Intelligence 41 (2013) 579–596

Contents lists available at ScienceDirect

Intelligence

A multitrait–multimethod study of assessment instruments for complex problem solving☆

Samuel Greiff[a,*], Andreas Fischer[b], Sascha Wüstenberg[a], Philipp Sonnleitner[a], Martin Brunner[c], Romain Martin[a]

[a] University of Luxembourg, Luxembourg, Luxembourg
[b] University of Heidelberg, Heidelberg, Germany
[c] Freie Universität Berlin, Berlin, Germany

ARTICLE INFO

Article history:
Received 17 January 2013
Received in revised form 16 July 2013
Accepted 17 July 2013
Available online 23 August 2013

Keywords:
Complex Problem Solving
Reasoning
MicroDYN
Genetics Lab
MicroFIN

ABSTRACT

Recently published studies on Complex Problem Solving (CPS) suggest that assessments of CPS using multiple complex systems are only moderately related to tests of classical cognitive abilities. Further, CPS assessments show incremental validity beyond tests of other cognitive abilities when predicting relevant outcomes. However, these empirical accounts have relied on single CPS assessment instruments. We do not know whether these findings will generalize to the construct level across different CPS assessment instruments. To answer this question, we tested a sample of $N = 339$ German university students who completed three CPS assessment instruments based on multiple complex systems (MicroDYN, the Genetics Lab, and MicroFIN) and the matrices subtest of the Intelligence Structure Test as measure of reasoning. Students further reported their school grades. Analyses including latent multitrait–multimethod models provided support for the conceptualization of CPS as a complex cognitive ability. Results indicated that different CPS assessment instruments showed sufficient convergent validity (with a consistency mostly between .50 and .60). In addition, we found evidence for the divergent validity of CPS from reasoning (reasoning predicted two CPS facets, knowledge and control, $\beta_{KNOW} = .49$ and $\beta_{CON} = .53$, respectively). In the prediction of academic achievement, CPS explained variance in natural science grades after we controlled for reasoning ($\beta_{CPS} = .22$), whereas social science grades were not predicted. Our findings suggest that the validity of CPS generalizes across different measurement instruments.

1. Introduction

Across the last century, the relevance of cognitive abilities has been demonstrated numerous times, and the assessment of cognitive abilities has been a major concern in areas such as education, the economy, public health, and politics. Cognitive abilities have been shown to be related to outcomes such as longevity (Gottfredson & Deary, 2004), individuals' personality (Salas & Cannon-Bowers, 2001), job success (Schmidt & Hunter, 2004), or low crime rates (Herrnstein & Murray, 1994). Results of tests of cognitive performance have been used to promote the selection of students for higher education (Kuncel, Hezlett, & Ones, 2001), to allocate individuals to jobs according to their ability profiles (Autor, Levy, & Murnane, 2003), or to enhance

☆ This research was funded by grants supported by the Fonds National de la Recherche Luxembourg (ATTRACT “ASSKI21”) and by the German Federal Ministry of Education and Research (LSA004 and 01JG1062). We are grateful to the TBA group at DIPF (http://tba.dipf.de) for providing the authoring tool CBA Item Builder and technical support.
* Corresponding author at: University of Luxembourg, 6, rue Richard Coudenhove Kalergi, 1359 Luxembourg-Kirchberg, Luxembourg. Tel.: +352 466644 9245.
E-mail addresses: samuel.greiff@uni.lu (S. Greiff), andreas.fischer@psychologie.uni-heidelberg.de (A. Fischer), sascha.wuestenberg@uni.lu (S. Wüstenberg), philipp.sonnleitner@uni.lu (P. Sonnleitner), martin.brunner@isq-bb.de (M. Brunner), romain.martin@uni.lu (R. Martin).

0160-2896/$ – see front matter
http://dx.doi.org/10.1016/j.intell.2013.07.012

cognitive performance by teaching specific strategies (Klauer & Phye, 2008).

After long relying on paper–pencil tests, a shift toward computer-based assessments has recently been initiated. In its earliest implementation about two decades ago, the main purpose of computer-based assessment was to increase standardization and efficiency in testing (Baker & O'Neil, 2002). This practice produced several advantages such as automatic scoring and adaptive testing, but so far, the assessment instrument itself has been limited to a transformation of paper–pencil tests into computer-based tests (Bunderson, Inouye, & Olsen, 1989; Williamson, Bejar, & Mislevy, 2006).

However, if computerized testing is used simply to present computerized versions of paper–pencil tests, the advantages that computers can offer will not be fully utilized (Baker & O'Neil, 2002). That is, computers enable researchers to assess abilities that are not assessable by paper–pencil tests (Kyllonen, 2009) and to develop tasks that interactively respond to examinees' inputs. According to Williamson et al. (2006), the highest added value of using computers in assessment is expected from interactive tasks. Further, Rigas, Carling, and Brehmer (2002) have identified dynamic and interactive task environments as a general source of innovation in cognitive ability testing.

Two major advantages of computer-based assessment—higher efficiency and the inclusion of interactive tasks—were acknowledged by international large-scale assessments such as the Programme for International Student Assessment (PISA; OECD, 2006). In PISA, a major shift from paper–pencil to computer-based test administration was recently implemented, and complex interactive task environments were included in the current assessment cycle (OECD, 2010). For instance, an assessment of problem solving in interactive and dynamically changing task environments to assess Complex Problem Solving (CPS; Funke, 2001; Rigas et al., 2002) was part of the international PISA 2012 survey (OECD, 2010). However, given the short history of computer-based assessment, our knowledge about constructs such as CPS is limited. In this study, we therefore focused on CPS and important questions related to it.

CPS fits into the category of broad cognitive abilities (Funke, 2010), which are viewed as essential for lifelong learning by the OECD (2010). It is assessed in complex simulations (Funke, 2001) that allow for dynamic interactions between examinees and task situations (Raven, 2000; Wirth & Klieme, 2003). This feature makes it impossible to assess CPS without a computer. CPS is usually decomposed into a phase of knowledge acquisition (from here on: knowledge; actively acquiring knowledge about the task; Mayer & Wittrock, 2006) and knowledge application (from here on: control; actively controlling the task; Novick & Bassok, 2005). Recent research on CPS has shown divergent validity with regard to reasoning (e.g., Greiff, Holt, & Funke, 2013; Greiff, Wüstenberg, et al., 2013; Wüstenberg, Greiff, & Funke, 2012) and working memory (e.g., Schweizer, Wüstenberg, & Greiff, 2013) as well as the predictive validity of CPS beyond other cognitive abilities (e.g., Greiff & Fischer, 2013; Wüstenberg et al., 2012). After some controversy with regard to its assessment (e.g., Kröner, Plass, & Leutner, 2005; Wüstenberg et al., 2012), CPS has recently experienced advances in terms of its scalability and psychometric properties. These assessment advances were substantially facilitated by the introduction of two formal frameworks—linear structural equations and finite state automata—and by the introduction of multiple complex systems (MCS; see below; Funke, 2010; Greiff, Wüstenberg, & Funke, 2012).

However, it is not quite appropriate to talk about the construct of CPS when describing these recent results. In fact, the results that we mentioned above were conducted only with single homogenous CPS assessment instruments. If we want to generalize these results to CPS as a construct independent of a particular measurement procedure, we need to use a variety of assessment instruments to measure CPS (Campbell & Fiske, 1959; Eid, Lischetzke, & Nussbeck, 2006).

In order to facilitate our understanding of CPS not only on the level of specific assessment instruments but on the construct level, we applied a combination of different CPS assessments instruments. Specifically, we employed three CPS instruments based on multiple complex systems (Greiff et al., 2012) to address (a) the convergent validity of these instruments by combining them in a multitrait–multimethod (MTMM) approach and (b) their divergent validity by relating CPS on the construct level to reasoning and to academic achievement. To this end, we will first outline the conceptual background behind these two research questions and continue with our presentation of empirical studies. We will conclude by discussing the relevance of CPS and its implications for research on cognitive abilities.

1.1. Research Question 1: Measurement of CPS by different assessment instruments

According to Baker and O'Neil (2002), CPS amplifies the learning of children and adults in a number of formal and informal settings. Further, Mayer and Wittrock (2006) point to the importance of CPS in educational settings aimed at making students better problem solvers. To this end, the OECD (2010) views CPS as a complex cognitive ability that has the interaction between task and examinee (and, thus, computer-based assessment) as a central component. Buchner (1995) defines 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).

In line with this definition, Funke (2001) and Raven (2000) argue that solving complex problems involves a series of complex cognitive operations, and that complex problems can be described by several characteristic features such as complexity, intransparency, interconnectedness, and dynamics. Coping with complex problems further involves monitoring (Osman, 2010) and learning (Leutner, 2002). It requires knowledge about when and how to