Educational Research and Innovation

The Nature of Problem Solving

USING RESEARCH TO INSPIRE 21ST CENTURY LEARNING

Edited by Benő Csapó and Joachim Funke

Centre for Educational Research and Innovation

OECD

Educational Research and Innovation

THE NATURE OF PROBLEM SOLVING

USING RESEARCH TO INSPIRE 21ST CENTURY LEARNING

Edited by Benő Csapó and Joachim Funke

OECD
BETTER POLICIES FOR BETTER LIVES

Please cite this publication as:
Csapó, B. and J. Funke (eds.) (2017), The Nature of Problem Solving: Using Research to Inspire 21st Century Learning, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264273955-en

ISSN: 2076-9660 (print)
ISSN: 2076-9679 (online)

Photo credits:
Fotolia/alphaspirit

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

Foreword

The demands on learners and thus education systems are evolving fast. In the past, education was about teaching people something. Now, it's about making sure that students develop a reliable compass and the navigation skills to find their own way through an increasingly uncertain, volatile and ambiguous world. These days, we no longer know exactly how things will unfold, often we are surprised and need to learn from the extraordinary, and sometimes we make mistakes along the way. And it will often be the mistakes and failures, when properly understood, that create the context for learning and growth. A generation ago, teachers could expect that what they taught would last a lifetime for their students. Today, teachers need to prepare students for more rapid economic and social change than ever before, for jobs that have not yet been created, to use technologies that have not yet been invented, and to solve social problems that we don't yet know will arise.

The dilemma for educators is that the kind of skills that are easiest to teach and easiest to test, are also the skills that are easiest to digitise, automate and outsource. There is no question that state-of-the-art disciplinary knowledge will always remain necessary. Innovative or creative people generally have specialised skills in a field of knowledge or a practice. And as much as "learning to learn" skills are important, we always learn by learning something. However, success in life and work is no longer mainly about reproducing content knowledge, but about extrapolating from what we know and applying that knowledge in novel situations. Put simply, the world no longer rewards people just for what they know – Google knows everything – but for what they can do with what they know. Problem solving is at the heart of this, the capacity of an individual to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious.

Conventionally our approach to problems in schooling is to break them down into manageable pieces, and then to teach students the techniques to solve them. But today individuals create value by synthesising disparate parts. This is about curiosity, open-mindedness, making connections between ideas that previously seemed unrelated, which requires being familiar with and receptive to knowledge in other fields than our own. If we spend our whole life in a silo of a single discipline, we will not gain the imaginative skills to connect the dots, which is where the next invention will come from.

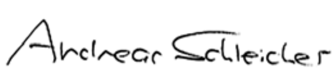
Perhaps most importantly, in today's schools, students typically learn individually and at the end of the school year, we certify their individual achievements. But the more interdependent the world becomes, the more we rely on great collaborators and orchestrators who are able to join others to collaboratively solve problems in life, work and citizenship. Innovation, too, is now rarely the product of individuals working in isolation but an outcome of how we mobilise, share and link knowledge. So schools now need to prepare students for a world in which many people need to collaborate with people of diverse cultural origins, and appreciate different ideas, perspectives and values; a world in which people need to decide how to trust and collaborate across such differences; and a world in which their lives will be affected by issues that transcend national boundaries. Expressed differently, schools need to drive a shift from a world where knowledge is stacked up somewhere depreciating rapidly in value towards a world in which the enriching power of collaborative problem-solving activities is increasing.

These shifts in the demand for knowledge and skills are well understood and documented, and to some extent they are even intuitive. Not least, many school curricula highlight the importance of individual and social problem-solving skills. And yet, surprisingly little is known about the extent to which education systems deliver on these skills. This is not just because school subjects continue to be shaped by traditional disciplinary contexts. It is also because educators have few reliable metrics to observe the problem-solving skills of their students - and what doesn't get assessed doesn't get done.

The OECD Programme for International Student Assessment (PISA) sought to address this. Its 2012 assessment contained the first international metric of individual problem-solving skills and the 2015 assessment took this further, assessing collaborative problem-solving skills. The results turned out to be extremely interesting, in part because they showed that strong problem-solving skills are not an automatic by product of strong disciplinary knowledge and skills. For example, Korea and Japan, which both did very well on the PISA mathematics test, came out even stronger on the PISA assessment of problem-solving skills. In contrast, top mathematics performer Shanghai did relatively less well in problem solving. Such results suggest that it is worth educators devoting more attention to how problem-solving skills are developed both in disciplinary and cross-disciplinary contexts.

But while problem solving is a fairly intuitive and all-pervasive concept, what has been missing so far is a strong conceptual and methodological basis for the definition, operationalisation and measurement of such skills. This book fills that gap. It explores the structure of the problem-solving domain, examines the conceptual underpinning of the PISA assessment of problem solving and studies empirical results. Equally important, it lays out methodological avenues for a deeper analysis of the assessment results, including the study of specific problem-solving strategies through log-file data.

In doing so, the book provides experts and practitioners with the tools to better understand the nature of problem-solving skills but also with a foundation to translate advanced analyses into new pedagogies to foster better problem-solving skills.



Andreas Schleicher

Director for Education and Skills
Special Advisor to the Secretary-General

Acknowledgements

Thanks to all the people who contributed to this book and helped to finalise a process that took much longer than expected! Special thanks to Julia Karl (Heidelberg) who helped us preparing a standardised and printable version of all manuscripts. And thanks to Marion Lammarsch (Heidelberg) for help with converting text from LaTex to Word. Thanks to the OECD staff, in particular Francesco Avvisati, Sophie Limoges and Rachel Linden, for their valuable support during the production process, and to Sally Hinchcliffe for the thoughtful language editing of the manuscript.

This book stems mainly from the collaboration of the members of the OECD Problem Solving Expert Group (PEG) of PISA 2012. This PEG group started its works in 2009 and finished their official work in 2014. The group consisted of the following eight members:

- Joachim Funke (Chair), Heidelberg University, Germany
- Benő Csapó, University of Szeged, Hungary (ex officio PGB representative)
- John Dossey, Illinois State University, United States
- Art Graesser, University of Memphis, United States
- Detlev Leutner, Duisburg-Essen University, Germany
- Richard Mayer, University of California, United States
- Tan Ming Ming, Ministry of Education, Singapore
- Romain Martin, University of Luxembourg, Luxembourg

Members of the PEG group have already been involved with problem solving for a long time, and their meetings under the umbrella of the PISA work inspired a number of other meetings and co-operative studies involving people from other organisations and institutions. Influenced by the creative atmosphere of the PEG meetings, some of the members met and presented their work together at other professional meetings as well. Amongst these were the annual meetings at Szeged University in the framework of Szeged Workshop of Educational Evaluation (SWEE; since 2009 a yearly repeated event), the TAO days in Luxemburg (March, 2011), the AERA meeting in New Orleans (April, 2011), two symposia at the EARLI biennial meeting in Exeter (September, 2011), the European Conference of Psychology in Istanbul (July, 2011), two symposia at the International Conference on Psychology in Cape Town (July, 2012), and more recently the "Celebrating Problem Solving" conference at the University of Szeged (November 2015). Many related journal articles have been published in the meantime – too many to be listed here.

These productive meetings brought together researchers interested in problem solving, assessment of cognitive skills, and technology based assessment, and so initiated empirical works in the overlapping areas of these special fields. For example, as already mentioned, one of

the major shifts from PISA 2003 problem solving to PISA 2012 problem solving was the shift from paper-and-pencil based to computer-based assessment that required strong interactions between item developers and the group taking care of the technical implementation. Within a rather short time scale, software tools had to be developed and implemented that allow for the necessities in international large-scale assessment studies (e.g. preparing for more than 100 different languages, different character sets including left-to-right and right-to-left, and different levels of computer equipment).

The PEG group was supported by a wonderful team from ACER (Australian Council for Educational Research, Melbourne, Australia): the “trio” consisting of Barry McCrae, Ray Philpot, and Dara Ramalingam. They prepared meetings and materials in a fantastic way and helped us through a jungle of dates, deadlines, and data. Ray Adams worked as Interim Chair in the beginning of the project. All of this contributed to the success of PISA 2012.

Maybe unique in the history of PISA expert groups, this community of researchers, while developing the assessment framework and creating the instruments discovered the potentials of an emerging field: the possibilities offered by computerised, dynamic, interactive assessment of problem solving. Using multimedia and simulation to present the test tasks, capturing students’ responses in novel ways, logging student’s activities and using log-file analyses for exploring cognitive processes, perseverance and motivation have opened new and exciting directions for research. They have been continuing their collaboration far beyond their task in the 2012 assessment cycle.

The individual chapters in this book have been reviewed by members of the group and by reviewers from the OECD. This process hopefully helped to improve the quality of the chapters. At the same time, these activities delayed the publication process a bit.

Lastly we would like to extend our thanks to Andreas Schleicher for his support throughout the publication process.

Table of contents

Figures

Tables

Chapter 8

Empirical study of computer-based assessment of domain-general complex problem-solving skills

By

Gyöngyvér Molnár

MTA-SZTE Research Group on the Development of Competencies, University of Szeged, Hungary.

Samuel Greiff

Maison des Sciences Humaines, Université du Luxembourg, Luxembourg.

Sascha Wüstenberg

TWT GmbH Science & Innovation, Germany.

and Andreas Fischer

Forschungsinstitut Betriebliche Bildung (f-bb), Nuremberg, Germany.

This study reviews the results of a recent project on problem solving. Taking a developmental and structural perspective, it contrasts static, paper-and-pencil tests with interactive, technology-based tests of thinking skills, with a special reference to reasoning skills including knowledge acquisition, knowledge application, and transfer of knowledge. Hungarian students aged 11 to 17 completed problem-solving tests in static scenarios (assessing domain-specific problem-solving skills from maths and science) and in interactive scenarios (assessing domain-general complex problem-solving skills). The students were also assessed for inductive reasoning and fluid intelligence, both by first generation tests. This chapter uses the results to elicit evidence for the development of dynamic problem solving and for the relationship between those 21st-century skills. Finally, it discusses the possibility of using traditional static tests to predict performance in third generation tests measuring dynamic problem solving.

This research was funded by the European Union (Project ID number: TÁMOP–3.1.9–11/1–2012–0001), the SZTE-MTA Research Group on the Development of Competencies, the K115497 OTKA research program, and by grants from the Fonds National de la Recherche Luxembourg (ATTRACT “ASKI21”) and the German Federal Ministry of Education and Research (BMBF FKZ 01JG1062). – Some of the analyses and figures as well as minor text parts presented in this article are reiterations of previous work and are taken directly from there (for instance, from Molnár, Greiff, & Csapó, 2013). At the same time, much of the theoretical thoughts and the analyses are extensions and are unique contributions of this chapter and have not been published elsewhere. Concerning terminology, please note that there exist different labels for the current subject (e.g., complex problem solving, dynamic problem solving, creative problem solving). In this chapter we use the prefix “dynamic”.

Introduction

How can we assess knowledge and skills essential in the 21st century? Can we predict students' performance in so-called "third generation" tests that measure, for instance, dynamic problem solving skills, from their performance in traditional, static and more academic testing situations (so-called "first generation" tests)? How do problem-solving skills develop over time, especially the skills involved in coping with interactive and dynamically changing problems? Are students ready for third generation testing or do they prefer first and second generation tests?

The assessment of 21st-century skills has to provide students with an opportunity to demonstrate their skills related to the acquisition and application of knowledge in new and unknown problem situations. These are the skills needed in today's society, characterised as it is by rapid change, where the nature of applicable knowledge changes frequently and specific content becomes quickly outdated (de Koning, 2000). There are three ways to assess 21st-century skills. First, they can be measured through traditional approaches using first generation tests – with designs based closely on existing static paper-and-pencil tests, such as the typical tests of domain-specific problem-solving-skills. Second, they can be measured with second generation tests using new formats, including multimedia, constructed response, automatic item generation and automatic scoring tests (Pachler et al., 2010). Finally, they can be assessed through third generation tests which allow students to interact with complex simulations and dynamically changing items, dramatically increasing the number of ways they can demonstrate skills such as the domain-general dynamic problem solving-skills in the study presented below. All three approaches are relevant to avoid methodological artefacts, since measurement of a thinking skill always involves reference to general mental processes independent of the context used (Ericson and Hastie, 1994).

In this chapter we take a developmental and a structural perspective to elaborate on technology-based assessment, contrasting static, paper-and-pencil tests with interactive, technology-based tests of thinking skills, with special reference to reasoning skills including knowledge acquisition, knowledge application and transfer of knowledge. We synthesise research related to identifying 21st-century skills, including these skills, both in static tests assessing domain-specific problem solving (DSPS) and in interactive scenarios assessing domain-general dynamic problem solving (DPS). We elicit evidence for the development of these domain-general 21st-century skills measured by a third generation test, and for the relationship between those skills and more domain-specific static problem-solving skills measured by first generation tests, inductive reasoning (IR) and fluid intelligence, which is considered a "hallmark" indicator of the general *g* factor (intelligence), both of which are also measured by first generation tests. Finally, we discuss the possibility of predicting performance in third generation tests of DPS from performance in traditional, static testing situations.

Technology-based assessment and new areas of educational assessment

Information and communication technologies have fundamentally changed the possibilities and the process of educational assessment. Research and development in technology-based assessment (TBA) go back three decades. In the 1990s the focus was on exploring the applicability of a broad range of technologies from the most common, widely available computers to the most expensive, cutting-edge technologies for assessment purposes (Baker and Mayer, 1999). A decade later, large-scale international assessments were conducted to explore the potential and the implementation of TBA, such as the National Assessment of Educational Progress (NAEP), the Programme for International Student Assessment (PISA), and the Progress in International Reading Literacy Study (Csapó et al., 2012; Fraillon et al., 2013; OECD, 2010, 2014a) with the aim of replacing traditional paper-and-pencil test delivery with assessments exploiting the manifold advantages of computer-based delivery.

The predominant subject of these studies was to compare the results of paper-and-pencil and computer-based assessments of the same construct (Kingston, 2009; Wang et al., 2008). The instruments studied were mostly first and third generation tests (Pachler et al., 2010). As technology has become more ubiquitous over the past decades, familiarity with computers (Mayrath, Clarke-Midura and Robinson, 2012) and test mode effects should no longer be much of an issue (Way, Davis and Fitzpatrick, 2006). In the last five years computer-based assessment has opened doors to exploring features not yet studied, such as domain-general dynamic problem-solving skills. This required the development of third generation tests that did not rely on standard, multiple-choice item formats (Ripley et al., 2009; Greiff, Wüstenberg and Funke, 2012; Greiff et al., 2014; OECD, 2014b). Finally, a relatively recent development is the online assessment of group processes in place of individual assessments, for example projects by the Assessment and Teaching of 21st Century Skills (ATC21S); Griffin, McGaw and Care (2012); PISA; National Assessment and Testing; Hesse et al. (2015); and the OECD (2013).

From static to dynamic problem solving with reference to reasoning skills

Reasoning is one of the most general thinking skills (Pellegrino and Glaser, 1982; Ropo, 1987) and often understood as a generalised capability to acquire, apply and transfer knowledge. It is related to, and the strongest component of, almost all higher-order cognitive skills and processes (Csapó, 1997), such as general intelligence (Klauer and Phye, 2008) and problem solving (Gentner, 1989; Klauer, 1996). In connection with reasoning, problem-solving skills have been extensively studied from different perspectives over the past decade, as they involve the ability to acquire and use new knowledge, or to use pre-existing knowledge in order to solve novel (i.e. non-routine) problems (Sternberg, 1994). Studies into the different approaches have showed the need to distinguish between domain-specific and domain-general problem-solving skills (Sternberg, 1995).

One of the arguably most comprehensive international large-scale assessments, the PISA survey, places special emphasis on DSPS and DPS processes. It measured DSPS in 2003 and DPS in 2012 as an additional domain beyond the usual fields of reading, science and mathematics (OECD, 2005, 2010, 2013; Greiff, Holt and Funke, 2013; Fischer et al., 2015). Here, problem solving was seen as a "cross-curricular" domain.

In DSPS situations, problem solvers need to combine knowledge acquired in and out of the classroom to reach the desired solution by retrieving and applying previously acquired knowledge in a specific domain. In this study, we treat DSPS as a process of applying domain-specific – mathematical and scientific – knowledge in three different types of new situations: 1) complete problems, where all necessary information to solve the problem is given at the outset; 2) incomplete problems relying on missing information that students are expected to have learnt at school; and 3) incomplete problems relying on missing information that was not learnt at school (see Molnár, Greiff and Csapó, 2013).

In contrast, DPS tasks require an additional series of complex cognitive operations (Funke, 2010; Raven, 2000) beyond those involved in DSPS tasks (e.g., Klieme, 2004; Wüstenberg, Greiff and Funke, 2012). In DPS tasks, students have to directly interact with a problem situation which is dynamically changing over time and use the feedback provided by the computer to acquire and apply new knowledge (Fischer, Greiff and Funke, 2012; Funke, 2014). This measures the competencies of knowledge acquisition and knowledge application.

Recent analyses provide evidence for the relation between domain-specific static and domain general-dynamic problem solving, especially between the acquisition and application of academic and non-academic knowledge; and their relation to general mental abilities.

Aims

The objective of the study described here is to review the results of a recent project on DPS (see Greiff et al., 2013; Molnár, Greiff and Csapó, 2013a, 2013b; Wüstenberg et al., 2012). We thus intend to answer five research questions:

1) Can DPS be better modelled as a two-dimensional construct with knowledge acquisition and knowledge application as separate factors than as a one-dimensional construct with both dimensions subsumed under one factor?

2) Can DPS be shown to be invariant across different grades implying that differences in DPS performance can be validly interpreted?

3) How does DPS develop over time during public education in different school types?

4) What is the relationship between DPS, DSPS and IR and do these relationships change over time?

5) Can we predict achievement in DPS (measured by a third generation test) by performance in more traditional testing situations (DSPS, IR and intelligence) assessing thinking skills including problem solving?

Methods

Sample

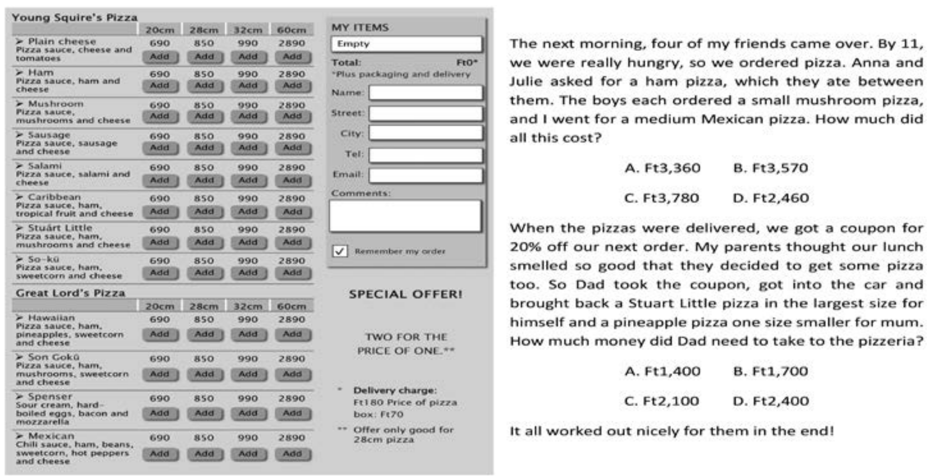
The samples of the study were drawn from 5th to 11th grade students (aged 11 to 17) in Hungarian primary and secondary schools. There were 300 to 400 students in each cohort. One-third of the secondary school students (9th to 11th grade) were from grammar schools, with the rest from vocational schools. Some technical problems occurred during online testing resulting in completely random data loss. Participants who had more than 50% of data missing (316 students) were excluded from the analyses. The final sample for the analyses contained data from 788 students.[1]

Instruments

The instruments of the study include tests of DPS, DSPS, IR, fluid reasoning (an important indicator of intelligence) and a background questionnaire. One comprehensive version of the DPS test was used regardless of grade. The test consisted of seven tasks created following the MicroDYN approach (see Greiff, et al., 2013; Chapter 6). In the first stage, participants were provided with instructions including a trial task. Subsequently, students had to explore several unfamiliar systems, in which they had to find out how variables were interconnected and draw their conclusions in a situational model (knowledge acquisition; Funke, 2001). In the final stage, they had to control the system by reaching given target values (knowledge application; see Greiff et al., 2013).

Three versions of the DSPS test were used with different levels of item difficulty, which varied by school grade. The test versions contained anchor items allowing performance scores to be represented on a single scale in each case. Approximately 80% of the 54 DSPS items were in multiple-choice format, the remaining items were constructed-response questions and all of the problems presented information in realistic formats. The DSPS test comprised three types of problems: 1) problems where all the information needed to solve the problem was given at the outset (knowledge application); 2) incomplete problems requiring the use of additional information previously learnt at school as part of the National Core Curriculum (knowledge application and knowledge transfer); and finally 3) incomplete problems requiring the use of additional information that had not been learnt at school and needed to be retrieved from real-life knowledge (knowledge application, knowledge transfer and knowledge creation; see Molnár et al., 2013a). The presentation of the DSPS problems looked similar in all three versions of the test. The left-hand column presented information in realistic formats (such as a map, picture or drawing) and on the right was a story of a family trip or a class excursion and a prompt students to solve problems (e.g., using the information provided and supplementing it with school knowledge) as they would arise during the trip" (Figure 8.1). All of the problems needed domain-specific knowledge from the field of mathematics or science to solve.

Figure 8.1 **Example of tasks in the domain-specific problem-solving test**



Source: Molnár et al. (2013a), "Inductive reasoning, domain specific and complex problem solving: Relations and development".

The IR test comprised both open-ended and multiple-choice items. It was divided into three subtests: number analogies (14 items) and number series (16 items) embedded in mathematical contexts, and verbal analogies (28 items; see Csapó, 1997; Figure 8.2).

The Culture Fair Test 20-R (CFT) test was used for measuring students' fluid intelligence, which is according to state-of-the-art intelligence theories such as the Cattell–Horn–Carrolltheory one of the most important indicators and markers of g (Weiß, 2006), the core of intelligence (Carroll, 2003). It consisted of 4 subscales with 56 figural items.

The background questionnaire contained questions regarding students' socio-economic background, academic achievement, school subject attitudes and parental education.

Figure 8.2 **Example of tasks in the inductive reasoning test**

Verbal analogies	CHAIR : FURNITURE = DOG : ? a CAT b ANIMAL C SPANIEL d TABLE e DOGHOUSE
Number analogies	20→32 :: 8→20 :: 11→____

Note: The original items were in Hungarian.

Source: Molnár et al. (2013a), "Inductive reasoning, domain specific and complex problem solving: Relations and development".

Procedure

The tests were completed in four sessions, each lasting approximately 45 minutes. In Session 1, students worked on the DPS test. In Session 2 students had to complete the DSPS test, in Session 3 they completed the intelligence test (CFT) and in Session 4 an IR test and the background questionnaire.

The online data collection was carried out by means of the TAO platform over the Internet. The testing took place in the computer labs of the participating schools, using existing computers and preinstalled browsers.

Confirmatory factor analyses (CFA) were applied to test the underlying measurement model of DPS (research question 1). A weighted least squares mean and variance adjusted (WLSMV) estimator and theta parameterisation were used to estimate model parameters, since all items were scored dichotomously (Muthén and Muthén, 2010). Tucker-Lewis Index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA) were proposed to assist in determining model fit (see Vandenberg and Lance, 2000). Nested model comparisons were conducted using the DIFFTEST procedure (Muthén and Muthén, 2010). All measurement models were computed with Mplus (Muthén and Muthén, 2010).

Measurement invariance was tested by multigroup analyses (MACS) within the structural equation modelling (SEM) approach (research question 2). The testing procedure for categorical data involves a fixed sequence of model comparisons (Vandenberg and Lance, 2000), testing three different levels of invariance (configural invariance, strong factorial invariance and strict factorial invariance) by comparing measurement models from the least to the most restrictive model (for more detail see Greiff et al., 2013).

Rasch's model was used for scaling the data, and then linear transformation of the logit metric was chosen. The means of 8th graders were set to 500 with a standard deviation of 100 (research question 3). A four-parameter logistic equation was used for the curve fitting procedures to estimate development. A coefficient of determination (R^2) was computed to express how well the model described the data (see Molnár et al., 2013a). SEM was also used to examine the direct and indirect effects between DPS, DSPS, IR and intelligence (research questions 4 and 5).

Results

Descriptive statistics

Internal consistencies across the DPS, DSPS, IR and CFT tests were generally high (DPSα = .92; $DSPS_{levels1to3}$ α = .73, .82, .65; IRα = .95; CFTα = .88, respectively; see Molnár et. al., 2013a and Greiff et al., 2013). However, there was a noticeable drop in reliability in the DSPS test used in grades 9 to 11 (Level 3) down to .65. Grade-level analyses reveal increased probability of measurement error among 9th and 10th grade students, whereas the reliability of the same test among 11th grade students ($DSPS_{grade11}\alpha$ = .72) proved to be higher (see Molnár et al., 2013a). For this reason, we included only the data for 5th to 8th grade and 11th grade students in all further analyses of DSPS.

Dimensionality of dynamic problem solving

The two-dimensional model of DPS measured by MicroDYN includes knowledge acquisition and knowledge application as separate factors. This was compared to a one-dimensional model that combined both dimensions under one general factor. In accord with our theoretical hypothesis, a two-dimensional model of DPS showed a better fit than a one-dimensional one (χ^2-difference test: χ^2 = 86.121; df = 1; $p < .001$). Fit indices were significantly better in the two-dimensional model (Table 8.1; see Greiff et al., 2013). In summary, this provided empirical support for the theoretically derived two-dimensional model of knowledge acquisition and knowledge application in DPS.

Table 8.1 **Goodness of fit indices for testing dimensionality of the dynamic problem solving model**

Model	χ^2	df	p	CFI	TLI	RMSEA
2 dimensional model	164.06	53	.001	.967	.978	.050
1 dimensional model	329.35	52	.001	.912	.944	.079

Note: df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker-Lewis Index.

Measurement invariance of the domain-general dynamic problem-solving instrument

In order to test measurement invariance of DPS, we tested configural invariance, strong factorial invariance and strict factorial invariance and compared the measurement models. All measurement models yielded a good fit as indicated by CFI, TLI and RMSEA (Table 8.2; CFI, TLI > .95; RMSEA < .06).

Neither the strong factorial invariance nor the strict factorial invariance models differed significantly from the configural invariance model, implying that measurement invariance was maintained (ΔCFI < .01 and non-significant χ^2-difference test, Table 2).

Table 8.2 **Goodness of fit indices for measurement invariance of DPS in the MicroDYN approach**

Model	χ^2	df	$\Delta\chi^2$	Δdf	p	CFI	TLI	RMSEA
Configural inv.	161.04	104	–	–	–	.975	.975	.051
Strong fact.inv.	170.10	115	22.29	23	>.10	.976	.982	.047
Strict fact.inv.	165.82	116	53.15	43	>.10	.978	.983	.045

Note: fact.inv. = factorial invariance; df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker-Lewis Index. Nested model comparisons (computing $\Delta\chi^2$) were conducted using the DIFFTEST procedure, a special $\Delta\chi^2$ difference test (see Muthén and Muthén, 2010).

As the DPS was measured as invariant, mean differences across groups attending different grades could be interpreted as true differences in the underlying DPS skills and were not due to psychometrical issues (Byrne and Stewart, 2006). This allowed us to make valid direct comparisons of skill levels between students and between classes and investigate the development of DPS in research question 3.

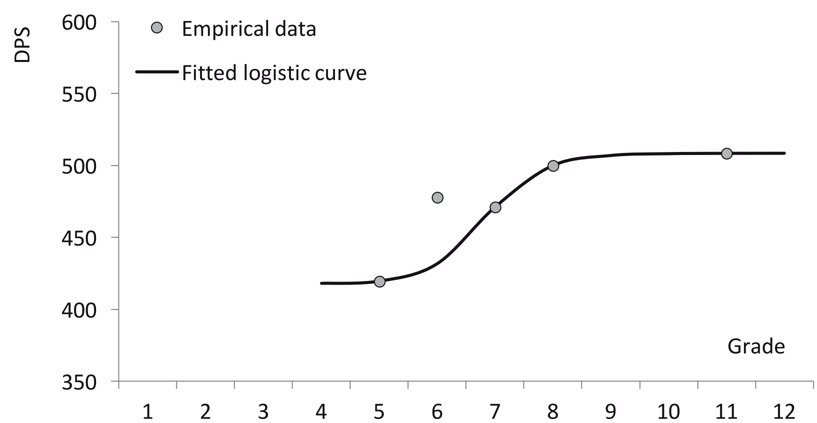
Development of domain-general dynamic problem solving

The features of developmental tendencies of DPS were in line with the findings of previous studies regarding thinking skills, (e.g., Adey et al., 2007) such as problem solving (Molnár et al., 2013a), and follow a regular developmental trend. The development spans several years and can be described with a logistic curve that fitted the empirical data adequately (R^2 = .91; see Figure 8.3). The fit was perfect (R^2 = 1.00) if we excluded 6th grade students from the analyses. The behaviour of 6th grade students calls for further study because the currently available empirical data do not account for this phenomenon.

The fastest growth, the point of inflexion, was observed in 7th grade (at the age of 12.8), offering opportunities for training in DPS. This is the sensitive period for stimulation, since the enhancement of thinking skills is most effective when “students’ development is still in progress, especially when they are in a fast-growing phase” (Molnár et al., 2013a). All in all, elementary school students from 5th to 8th grade showed noticeable development in DPS; fostering DPS could be very efficient during this period. This trend seemed to change after 8th grade, when development slowed down. However, extrapolation of the fitted logistic curves indicated that substantial development took place before 3rd grade and some improvement can also be expected after 11th grade (Figure 8.3).

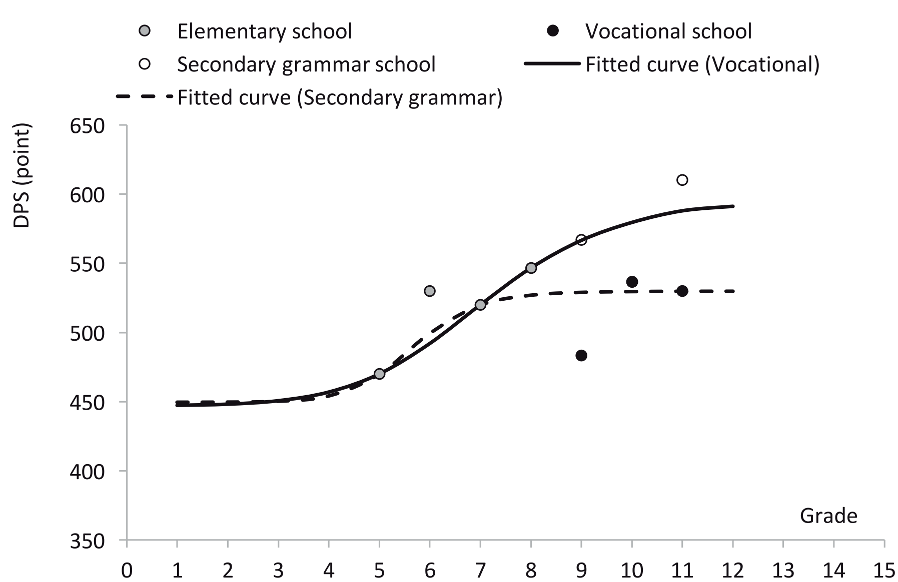
Looking at the analyses of the different school types, the coefficient of determination decreased in the case of vocational school ($R^2 = .72$) and increased ($R^2 = .96$) for grammar school data. The slope and maximum of the developmental curve changed notably after primary school. Even the 9th graders in grammar schools performed better than 11th graders in vocational schools, and 11th graders in vocational schools performed worse than 8th graders in elementary schools. The performance differences ($t = -8.59$, $p < .01$) between vocational and grammar school students proved to be stable over time (Figure 8.4; Molnár et al., 2013b).

Figure 8.3 **Developmental curve of dynamic problem solving**



Source: Molnár, G., S. Greiff and B. Csapó (2013b), "Relations between problem solving, intelligence and socio-economic background", paper presented at the EARLI, Munich, 27-31 August 2013.

Figure 8.4 **Developmental curve of dynamic problem solving by school type**



Source: Molnár, G., S. Greiff and B. Csapó (2013b), "Relations between problem solving, intelligence and socio-economic background", paper presented at the EARLI, Munich, 27-31 August 2013.

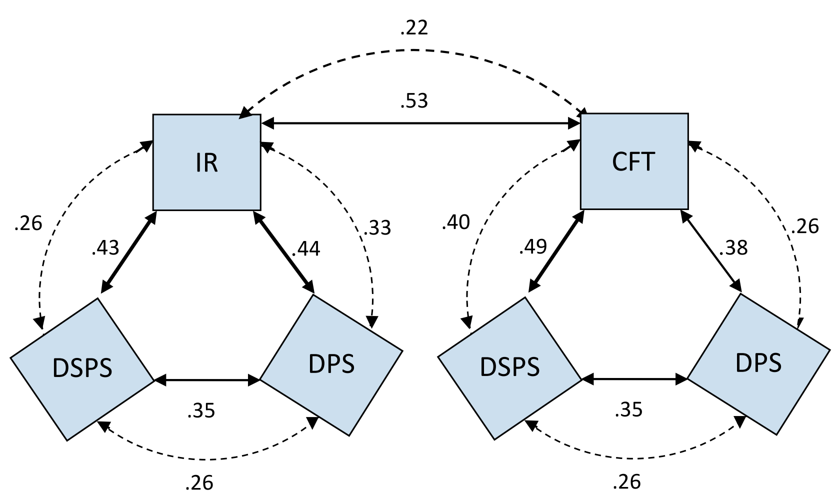
The relationship between inductive reasoning, intelligence, domain-specific and domain-general dynamic problem solving

The bivariate correlations between inductive reasoning, DSPS and DPS were similar to those between intelligence, DSPS and DPS. All of them were moderate ranging from .35 to .49 (Figure 8.5). The relationships proved to be similar between IR and either DSPS or DPS ($r = .43$ and $.44$, $p < .01$, respectively) and between intelligence and DSPS ($r = .49$, $p < .01$). They were significantly stronger ($z = 1.80$, $p < .05$) than the correlation between DSPS and DPS ($r = .35$, $p < .01$) or between intelligence and DPS ($r = .38$, $p < .01$). The relationship between IR and intelligence, measuring inductive reasoning skills (e.g. generalisation, discrimination) in academic and in figural content, respectively, proved to be the strongest ($r = .53$, $p < .01$).

Partial correlations were significantly lower as all bivariate relationships were influenced by the third construct, that is, either DPS; DSPS, or reasoning/intelligence ($r_{ir\_DSPS} = .26$; $r_{ir\_DSPS} = .33$; $r_{DSPS\_DPS} = .26$, $r_{CFT\_DSPS} = .40$; $r_{CFT\_DPS} = .26$; $p < .01$ respectively).

To analyse the stability of these processes, three different cohorts were selected and analysed separately: 5th graders, whose skills showed some development, but who had not yet entered the fast-growing phase, 7th graders, who were in the fast-growing developmental phase regarding DPS and 11th graders, who had left behind the fast-growing phase and were approaching the end of compulsory education. The correlation patterns showed instability across the three cohorts as they proved to be more homogeneous within grades than across grades. The strengths of the correlations of any two constructs were not generally influenced by the third construct (see Molnár et al., 2013a).

Figure 8.5 **Correlations between inductive reasoning, intelligence, domain-specific and domain-general problem solving**



Note: All coefficients are significant at $p < .01$; partial correlations are depicted as dotted lines. IR: inductive reasoning; CFT: culture-fair test of intelligence; DSPS: domain-specific problem solving; DPS: domain-general problem solving.

Source: Molnár et al. (2013a).

On the whole, the correlation between DPS and IR was the highest and the most stable over time. It was not due to students' level of DSPS skills. A possible explanation for this phenomenon could be that the basic mechanisms of IR, such as finding similarities and dissimilarities, or generating rules based on observation, are also the basic cognitive processes involved in DPS.

The relationship between DPS and DSPS became stronger (from no significant correlation to moderate but significant correlation) over time. This can be explained by the fact that the strategies used in DSPS and DPS situations become more similar, the mechanisms involved getting closer, over time. Older students have more opportunities to acquire and apply knowledge in a self-regulated way so DPS may be a prerequisite to DSPS. DSPS – if it is embedded in a domain-specific context as in our study – is based on knowledge application, whereas DPS – as captured in the MicroDYN approach – is a prerequisite to gaining and applying new knowledge.

The predictive power of first generation tests measuring thinking skills on DPS

The analysis regarding DSPS and DPS indicated a moderate correlational relationship between performance in static domain-specific and performance in domain-general dynamic problem solving environments (see research question 4). Thus, we could conclude that more traditional first generation tests of problem solving (DSPS) could predict performance in third generation tests of problem solving (DPS) and the other way around in a synchronised manner. To analyse the causal and one-way predictive force of first generation tests on third generation tests, SEM analyses are needed.

In the first, simplest SEM model, DPS measured by a third generation test was regressed on DSPS assessed by a first generation static test. The measurement model showed a good fit for the overall sample (CFI = 1.00, TLI = 1.00, RMSEA = .00). The standardised path coefficient was $\beta = .33$. A significant amount of variance remained unexplained.

In the second model, DSPS and CFT were used as predictors for DPS. The measurement model still showed a good fit for the overall sample (CFI = 1.00, TLI = 1.00, RMSEA = .00). The standardised path coefficients were .19 for DSPS and .29 for CFT/intelligence. That is, DSPS predicted performance in DPS beyond CFT. However, a significant amount of variance still remained unexplained and path coefficients were only moderately high.

In the third model, DPS was regressed on DSPS, CFT and IR. The standardised path coefficients dropped (to .13, .18 and .26 respectively), indicating the role of IR in the predictive power of DPS. The fit of the measurement model was good. A significant amount of variance remained unexplained in this model as well. Accordingly – knowing the limitations of the analyses, namely that although the tests measure problem solving, they do not measure the same construct in a strict sense (see above) – it is possible to predict performance in dynamic problem solving, from performance in traditional, static tests of DSPS, IR or intelligence, but the power of prediction is limited.

Thus, third generation tests are designed to require more cognitive skills and therefore additional aspects of problem solving that are relevant in today's life but are not captured by classical first generation tests of domain-specific skills..

Discussion

Our aim was to review the results of a recently conducted project that investigated different thinking skills and their relevance in educational settings (see Greiff et al., 2013; Molnár et al., 2013a, 2013b). We concentrated on dynamic problem solving, which we approached from an assessment perspective. We analysed its dimensionality and measurement invariance, and tested how it is influenced by other cognitive constructs such as domain-specific problem solving, fluid intelligence and reasoning assessed by domain-specific problem solving, the Culture Fair Test and inductive reasoning tests in an educational context.

Generally, the results of the study provided support for a view of DPS as a set of indispensable skills for the 21st century having relevance to educational settings from a developmental perspective. Our analysis shows that DPS can be better understood as involving two factors, knowledge acquisition and knowledge application, rather than subsuming both dimensions under one factor. DPS proved to be measurement invariant across different school grades, implying that differences in DPS performance can be validly interpreted. DPS developed following a regular trend, spanning several years of compulsory schooling, and the development curve varied across different school types. The correlations between DPS, DSPS, and IR were moderate and changed over time, as did those between DPS, DSPS and intelligence. With some limitations, it was possible to predict performance in dynamic problem solving measured by a third generation test by performance in traditional, static testing situations assessing thinking skills including problem solving (see also Fischer et al., 2015).

Dimensionality

More specifically, the MicroDYN approach gave students the opportunity to demonstrate their level of skills regarding both acquiring and applying knowledge. This confirmed previous research results reporting substantial but not perfect correlations between knowledge acquisition and knowledge application (e.g. Bühner et al., 2008; Kröner et al., 2005; Wüstenberg et al., 2012) regarding problem solving, which was identified by Mayer and Wittrock (2006) and confirmed to be significant by PISA (OECD, 2010, 2014b). Our result corroborated the hypothesis that knowledge acquisition is a necessary but not a sufficient condition of knowledge application (Greiff et al., 2013). These are two general and overarching processes of problem representation and solution at a broad level. However, one could object that knowledge acquisition and its application are themselves composed of several secondary processes or skills, each of which may be relevant in educational settings, and that those were neglected in the current study (see Greiff et al., 2013). Future research should address this question and define the components of lower-level processes.

Invariance

The measurement of these two dimensions was invariant across the students tested, from 5th to 11th grade. That means individual differences in DPS scores can be interpreted as true differences in the construct allowing direct comparisons of students in different grades and building a DPS scale from 5th grade to 11th grade. Such measurement invariance is a prerequisite for analyses on developmental patterns. However, this does not mean that the two factors of DPS, knowledge acquisition and application, relate to each other in the same way across all grades. Further research is needed to describe the changes in the relationship between knowledge acquisition and knowledge application with respect to DPS across different school grades and into adulthood.

Development

The development of DPS took place mostly during the compulsory schooling years, offering an opportunity to explicitly foster this higher-order thinking skill. As the greatest development was observed between 6th and 8th grade, that is the most sensitive and effective period in which to enhance students' DPS skills (Molnár et al., 2013a). This supports previous research results reporting relatively slow development of thinking skills (see, e.g., Csapó, 1997), and a lack of explicit training (de Koning, 2000; Molnár, 2011). Development occurs spontaneously as a "by-product" of schooling rather than as a result of explicit training (de Koning et al., 2002; Molnár et al., 2013a). Our results also confirm the importance of the elementary school years in the fostering of thinking skills as a possible way of making education more effective (Adey et al., 2007). To achieve this aim in practice, explicit training (Molnár, 2011) or different teaching methods (e.g. Shayer and Adey, 2002) have been suggested and proved to be effective tools.

The developmental curves differed notably after 8th grade in the different school types in the Hungarian school system. Mean performance differences between classes of vocational and grammar school students could be expressed in several developmental years and proved to remain stable over time. This confirmed previous research results showing high performance differences even in non-curriculum-related situations such as DPS, and reporting a drop in performance and motivation in 9th grade test scores, creating a gap between Hungarian vocational and grammar school students as a result of the strong selection process after primary school (OECD, 2005; Csapó, Molnár and Kinyó, 2008). Our findings support the claim that the effect of the extreme selectiveness exercised in curriculum-related situations, i.e. when students need to use knowledge acquired at school in order to find the desired solutions, can be detected in more general problem-solving situations as well, when domain-general cognitive processes are needed instead of content knowledge and rote learning (Buchner, 1995; Molnár et al., 2013a). Our results indicated a hypothetical significant role of IR and intelligence in the development of DSPS and DPS.

Relationships between thinking skills

The role of IR and intelligence in the development of DSPS and DPS proved to be significant, indicating that IR and intelligence are correlated with, but yet distinct from, the knowledge acquisition and application phase of DPS. As a result of the analyses, it can be assumed that interventions affecting one skill may affect the other as well. For instance, if DPS is fostered, DSPS and IR will be developed indirectly as well. This result draws attention to the importance of developing knowledge acquisition and knowledge application skills that also go along with improved reasoning skills (Hamers, De Koning and Sijtsma, 2000; Klauer, 1996), suggesting that together with IR (de Koning, 2000; Resnick, 1987), DPS should become an integral part of school agendas and should be incorporated into a broad range of school-related learning activities. This highlights the importance of developing of special methods to enhance DPS, such as guided discovery, or using DPS tasks as assessment and eventual enhancement tools to foster students' IR and domain-specific knowledge acquisition and application skills (see Molnár et al., 2013a).

The fact that strength of the relationships between the four thinking skills (IR, intelligence, DSPS and DPS) changed over time signals that good problem solvers somehow learn to know how to solve both DPS and DSPS tasks, or that problem solving strategies used in DSPS situations become more and more similar to the strategies used in domain general DPS problems (see also Fischer et al., 2015). This further emphasises the importance of explicit training of DPS. After a while the same mechanisms, the same knowledge acquisition and application skills are made use of when solving different kinds of problems (Molnár et al., 2013a), while the role of experience and knowledge in specific academic areas for solving problems decreases over time.

Predictive power of first generation tests

DPS, as measured with a state-of-the-art third generation test, was predicted by DSPS and intelligence measured by more traditional first generation tests. This confirms the results in research question 4 in general, while some variance remained unexplained. This may point towards the explanation for research question 4 and the results in this study and in the literature, specifically that DPS measures skills not measured by DSPS or intelligence, and predicts performance beyond reasoning in school grades (Wüstenberg et al., 2012). The results highlight once again the importance of DPS and, as a by-product, the importance of third generation testing, since DPS cannot be captured by first generation tests.

The present study contributes to the issue of assessing 21st-century skills by using third generation tests of cross-curricular skills indispensable for successful participation in modern Western society, thus overcoming some of the limitations associated with understanding only specific academic abilities such as reading, writing, and maths. We conclude by emphasising 1) the possibility of using and measuring new constructs by third generation tests; and 2) the potential of thinking skills such as dynamic problem solving as educationally relevant constructs that can and should be fostered over several years during public education. Encouraging the development of knowledge acquisition and knowledge application, that is, enhancing thinking skills relevant for both academic and non-academic contexts, is one of the most important and most challenging educational goals lying ahead.

Notes

1 The same data were used as in Molnár et al. (2013a) and Greiff et al. (2013) but unique analyses were conducted here.

References

Adey, P., B. Csapó, A. Demetriou, J. Hautamaki, and M. Shayer (2007), "Can we be intelligent about intelligence? Why education needs the concept of plastic general ability", *Educational Research Review*, Vol. 2/2, pp. 75-97, http://dx.doi.org/10.1016/j.edurev.2007.05.001.

Baker, E.L. and R.E. Mayer (1999), "Computer-based assessment of problem solving", *Computers in Human Behavior*, Vol. 15/3-4, pp. 269-282.

Buchner, A. (1995), "Basic topics and approaches to the study of complex problem solving", in P.A. Frensch and J. Funke (eds.), *Complex Problem Solving: The European Perspective*, Erlbaum, Hillsdale, NJ, pp. 27 63.

Bühner, M., S. Kröner and M. Ziegler (2008), "Working memory, visual-spatial intelligence and their relationship to problem-solving" *Intelligence*, Vol. 364, pp. 672-680.

Byrne, B.M. and S.M. Stewart (2006), "The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process", *Structural Equation Modeling*, Vol. 13/2, pp. 287-321.

Carroll, J.B. (2003), "The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors" in H. Nyborg (ed.), *The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen*, Pergamon, Amsterdam, pp. 5-21.

Csapó, B. (1997), "The development of inductive reasoning: Cross-sectional assessments in an educational context", *International Journal of Behavioral Development*, Vol. 20/4, pp. 609-626.

Csapó, B., G. Molnár and L. Kinyó (2008), "Analysis of the selectiveness of the Hungarian educational system in international context", paper presented at the 3rd IEA International Research Conference in Taipai, Taiwan, 16-20 September 2008.

Csapó, B., J. Ainley, R. Bennett, T. Latour and N. Law (2012), "Technological issues of computer-based assessment of 21st century skills" in P. Griffin, B. McGaw and E. Care (eds.), *Assessment and Teaching of 21st Century Skills*, Springer, New York, pp. 143-230.

De Koning, E. (2000), *Inductive Reasoning in Primary Education: Measurement, Teaching, Transfer*, Kerckebosch, Zeist.

De Koning, E., J.H.M. Hamers, K. Sijtsma and A. Vermeer (2002), "Teaching inductive reasoning in primary education", *Developmental Review,* Vol. 22, pp. 211–241.

Ericsson, K.A. and R. Hastie (1994), "Contemporary approaches to the study of thinking and problem solving", in R.J. Sternberg (ed.), *Thinking and Problem Solving*, 2nd edition, Academic Press, New York, pp. 37-79.

Fischer, A., S. Greiff and J. Funke (2012), "The process of solving complex problems", *Journal of Problem Solving*, Vol. 4/1, pp. 19-42.

Fischer, A., S. Greiff, S. Wüstenberg, J. Fleischer, F. Buchwald, and J. Funke (2015), "Assessing analytic and interactive aspects of problem solving competency", *Learning and Individual Differences*, Vol. 39, pp. 172-179, http://dx.doi.org/10.1016/j.lindif.2015.02.008.

Fraillon, J., et al. (2013), "Measuring computer and information literacy across countries", paper prepared for the 5th IEA International Research Conference, Singapore, 26-28 June 2013, www.iea.nl/fileadmin/user_upload/IRC/IRC_2013/Papers/IRC-2013_Fraillon_etal.pdf.

Funke, J. (2014), "Analysis of minimal complex systems and complex problem solving require different forms of causal cognition", Frontiers in Psychology, Vol. 5, http://dx.doi.org/10.3389/fpsyg.2014.00739.

Funke, J. (2010), "Complex problem solving: A case for complex cognition?", *Cognitive Processing*, Vol. 11/2, pp. 133-142, http://dx.doi.org/10.1007/s10339-009-0345-0.

Funke, J. (2001), "Dynamic systems as tools for analysing human judgement", *Thinking and Reasoning*, Vol. 7/1, pp. 69-89, http://dx.doi.org/10.1080/13546780042000046.

Gentner, D. (1989), "The mechanisms of analogical learning", in S. Vosniadou and A. Ortony (eds.), *Similarity and Analogical Reasoning*, Cambridge University Press, London. pp. 199-241.

Greiff, S. (2012), "Assessment and theory in complex problem solving – A continuing contradiction?", *Journal of Educational and Developmental Psychology*, Vol. 2/1, pp. 49-56.

Greiff, S. and A. Fischer (2013), "Der Nutzen einer komplexen Problemlösekompetenz: Theoretische Überlegungen und Empirische Befunde" (On the usefulness of complex problem solving competency: Theoretical considerations and empirical results), *Zeitschrift für Pädagogische Psychologie*, Vol. 27/1, pp. 27-39.

Greiff, S., D.V. Holt and J. Funke (2013), "Perspectives on problem solving in cognitive research and educational assessment: Analytical, interactive, and collaborative problem solving", *Journal of Problem Solving*, Vol. 5, pp. 71-91, http://dx.doi.org/10.7771/1932-6246.1153.

Greiff, S., et al. (2014), "Domain-general problem solving skills and education in the 21st century", *Educational Research Review*, Vol. 13, pp. 74-83.

Greiff, S., S. Wüstenberg, and J. Funke (2012), "Dynamic problem solving: A new measurement perspective", *Applied Psychological Measurement*, Vol. 36/3, pp. 189-213.

Greiff, S., et al. (2013), "Complex problem solving in educational settings – Something beyond *g*: Concept, assessment, measurement invariance, and construct validity, *Journal of Educational Psychology*, Vol. 105/2, pp. 364-379, http://dx.doi.org/10.1037/a0031856.

Griffin, P., B. McGaw and E. Care (eds.) (2012), *Assessment and Teaching of 21st Century Skills*, Springer, Dordrecht.

Hamers, J.H.M., E. De Koning and K. Sijtsma (2000), "Inductive reasoning in third grade: Intervention promises and constraints", *Contemporary Educational Psychology*, Vol. 23/2, pp. 132-148.

Hesse, et al. (2015), "A framework for teachable collaborative problem solving skills", in P. Griffin and E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 37-56). Springer, Dordrecht.

Kingston N. M. (2009), "Comparability of computer- and paper-administered multiple-choice tests for K–12 populations: A synthesis", *Applied Measurement in Education*, Vol. 22/1, pp. 22-37.

Klauer, K.J. (1996), "Begünstigt induktives Denken das Lösen komplexer Probleme? Experimentellen Studien zu Leutners Sahel-Problem" (Has inductive thinking a positive effect on complex problem solving?), *Zeitschrift für Experimentelle Psychologie*, Vol. 43/1, pp. 361-366.

Klauer, K.J. and G.D. Phye (2008), "Inductive reasoning: A training approach", *Review of Educational Research*, Vol. 78/1, pp. 85-123.

Klieme, E. (2004), "Assessment of cross-curricular problem-solving competencies" in J.H. Moskowitz and M. Stephens (eds.), *Comparing Learning Outcomes: International Assessments and Education Policy*, Routledge Falmer, London, pp. 81–107.

Kröner, S., J.L. Plass and D. Leutner (2005), "Intelligence assessment with computer simulations" *Intelligence*, Vol. 33/4, pp. 347-368, http://dx.doi.org/10.1016/j.intell.2005.03.002.

Mayer, R.E. and M.C. Wittrock (2006), "Problem solving", in P. A. Alexander and P. H. Winne (eds.), *Handbook of Educational Psychology*, Erlbaum, Mahwah NJ, pp. 287-303.

Mayrath, M., J. Clarke-Midura and D. Robinson (eds.) (2012), *Technology-Based Assessment for 21st Century Skills: Theoretical and Practical Implications from Modern Research*, Springer-Verlag, New York.

Molnár, G. (2011), "Playful fostering of 6- to 8-year-old students' inductive reasoning", *Thinking Skills and Creativity*, Vol. 6/2, pp. 91-99.

Molnár, G., S. Greiff and B. Csapó (2013a), "Inductive reasoning, domain specific and complex problem solving: Relations and development", *Thinking Skills and Creativity*, Vol. 9/8, pp. 35-45, http://dx.doi.org/10.1016/j.tsc.2013.03.002.

Molnár, G., S. Greiff and B. Csapó (2013b), "Relations between problem solving, intelligence and socio-economic background", paper presented at the EARLI, Munich, 27-31 August 2013.

Muthén, L.K. and B.O. Muthén (2010), *Mplus User's Guide*, Muthén & Muthén, Los Angeles.

OECD (2014a), *PISA 2012 Results: What Students Known and Can Do (Volume 1, Revised Edition, February 2014): Student Performance in Mathematics, Reading and Science,* OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264208780-en.

OECD (2014b), *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264208070-en.

OECD (2013), *PISA 2015: Draft Collaborative Problem Solving Framework*, OECD, Paris, www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf.

OECD (2010), *PISA 2012 Field Trial Problem Solving Framework*, OECD, Paris.

OECD (2005), *Problem Solving for Tomorrow's World: First Measures of Cross-Curricular Competencies from PISA* 2003, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264006430-en.

Pachler, N., et al. (2010), "Formative e-assessment: Practitioner cases" *Computers & Education*, Vol. 54/3, pp. 715-721.

Pellegrino, J.W. and R. Glaser (1982), "Analyzing aptitudes for learning: Inductive reasoning" in R. Glaser (ed.), *Advances in Instructional Psychology*, Vol. 2. Erlbaum, Hillsdale, NJ, pp. 269-345.

Raven, J. (2000), "Psychometrics, cognitive ability, and occupational performance", *Review of Psychology*, Vol. 7, pp. 51-74.

Resnick, L. (1987), *Education and Learning to Think*, National Academy Press, Washington, DC.

Ripley, M., et al. (2009), *Review of Advanced e-Assessment Techniques (RAeAT) Final Report*, Joint Information Systems Committee.

Ropo, E. (1987), "Skills for learning: A review of studies on inductive reasoning", *Scandinavian Journal of Educational Research*, Vol. 31/1, pp. 31-39.

Shayer, M. and P. Adey (Eds.) (2002), *Learning Intelligence: Cognitive Acceleration Across the Curriculum from 5 to 15 Years*, Open University Press, Milton Keynes.

Sternberg, R.J. (1995), "Expertise in complex problem solving: A comparison of alternative concepts" in P.A. Frensch and J. Funke (eds.) (1995), *Complex Problem Solving: The European Perspective*, Erlbaum Associates, Hillsdale, NJ, pp. 295-321.

Sternberg, R. (ed.) (1994), *Thinking and Problem Solving*, Academic Press, San Diego.

Vandenberg, R.J and C.E. Lance (2000), "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research", *Organizational Research Methods*, Vol 3/1, pp. 4-70

Wang, S., et al. (2008), "Comparability of computer-based and paper-and-pencil testing in K–12 reading assessment: A meta-analysis of testing mode effects", *Educational and Psychological Measurement*, Vol. 68/1, pp. 5-24.

Way, W.D., L.L. Davis and S. Fitzpatrick (2006), "Score comparability of online and paper administrations of the Texas Assessment of Knowledge and Skills", paper presented at the Annual Meeting of the National Council on Measurement in Education. San Francisco, April 2006.

Weiß, R. H. (2006), *Grundintelligenztest Skala 2 – Revision – (CFT 20-R)*, Hogrefe, Göttingen.

Wüstenberg, S., S. Greiff and J. Funke (2012), "Complex problem solving. More than reasoning?", *Intelligence*, Vol. 40/, pp. 1-14, http://dx.doi.org/10.1016/j.intell.2011.11.003.