in Stadler, Becker, Gödker, Leutner, & Greiff, 2015).

In one study using a between-subject design (Moser, Schatz, Neidzwski, & Ott, 2011), a battery of cognitive tests was administered to high school athletes that were randomly assigned to one of two conditions (analyses with $n = 316$). The setting in the group condition was a high school computer laboratory, in which up to 20 participants took the test concurrently, whereas participants in the individual setting were tested in a neuropsychology clinic individually. Scores on all measures indicated better performance in the individual setting ("verbal memory": $d = 0.34$, "visual memory": $d = 0.41$, "motor processing speed": $d = 0.42$, "reaction time": $d = 0.38$). Moser et al. (2011) argued that the presence of other test takers could have been distracting because the tasks were novel and demanding. Another study using a figural matrices test in a quasi-experimental design (Ihme et al., 2009) found setting differences just the opposite. The performance in the laboratory group ($n = 57$ local psychology students) was similar to the performance in an individual online condition with $n = 212$ external psychology students ($d = 0.07$), yet better than another individual online condition with $n = 212$ unselected web users ($d = -0.46$). Those effects, however, disappeared ($|d| \leq 0.07$) after controlling for demographic differences due to the non-random assignment of the groups.

A series of experimental studies that compared individual and group administration of an intelligence test showed a similarly ambiguous picture. In one study (Scheffl, 2011), 154 school children aged 6 to 15 were tested under both conditions. One half (randomly chosen) of the sample began in the individual setting; the other half began in the group setting. The between-subject comparison of the initial assessment was mostly in favor of the individual condition across the five subtests that were administered: "general knowledge" ($d = 0.34$), "practical calculating" ($d = 0.97$), "finding synonyms" ($d = 0.23$, not statistically significant), "abstracting functions" ($d = 1.02$), and "social understanding and factual reflecting" ($d = 0.72$; effect sizes calculated from group A *individual setting* and group B *group setting*; Scheffl, 2011, p. 51). These particularly large effects were possibly due to insufficient time for supplementary instructions for slow learners in the group setting. In a similar study (Neustifter, 2011) with a different test of reasoning and spatial intelligence, there was no significant setting-difference ($d = 0.16$; $n = 111$ students aged 15 to 19). This null result resembles earlier comparisons of individual and group administrations (e.g., Farnsworth, 1928; cf. Becker et al., in press).

Flexibility regarding the time and place of computer-based administrations does not seem to be related to cognitive test performance. For example, Schreiner, Reiss, and Schweizer (2014) administered a computer-based working memory task in two different settings. Participants ($n = 211$) were tested individually in both settings. There were no significant differences between the laboratory and the more flexible online condition (reaction time: $d = 0.05$; accuracy: $d = -0.13$).

In summary, there are several studies reporting no setting differences, whereas some other studies provide evidence for better performance in individual settings. Likely causes for lower performance in group settings include insufficient instructions (Scheffl, 2011), distractions such as noise (Milgram & Milgram, 1976; Moser et al., 2011), and unfamiliar hardware and surroundings (Moser et al., 2011). However, studies so far have largely been limited to different facets of intelligence or to working memory, whereas studies that focus on abilities that require a longer sequence of actions in each task such as CPS are scarce at best. Furthermore, existing studies usually report mean differences between conditions, but there is little or no information provided on the behavioral antecedents (e.g., time on task) of overall performance differences. Using CPS with its aforementioned features and the possibility of focusing on process characteristics will provide new insights regarding the extent and possible explanations for