UNIVERSITY OF SZEGED
DOCTORAL SCHOOL OF EDUCATION
INFORMATION AND COMMUNICATION TECHONOGIES IN EDUCATION

HAO WU

Identifying the Influencing Factors of Problem Solving: A Cross-national Study

PHD DISSERTATION

SUPERVISOR:
GYÖNGYVÉR MOLNÁR DSC
PROFESSOR OF EDUCATION

SZEGED, 2019

Table of Contents

Table of Contents.. 1

Acknowledgement...5

List of Abbreviations...6

List of Figures.. 8

List of Tables...11

1. Introduction.. 13

2. The concept and assessment of problem solving.. 18

  2.1 Problem and problem solving: definitions and types of problem solving......18

  2.2 The theory of CPS & sub-types of CPS.. 20

    2.2.1 Domain-general & domain-specific problem solving.........................23

    2.2.3 Well-defined & ill-defined problem solving......................................27

  2.3 The development and different approaches to measure problem solving......28

    2.3.1 The possibilities provided by computer-based assessment beyond paper-and-pencil testing... 28

    2.3.2 Assessment of static problem solving.. 30

    2.3.3 Assessment of interactive problem solving..34

    2.3.4 Beyond individual testing of problem solving: assessment of collaborative problem solving...40

    2.3.5 The challenges and potential of analyzing log-file data in problem solving environment...43

  2.4 Factors influencing problem solving achievement in educational context.... 44

    2.4.2 Information processing..47

    2.4.3 Non-cognitive variables.. 52

  2.5 Students' exploration strategies in interactive problem solving in the MicroDYN approach...55

  2.6 Cross-cultural assessment of problem solving..56

    2.6.1 Problem-solving research in China.. 56

2.6.2 Computer-based assessment of problem solving in China and in Hungary........58

3. Research questions and hypotheses........62

4. Pilot study: The feasibility of computer-based assessment of reasoning skills in China........68

4.1 Aims, research questions and hypotheses for the pilot study........68

4.2 Methods........69

4.2.1 The sample of the pilot study........69

4.2.2 Instruments........69

4.2.3 Procedures........81

4.2.4 Data analysis........81

4.3 Results........82

4.3.1 Descriptive statistics: Computer-based thinking skill assessment is feasible in China........82

4.3.2 The behaviour of the IPS test in China: dimensionality and difficulty83

4.3.3 Students' test-taking motivation in IPS environment........86

4.3.4 The relationships between students' performance in problem solving, inductive reasoning, creativity, working memory and learning strategies...91

4.3.5 The influencing factors of problem solving........92

4.4 Conclusions of the pilot study........93

4.5 Limitations of the pilot study........94

5. Cross-national comparison study of problem solving and its influential factors in educational context........96

5.1 Aims, research questions and hypotheses of the comparison study........96

5.2 Methods........97

5.2.1 The sample of the comparison study........97

5.2.2 Instruments........98

5.2.3 Procedures........100

5.2.4 Data analyses........100

5.3 Descriptive statistics of the thinking skill tests........101

5.3.1 Descriptive statistics of the thinking skill tests– China.....................101

5.3.2 Descriptive statistics of thinking skill tests– Hungary...................... 107

5.4 Cross-cultural comparison of problem solving, inductive reasoning and combinatorial reasoning.. 111

5.4.1 Dimensionalities of IPS, inductive reasoning and combinatorial reasoning in Chinese and Hungarian contexts.. 111

5.4.2 Measurement invariance across nationality and gender....................114

5.4.3 Students developmental level of problem solving, inductive reasoning and combinatorial reasoning in China and in Hungary.............................116

5.4.4 Similarities and differences in the relationships between problem solving, inductive reasoning and combinatorial reasoning in China and Hungary..117

5.5 Cross-cultural comparison of the relationships between problem solving performance and students' developmental level of working memory, test-taking motivation and learning strategies..120

5.5.1 The influential level of working memory...120

5.5.2 The influential level of students' test-taking motivation in interactive problem-solving environment.. 121

5.5.3 The influential level of parents' educational level........................... 125

5.5.4 The influential level of students' learning strategies.........................127

5.6 Overall models with all factors included...129

5.7 Discussions and limitations... 134

6. Cross-national comparison study of students' problem solving strategies: Log-file analysis... 136

6.1 Background.. 136

6.2 Aims, research question and hypothesis..136

6.3 Methods...137

6.3.1 Participants... 137

6.3.2 Instruments... 137

6.3.3 Scoring...138

6.3.4 Data analyses..140

6.4 Results and discussions...140

6.4.1 Cross-cultural comparison of the theoretically effective strategy usage in problem-solving environment..140

6.4.2 Cross-cultural comparison of the efficacy of VOTAT strategies......143

6.4.3 Distinguishing qualitatively different explorer class profiles in China and in Hungary..147

6.5 Conclusions and limitations of the log-file analyses...................................151

7. Summary of the Study..154

7.1 Summary of the results and answering the research questions...................154

7.2 Discussions of the findings..156

7.3 Limitations..159

7.4 Originality...159

Publications related to the dissertation...161

References..164

Appendixes...182

Appendix A: The Hungarian and the Chinese version of the IPS problem Cat.183

Appendix B: Sample items from the Chinese version of the inductive reasoning test..185

Appendix C: Sample items from the Chinese version of the combinatorial reasoning test...187

Appendix D: Sample item from the Chinese version of the working memory test188

Appendix E: Sample item from the Chinese version of the creativity test........189

Appendix F: Background questionnaire (in English)..190

Acknowledgement

I would first like to thank my supervisor Prof. Gyöngyvér Molnár for providing me with the wonderful opportunity to do my PhD research under her supervision. Thank you for all the support and patience in guiding me through my PhD project. Your wise advice on my research, career, and life have been priceless.

I would like to express my gratitude to my internal defense reviewers, Andrea Magyar and Anita Pásztor-Kovács, for their precious comments and suggestions. My sincere appreciations also go to the Doctoral School of Education at the University of Szeged. I am so glad to have worked with a group of outstanding researchers. I am also thankful to the staff members in the institute for their kind help with my studies.

A special gratitude goes to my wife, Feng Wang. Studying and living in a foreign country are not always an easy thing to do, but you make me feel Hungary like home. Marrying you was the luckiest thing in my life.

Thank you, mom and dad, for always encouraging and supporting me, and for all the sacrifices that you have made for me. I love you with all my heart, and I am proud to be your son. Now me and Feng are planning to have our own children. I hope we can be the perfect parents like you.

Last but not least, to my friends in Szeged and Budapest, thank you for the happy times we had together. We came from all corners of the world. I have been so lucky to have you on this journey.

List of Abbreviations

aBIC	Adjusted Bayesian Information Criterion
AIC	Akaike information criterion
ANOVA	Analysis of Variance
BIC	Bayesian Information Criterion
CFI	Comparative Fit Index
CN	China
CPS	Complex Problem Solving
CR	Combinatorial Reasoning
df	Degree of Freedom
eDia	Electronics Diagnostic Assessments System
HU	Hungary
ICT	Information Communication Technology
IPS	Interactive Problem Solving
IR	Inductive Reasoning
IRT	Item Response Theory
LCA	Latent Class Analysis
L-M-R	Lo-Mendell-Rubin Adjusted Likelihood Ratio
M	Mean
NAEP	National Assessment of Educational Progress
OECD	Organization for Economic Co-operation and Development
PISA	Programme for International Student Assessment
PS	Problem Solving
RMSEA	Root Mean Square Error of Approximation
SD	Standard Deviation
SEM	Structural Equation Modeling
SRMR	Standardized Root Mean Square Residual
TIMSS	Trends in International Mathematics and Science Study

TLI	Tucker-Lewis Index
VOTAT	Vary One Thing at A Time
WM	Working Memory

List of Figures

Fig. 2.1 Components of CPS

Fig. 2.2 Process for information processing

Fig. 2.3 Sample assessment item from PISA 2012 problem-solving assessment

Fig. 2.4 Example of a static problem-solving item in PISA 2003

Fig. 2.5 Example of a static problem-solving item in PISA 2012

Fig. 2.6 A simple finite state automata using the MicroFIN approach

Fig. 2.7 Interactive problem-solving task “MP3 Player”: Stimulus information

Fig. 2.8 Structure of a typical MicroDYN task with direct effects, side effect and eigendynamics

Fig. 2.9 Interactive Problem solving task “Climate Control”: Stimulus information

Fig. 2.10 Sample item from the PISA 2015 computer-based collaborative problem-solving assessment

Fig. 2.11 Factors of three categories for problem solver

Fig. 2.12 The working memory model

Fig. 2.13 Structure of inductive reasoning

Fig. 2.14 Developmental curve of complex problem solving

Fig. 2.15 Students' performance in the PISA 2012 problem solving assessment

Fig. 2.16 Relative performance in problem solving

Fig. 4.1 Screenshot of the MicroDYN task Cat – First phase

Fig. 4.2 Screenshot of the MicroDYN task Cat – Second phase

Fig. 4.3 Sample items for the inductive reasoning test

Fig. 4.4 Sample item for the creativity test

Fig. 4.5 Stimuli for the picture meaning tasks

Fig. 4.6 Sample item for the working memory test

Fig. 4.7 Sample item for the motivation test

Fig. 4.8 Item-person map for the IPS test

Fig. 4.9 Item-person map for the inductive reasoning test

Fig. 4.10 Item-person map for the motivation questions

Fig. 4.11 Item-person map with gender difference

Fig. 4.12 Structural equation model of problem solving and its influencing factors

Fig. 5.1 Sample item for the combinatorial reasoning test-figural

Fig. 5.2 Sample item for the combinatorial reasoning test-verbal

Fig. 5.3 Item-person map for the IPS test – Chinese sample

Fig. 5.4 Item-person map for the inductive reasoning test – Chinese sample

Fig. 5.5 Item-person map for the combinatorial reasoning test – Chinese sample

Fig. 5.6 Correlations between the thinking skills – Chinese sample

Fig. 5.7 Item-person map for the IPS test – Hungarian sample

Fig. 5.8 Item-person map for the inductive reasoning test – Hungarian sample

Fig. 5.9 Item-person map for the combinatorial reasoning test – Hungarian sample

Fig. 5.10 Correlations between thinking skills – Hungarian sample

Fig. 5.11 A structural model presenting the relationships among problem solving, combinatorial reasoning and inductive reasoning – Chinese sample

Fig. 5.12 A structural model presenting the relationships among problem solving, combinatorial reasoning and inductive reasoning – Hungarian sample

Fig. 5.13 Item-person map for the motivation questions

Fig. 5.14 Item-person map with nation & gender differences

Fig. 5.15 Relationships between parents' education, learning strategies, and problem solving – Chinese sample

Fig. 5.16 Relationships between parents' education, learning strategies, and problem solving – Hungarian sample

Fig. 5.17 Relationships between each measured skill and factor – Chinese sample

Fig. 5.18 Relationships between each measured skill and factor – Hungarian sample

Fig. 6.1 Efficacy of the most frequently employed VOTAT strategies on problems with two input variables

Fig. 6.2 Efficacy of the most frequently employed VOTAT strategies on problems with three input variables

Fig. 6.3 Three qualitatively different class profiles for the Chinese sample

Fig. 6.4 Three qualitatively different class profiles for the Hungarian sample

List of Tables

Table 4.1 Test-taking motivation questionnaire questions

Table 4.2 Learning strategies questionnaire questions

Table 4.3 Mean, standard deviation and Cronbach's alpha for each thinking skill test

Table 4.4 Goodness of fit indices for testing dimensionality of IPS

Table 4.5 Model fits for rating scale model and partial credit model

Table 4.6 Correlation matrix between the measured thinking skills

Table 5.1 Descriptive statistics for each assessed thinking skill and their subscales – Chinese sample

Table 5.2 Descriptive statistics for each assessed thinking skill and their subscales – Hungarian sample

Table 5.3 Goodness of fit indices for testing the dimensionality of IPS

Table 5.4 Goodness of fit indices for testing the dimensionality of combinatorial reasoning

Table 5.5 Goodness of fit indices for testing the dimensionality of inductive reasoning

Table 5.6 Goodness of fit indices for testing invariance across nationality and gender of problem solving

Table 5.7 Comparison of the ability levels of the assessed thinking skills

Table 5.8 Correlations between working memory and thinking skills

Table 5.9 Model fits for rating scale model and partial credit model

Table 5.10 Distribution of participants' parental educational levels

Table 5.11 Performance differences for students with different parental educational levels – Chinese sample

Table 5.12 Performance differences for students with different parental educational levels – Chinese sample

Table 5.13 Correlations between mothers' and fathers' educational level and thinking skills

Table 6.1 Percentage of the use of theoretically effective and non-effective strategy

Table 6.2 Problem-solving performance for the theoretically effective strategy users

Table 6.3 Problem-solving performance for the non-theoretically effective strategy users

Table 6.4 Percentage of VOTAT strategy use

Table 6.5 Fit indices for latent class analyses

Table 6.6 Frequencies and average latent class probabilities

1. Introduction

Problem solving is present in our daily activities. For instance, what to wear in the morning, how to use our new electric devices, how to get to our reserved restaurant by public transport, how to arrange our schedule to reach the highest work efficiency, how to communicate with people in a foreign country, etc. In most of the cases, solving the problems that keep occurring in our study, work, and daily life, is a must-do task. These situations require problem solving. Generally, problem solving is the thinking that occurs if we want "to overcome barriers between a given state and a desired goal state by means of behavioral and/or cognitive, multistep activities" (Frensch, & Funke, 1995, p. 18).

In the past, education aimed at teaching students knowledge and skills, but nowadays education is "[...]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" (Schleicher, 2017, p. 3). Nowadays, our society and environment keep changing all the time, the technologies used in almost every industry sector are also developing quickly. This situation means that the content of applicable knowledge evolves rapidly. People are facing problems almost every day: simple or complex ones, fixed or dynamically changing ones, familiar or completely new ones.The ability to solve problems in a timely manner and properly is gradually becoming one main factor for peoples' career and life. "Adapting, learning, daring to try out new things and always being ready to learn from mistakes are among the keys to resilience and success in an unpredictable world" (OECD, 2014a, p. 13). Problem solving is thus considered to be one of the most important 21st-century skills (the required skills for success in 21st century society and workplaces; Dede, 2010). Improving students' problem-solving skills has become one of the main aims and challenges in contemporary education (Greiff, Holt, & Funke, 2013).

Most of the time, the problems we face are different. Some of them are simple, some of them are complex; some of the problems have clear goals, but sometimes we need to figure out the goals by ourselves. Sometimes the problems are similar or even the same as the problems we have solved before; but it is also very possible we get some problems that we have never encountered before. Some problems are static, while some of them are dynamically changing. Both the problem-solving processes and the mental steps can differ widely while solving different types of problems. This study focuses on one specific kind of problem solving, which is called interactive problem solving (IPS) in the MicroDYN approach. IPS is a type of problem solving which requires the problem solver to do the direct interaction with the problem in order to uncover and discover relevant information (OECD, 2010).

Problem solving is a transversal skill (Greiff et al., 2014), operating several mental activities and cognitive skills (see Molnár, Greiff, & Csapó, 2013). Thus, defining its direct development in general is not a realistic research task, but we can define the development of solving different kinds of problems and we can map the skills, which have an effect on its development.

This study aims to detect the component skills of interactive problem solving in the MicroDYN approach. It tests structural relations and the predictive value of cognitive background variables on the development of IPS, thus provide the possibility to develop an effective training program on problem solving based on the research results regarding the component skills of IPS. Detecting and identifying differences in the mental and cognitive aspects for students with different cultural backgrounds in the IPS environment (see e.g. OECD, 2014a; Wüstenberg, Greiff, Molnár, & Funke, 2014) is also one of the emphases of this study. PISA (Programme for International Student Assessment) is an influential international educational assessment project launched by OECD (Organisation for Economic Co-operation and Development). Based on the PISA results regarding the developmental level of problem solving skills (OECD, 2014a), a high and an average achiever country have been selected for the study, thus China (mainland) and Hungary were involved in this cross-national study. Generally, the study aims to contribute to further understanding of the cognitive construction of interactive problem solving in the MicroDYN approach; and to highlight the similarities and differences regarding development and component skills of IPS in European and Asian educational contexts.

The literature review part of the dissertation starts with a discussion of the definitions of the basic terms of this study: problem and problem solving. There are still no universally accepted definitions available for these two terms. The dissertation reviews the definitions which have been offered in the past, and discusses different definitions proposed at different times or with different theoretical bases. It also introduces different types of problem solving (e.g. domain-specific & domain-general; static & dynamic; interactive; ill-defined & well-defined).

We also introduce the development of different approaches which focus on measuring problem solving and we provide an overview of the evolving problem-solving assessment methods: from paper-and-pencil to computer-based, and from first generation computer-based to third generation computer-based assessment of problem solving. Beyond the illustration of the different assessment methods for varying types of problem solving (analytical, interactive, collaborative), we discuss more deeply the challenges and potentials in computer-based assessment of problem solving, where log-file analysis – an unique feature provided by modern technologies – plays an important role.

Several thinking skills and non-thinking skill factors have the potential to influence one's problem-solving achievement (Greiff et al., 2014; Molnár, Greiff, & Csapó, 2013). This dissertation selects those cognitive skills (e.g., inductive reasoning, combinatorial reasoning, creativity) and socio-economic or affective factors (e.g., gender, parental education, test-taking motivation) which proved to be the most influential ones in small or large scale national studies (Casakin, 2007; Greiff et al., 2015; Molnár, Greiff, & Csapó, 2013; OECD, 2014a) and which have been the most frequently mentioned in the literature (Bisanz, Bisanz, & Korpan, 1994; Hamers, De Koning, & Sijtsma, 2000; Herrmann, 1995; Molnár, Greiff, Wüstenberg, & Fischer, 2017). We illustrate the relations between these selected, mostly reasoning skills and different background factors with problem solving.

Finally, the theoretical section reviews the available research results regarding Chinese and Hungarian students' problem solving performance in cross-cultural context. We strongly lean on the referring PISA results, as both China (represented by Shanghai) and Hungary participated in the PISA 2012 problem-solving assessment. This part analyzes and evaluates their performance in this assessment, and analyzes their advantages and disadvantages. It also reviews the problem-solving research in China and in Hungary, separately. And finally, this part discusses why this study is important and necessary to be conducted across China and Hungary.

Base on the literature review, the theoretical analyses and the research aims, the dissertation formulates eight research questions and puts forward corresponding hypotheses. Using the Hungarian experiences regarding assessment of IPS in educational context (Molnár, Greiff, & Csapó, 2013; Molnár et al., 2017), two studies, one pilot study in China and one comparison study across China and Hungary, were designed and conducted to answer the research questions and testify the hypotheses.

The pilot study was implemented with the aim to test and prove the feasibility of computer-based online assessment of thinking skills via first, second and third generation tests (Molnár et al., 2017) in China. Beyond the regular information such as the sample of the study and the procedures, the instruments part of the method section introduces the used assessment instruments in a very detailed form, including the original sources, development histories, operating methods and provides sample items. As a result of the pilot study conducted in China, we concluded that computer-based online assessment of thinking skills is feasible in China independently of the type of tests (first, second or third generation).

The design and realization of the cross-national comparison study relays on all of the experiences we have conducted in the Chinese pilot study and was carried out earlier regarding the Hungarian online assessments of IPS. In the pilot study section we also follow the regular introduction of an empirical project, thus the method part discusses the sample, the instruments (some modification were made after the pilot study), and the procedures. The structure of the result part follows the order of the research questions. Because of their novelty, complexity and importance, the results referring to the research questions on Chinese and Hungarian students' exploration strategies in interactive problem-solving environment using log-file analyses with state-of-art procedures for answering research questions (which could not have been answered by means of traditional assessment techniques and analyzing methods) are discussed in a separate section.

The final part of the dissertation summarizes the results and briefly answers shortly the research questions. Furthermore, the limitations and originality of this study are also stated here.

In conclusion, according to Lord Kelvin, "if you can't measure it, you can't improve it", this study measures Chinese and Hungarian students' problem-solving skills as well as the relative internal and external factors, in order to explore the nature of problem solving in both cultures. The results provide the possibility for improving students' problem-solving skills in the future.

2. The concept and assessment of problem solving

2.1 Problem and problem solving: definitions and types of problem solving

Several definitions and theoretical models of problem solving have been published in the literature (for an overview see Sternberg, 2013; Frensch, & Funke, 1995). Researchers have proposed several different definitions regarding the basic terms, even problem and problem solving themselves.

Before discussing the definition of problem solving, we should answer the following question: what is a problem and what makes a problem a problem? Although there exists no universal definition of what constitutes a problem, we intend to give an answer to the question by collecting some of the common features of the different definitions.

Some of the definitions focus on the absence of an observable response. For instance, Davis (1973) stated that a problem is "a stimulus situation for which an organism does not have a ready response" (Davis, 1973, p. 12). Similarly, Woods, Crow, Hoffman, & Wright (1985, p. 1) believed that a problem is a "stimulus situation for which an organism does not have a response," while a problem arises "when the individual cannot immediately and effectively respond to the situation". Meanwhile, some definitions focus on the absence of a non-observable thought or cognition, such as Newell & Simon (1972, p. 72) who stated that "a person is confronted with a problem when he wants something and does not know immediately what series of actions he can perform to get it"; and Hayes (1980, p. i) illustrated that "whenever there is a gap between where you are now and where you want to be, and you don't know how to find a way to cross that gap, you have a problem."

Even though the definitions of the terminus *problem* are highly varied, we can find some common features of them. Most of the definitions focus on the distance between the task and the solver but not the task itself. That is, a problem arises if there is a gap between task and problem solver, or a barrier between the given state and the problem solver's goal state. Therefore, it can be said that a problem is defined by the interaction between task characteristics and person characteristics (Frensch, & Funke, 1995).

The definition and research methodology of problem solving have also kept evolving in the past decades. At the early stage, problem solving was commonly described as a passive, reproductive and domain-general stepwise process, which is based on trial and error (Newell, & Simon, 1972; VanLehn, 1989). By contrast, Gestalt psychologists consider problem solving to be an active and productive process, where insight, reorganization and functional fixedness play an important role (Fiore, & Schooler, 1998). With the in-depth study, researchers confirmed that problem solving is a cognitive, rather than behavioral or neurophysiological, activity. Similarly, the PISA 2012 assessment defined problem solving as "cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious." (OECD, 2013, p. 122); and Fischer, Greiff and Funke (2012, p. 19) defined problem solving as "knowledge acquisition and knowledge application concerning the goal-oriented control of systems that contain many highly interrelated elements". In the knowledge acquisition phase of problem solving, the problem solver understands the problem itself, and also stores the acquired information (Funke, 2001; Novick, & Bassok, 2005). While in the knowledge application phase the problem solver applies the acquired knowledge to realize the transition from given state to goal state (Novick, & Bassok, 2005). This study adopted Frensch and Funke's (1995, p. 18) definition of problem solving which states that problem solving is a goal-directed "thinking that occurs to overcome barriers between a given state and a desired goal state by means of behavioral and/or cognitive, multistep activities." The problem solver has to organize information and to deal with ill-defined or more or less well-defined goals. The problem solver does not immediately know how to solve the problem or how to reach the goal (Frensch, & Funke, 1995).

Most of the early research on problem solving were typically conducted with relatively simple and laboratory tasks which were novel to subjects,.For example, Ewert and Lambert (1932) used the disk problem (now it has a more common name: Tower of Hanoi) in their problem-solving research. These kinds of simple novel tasks usually had clearly defined optimal solutions, and problem solvers were likely to solve it within a relatively short time frame. The researchers expected that the simple tasks could capture the main properties of real problems, so they could use such experiments to explore the problem solver's cognitive processes while solving real world problems.

However, starting from the 1970s, it has been gradually realized that empirical findings and theoretical concepts which came from simple laboratory tasks were not generalizable to more complex real-life problems (Frensch, & Funke, 1995). New, alternative ways and directions have been provided by the development of computers and the information-processing approach, namely the *problem-space theory* of Newell and Simon (1972). Moreover, these realizations have led to rather different research directions in North America and in Europe (Anzai, & Simon, 1979; Bhaskar, & Simon, 1977; Funke, 1991). North American research on problem solving has typically concentrated on domain-specific problem solving, examining the development of expertise in separate, natural knowledge domains, such as mathematics (Sokol, & McCloskey, 1991), reading (Stanovich, & Cunningham, 1991), computer skills (Kay, 1991), etc., and on the task – problem solver interaction; while much of the European research is about general problem solving, which focuses on solving processes of complex, real-world, unknown problems (commonly performed with computerized scenarios), and on the characteristic features of problems to be solved (Funke, 1991). The approach on complex problem solving proposed by Frensch and Funke (1995) is an integrated whole of studies based on the European and American models on problem solving (see section 2.2).

2.2 The theory of CPS & sub-types of CPS

Based on the empirical findings, Frensch and Funke (1995) constructed a theoretical framework which summarizes the basic components of complex problem solving and the interrelations among the components (Fig 2.1). The framework contains three separate components: problem solver, task, and environment. Based on the differences between these components, complex problem solving can be divided into several sub-types.

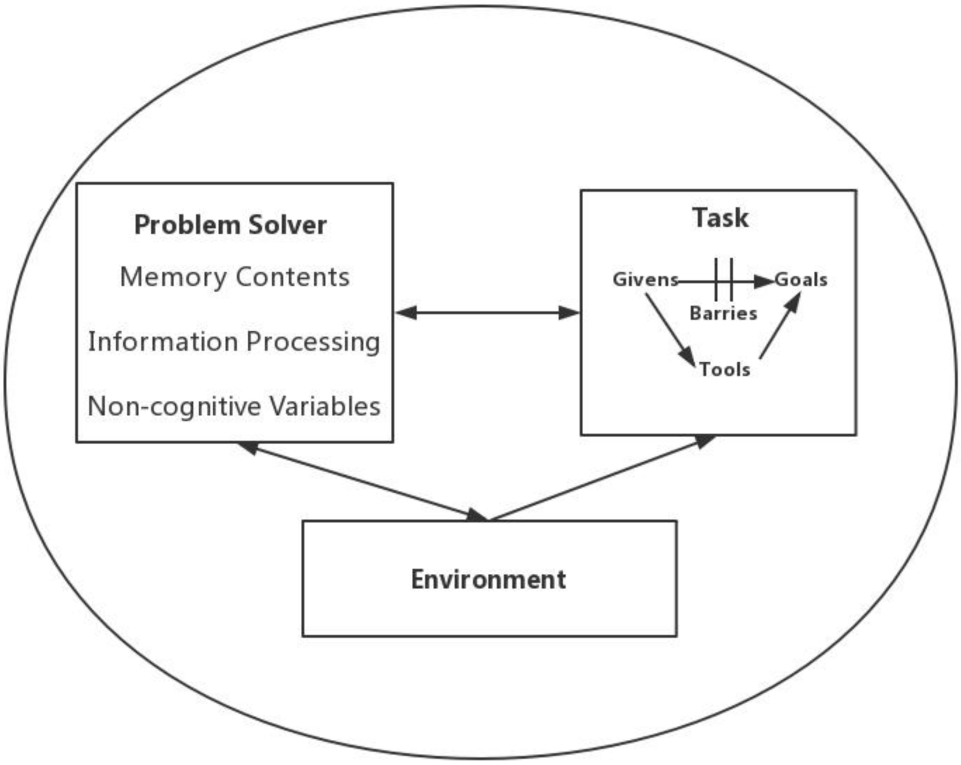

Problem Solver
Memory Contents
Information Processing
Non-cognitive Variables
Task
Givens
Goals
Barries
Tools
Environment

Fig. 2.1 Components of CPS (based on Frensch, & Funke, 1995)

The problem solver. The problem-solving process is influenced by the problem solver' memory content, dynamic information processing, and non-cognitive variables. Generally, the two sub-processes of problem solving are knowledge acquisition and knowledge application (Funke, 2001; Greiff, & Funke, 2017; Wüstenberg, 2013). The problem solvers' existing knowledge (memory content; if the problem is relevant to it) can help them to omit or shorten the time and effort for doing the sub-processes of knowledge acquisition. Generally, problem solving can be classified as *domain-general problem solving* and *domain-specific problem solving,* depending on what kind of background knowledge is required in the problem solving process (see section 2.2.1).

Compared with the problem solver's memory contents, the dynamic activities in the problem solving process have been described as information processing (Frensch, & Funke, 1995). According to Newell (1993), the main steps for the information process are as following:

1. Select a state; select a strategy.

2. Apply strategy to state, producing new state.

3. Decide if a goal state; decide to quit; decide to save the new state.

A model has been built according to the definition of problem solving – overcome the barriers between the given state and the goal state (Frensch, & Funke, 1995) – and the statement above from Newell (1993). Figure 2.2 visualizes this approach.

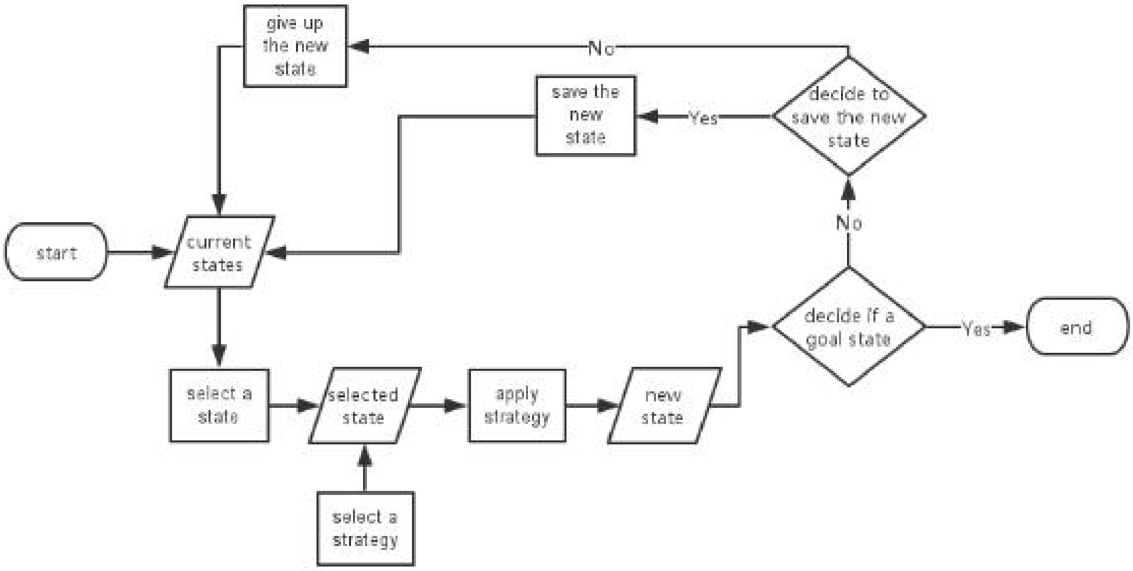

give up the new state
No
save the new state
Yes
decide to save the new state
No
start
current states
decide if a goal state
Yes
end
select a state
selected state
apply strategy
new state
select a strategy

Fig. 2.2 Process for information processing [created based on Newell's (1993) theory]

Beyond the cognitive ones, some of the non-cognitive variables (e.g. motivation, family background, etc.) are also able to effect problem-solving performance. A more elaborated discussion is provided in section 2.4.3.

Task. Task is described as the barriers that exist between a given state and a goal state (Frensch, & Funke, 1995). In some cases, the task is able to change dynamically (or interactively), while sometimes problems are static and non-changeable. These two cases are described as *interactive problem solving* (or in the literature also called *dynamic problem solving* or *complex problem solving*) and *static problem solving,* respectively.

The *environment* plays also an important role in the problem solving process. Environmental factors include e.g. feedback and feedback delay, expectations, cooperation and peer pressure. The environment is able to affects both the problem solver and the task. The environment can constrain the problem solver's information processes and influence the problem solver's accessible knowledge. During the problem solving process, some of the environmental factors can be improved, or at least modified by the problem solvers (Frensch, & Funke, 1995). Meanwhile, the environment can affects the task by several ways such as providing additional information, promoting or constraining the tool using, etc. The type of interaction between the environment and the problem itself (task) is one-way, that is the task does not have enough power to affect the environment (Frensch, & Funke, 1995).

2.2.1 Domain-general & domain-specific problem solving

Historically, there are two main categorizations in the problem solving research, namely domain-specific problem solving and domain-general problem solving. In this regard the emphases for American and European problem-solving researchers are different. In general, American researchers typically focus on students' problem solving in separate, natural knowledge domains (Bennett, Persky, Weiss, & Jenkins, 2007). These researchers frequently focus on the development of problem-solving skills within a certain domain, that is on the development of expertise. There are some areas that have attracted attention in North-American research in such diverse fields as reading, writing, calculation mathematics, political, computer skills etc. (e.g., Kay, 1991; Sokol, & McCloskey, 1991; Stanovich, & Cunningham, 1991). In contrast, many of the European researchers are more focused on the "latively complex, semantically rich, computerized laboratory tasks that are constructed to be similar to real life problems", which has also been described as and domain-general problem solving (Frensch, & Funke, 1995, p. 16).

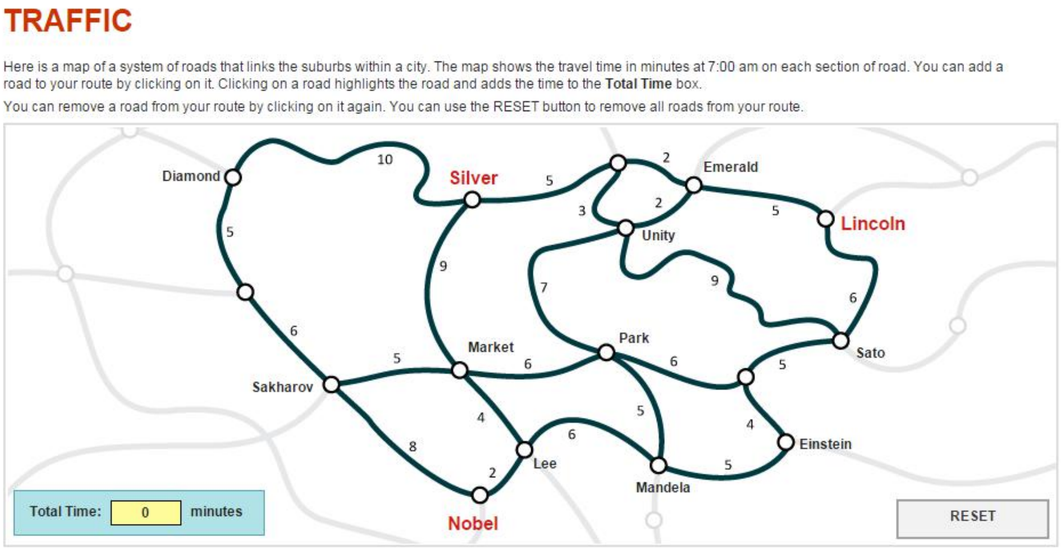
Every problem solver has his/her own background knowledge system and storage. When dealing with some problems, their knowledge can help them overcome the barriers. As introduced above, domain-specific and domain-general knowledge is required in domain-specific and domain-general problem solving processes, respectively.

Both domain-specific and domain-general problem solving are important in students' studies and life. To compare these two kinds of problem solving, Greiff Fischer, Stadler, & Wüstenberg (2014) stated that domain-specific problem solving skills and specific factual knowledge are very important when a problem solver is dealing with a routine task or a problem in a specific field, but they also pointed out that domain-specific problem solving is limited to narrow domains. In the 21th century, students' challenges and problems will not only be limited to school life, but more importantly will appear in in real life situations. Greiff et al. (2014) believed that modern education should not only provide students with factual knowledge and domain-specific problem-solving strategies, but should also provide a broader set of skills required in our modern societies. According to this requirement, domain-general problem solving also has an extremely significant meaning for students in the current age.

The assessment of domain-specific problem solving – compared with domain-general problem solving – usually had deeper connection with the problem solver's knowledge storage (Frensch, & Funke, 1995). For instance, in the NAEP assessment project launched by the United States there were some problems which required both of students' knowledge in one certain domain (e.g. Mathematics; see Dossey, 1993) and their ability of solving problems. But one obvious question for this kind of problem-solving assessment is students' performance in the relevant field or subject-area. Even a student with poor problem-solving skills can still have a certain probability to get a good performance in this kind of assessment if he/she has enough background knowledge in the concerned area. Frensch and Funke (1995) also believed that this kind of knowledge-based task focuses primarily on students' learning outcomes but not problem-solving ability.

According to Frensch and Funke (1995), problem-solving assessment should try to avoid the impact of examinee's knowledge storage. Based on their theory, in the problem solving assessment everyone should be novice (Frensch, & Funke, 1995). The emphasis of this study is on domain-general problem solving which focuses on real-world problems. In some assessment items which focus on the domain-general problem solving, like the assessment based on the MicroDYN approach or the assessment tasks in PISA, the researchers tried to design the tasks in a brand-new area about which no participants could have the relevant knowledge (Greiff et al., 2013; OECD, 2014a).

For instance, in the item "traffic" in the PISA 2012 problem-solving assessment (see Fig. 2.3), the map used was designed as completely fictional (including the towns' names, the distances between each town, the circulation map, etc.). So it can exclude the possibility that students are familiar with the geography situation which was shown in this map and solve the problems according to their own geography knowledge (OECD, 2014a).



TRAFFIC

Here is a map of a system of roads that links the suburbs within a city. The map shows the travel time in minutes at 7:00 am on each section of road. You can add a road to your route by clicking on it. Clicking on a road highlights the road and adds the time to the **Total Time** box.

You can remove a road from your route by clicking on it again. You can use the RESET button to remove all roads from your route.

Fig. 2.3 Sample assessment item from PISA 2012 problem-solving assessment (OECD, 2014a, p. 41)

Because of the cultural and curricular differences, *in the present study we preferred domain-general problem solving to domain-specific problem solving.*

Furthermore, the assessments in this study have been carefully designed to keep the impact from the students' knowledge storage to the minimum.

2.2.2 Static & interactive problem solving

In static problem solving environment (also referred to as analytical problem solving in some literature if all necessary pieces of information are provided for solving the problem, e.g. Csapó, & Funke, 2017; Fleischer, Buchwald, Leutner, Wirth, & Rumann, 2017), the task itself, and thus the information provided for the problem solver stay static and non-changeable. The problem solvers need to analyze these pieces of information to be able to solve the problem (Greiff et al., 2013). This kind of problem solving relies more on problem solver's reasoned application of existing knowledge (OECD, 2004). The classic problem solving task "Tower of Hanoi problem" (Ewert, & Lambert, 1932) is a typical example for a static problem. However, these analytical problem solving tasks are sometimes considered "simple" problems because they are not generalizable to the complex and interactive real-life problems (Ramalingam, Philpot, & McCrae, 2017).

Along with the development of problem-solving theory and technology (especially computer technology), interactive problem solving gradually replaced analytical problem solving as the mainstream in academic research (Frensch, & Funke, 1995). Interactive problem solving (IPS) is "characterized by the interaction between a problem solver and the problem to generate and integrate information about the problem" (Greiff et al., 2013, p. 76). And in IPS, relevant information needs to be actively generated, problem solvers need to have direct interaction with the problem to uncover and discover relevant information (OECD, 2010).

In the interactive problem-solving environment, the problem solver has to use knowledge acquisition and knowledge application to apply non-routine actions to reach a certain goal state (Greiff, Holt, & Funke, 2013). Interaction is the key feature for the problem solver to discover the necessary information in an interactive problem-solving environment. Nowadays, in our daily environments, the number of interactive situations is increasing (Greiff et al., 2013), such as the interactions with software interfaces, hardware (e.g., smart phones, copiers, and vehicles) tools or other people (e.g., sales and clients). The common feature of these examples is the problem solvers' needs to actively interact with the systems/individuals to generate the new information they need, then they can build a problem representation and launch a goal-directed solution process (Greiff et al., 2013). Therefore, it can be said that IPS is closer to the real world than static (analytical) problem solving. Thus, *this study focuses on interactive problem solving.*

In some studies, interactive problem solving (problem solving in the MicroDYN approach) was also named as dynamic or complex problem solving (e.g. Greiff, Wüstenberg, & Funke, 2012). Ramalingam, Philpot, and McCrae (2017) argued that the term "dynamic" was not appropriate in this case since this term can imply that the task can change autonomously without direct action from the problem solver. The term "interactive", by contrast, can highlight that the problem solver-task interaction is required to discover undisclosed information. Similarly, Csapó and Funke (2017) also stated that dynamic problem solving should be called interactive problem solving in certain contexts (e.g. in the MicroDYN approach) in which interactive nature is emphasized. In order to highlight the interaction between the task and the problem solvers, we decided to use interactive problem solving in this study.

2.2.3 Well-defined & ill-defined problem solving

Problem solving is able to be classified as well-defined and ill-defined based on the types of unsolved problem. Well-defined problems have "a clear set of means for reaching a precisely described goal state" (Dörner, & Funke, 2017, p. 1153). This kind of problems contains clearly-defined initial and goal state, and the clear means of moving from the initial state to the goal states (Hong, 1998). For example, the majority of mathematics and physics problems belong to this category; the problem solvers clearly know the final state they need to reach, and understand what they need to do in order to move towards the goal state (e.g. students know they need to run the mathematical operations to get a correct answer to solve the given mathematical problem).

By contrast, ill-defined problems have no clear problem definition and goal state, the means of getting close and reach the goal state are not clear either (Dörner, & Funke, 2017), and the information available for the problem solver is usually incomplete and inaccurate or ambiguous (Allaire, & Marsiske, 2002; Hong, 1998). People need to solve ill-defined problem frequently in reallife. Some tasks such as building social connections with other people, making business decisions at work, belong to this category.These tasks do not have a clearly defined goal state, and it is uncertain which rules and principles are needed for the solution.

2.3 The development and different approaches to measure problem solving

2.3.1 The possibilities provided by computer-based assessment beyond paper-and-pencil testing

Currently, computer-based assessment is becoming increasingly popular. The research and development of computer-based assessment can go back almost three decades. Researchers started to explore the applicability of computer technologies (for the most common to most cutting-edge) for assessment purposes as early as the 1990s (Baker, & Mayer, 1999). From the 2000s, several large-scale international assessments have been conducted to explore the feasibility and the potential of computer-based assessment, such as the Programme for International Student Assessment (PISA), the Progress in International Reading Literacy Study (PIRLS), and the National Assessment of Educational Progress (NAEP) (Csapó, Ainley, Bennett, Latour, & Law, 2012; Fraillon Ainley, Gebhardt, & Schulz, 2013; OECD, 2010, 2014b). Csapó et al. (2012) presented that information-communication technology offers so many possibilities for modern education. They believed that computer-based assessment will replace traditional paper-and-pencil testing sooner or later. Similarly, Lau and Cheung (2010) stated that "more and more paper-and-pencil measures have their computerized counterparts". Csapó et al. (2012) summarized the advantages of computer-based assessment as follows:

"1) developing tests, making the questions easier to generate automatically or semiautomatically, share, review, and revise;

(2) delivering tests, obviating the need for printing, warehousing, and shipping paper;

(3) presenting dynamic stimuli like audio, video, and animation, making obsolete the need for the specialized equipment currently used in some testing programs that assess such constructs as speech and listening;

(4) scoring constructed responses on screen, allowing marking quality to be monitored in real time and potentially eliminating the need to gather examiners together;

(5) scoring some types of constructed responses automatically, reducing the need for human reading; and

(6) distributing test results, cutting the costs of printing and mailing reports." (Csapó et al., 2012, p. 149)

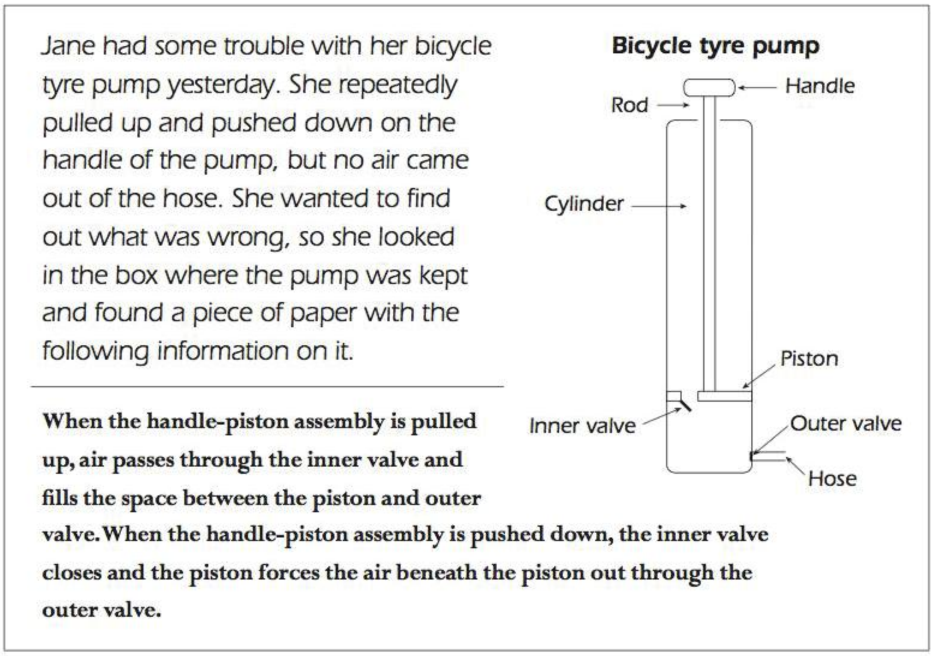
There are three ways of using computer to measure cognitive skills (Molnár et al., 2017). The first generation computer-based test was close to static paper-pencil based assessment. In most of cases, this kind of computer-based test does not have too many differences compared to the traditional paper-and-pencil assessment. The stimulus for this kind of test are still static, only the delivery platform has evolved to computers (Molnár et al., 2017; Pachler, Daly, Mor, & Mellar, 2010). The second generation computer-based test started to use new stimuli formats which cannot be realized in paper-and-pencil assessment, such as multimedia. Meanwhile, it also employs some techniques, including but not limited to constructed response, automatic item generation, and automatic scoring (Molnár et al., 2017; Pachler et al., 2010), thus boosts the assessment efficiency and makes some forms of assessment (e.g. self-assessment, adaptive testing, etc.) possible to be implemented. Many cognitive skill assessment types are designed in the scope of second generation computer-based testing (e.g. Kambeyo, & Wu, 2018; Molnár, Greiff, & Csapó, 2013).

However, second generation computer-based tests are still not able to implement the assessment of interactive problem solving, because they cannot realize “interaction”, the key feature of IPS. Therefore, the third generation computer-based test was created to “allow students to interact with complex simulations and dynamically changing items” (Molnár et al., 2017, p. 126). Currently, information and communication technologies provide new opportunities that can revolutionize the educational assessment and evaluation process (Csapó, Lőrincz, & Molnár, 2012). One of the major advantages is that the current ICT technology is able to support computer-based testing to provide a unique assessment environment, where both dynamic and interactive situations are available (Greiff et al., 2014). Due to this advantage, the third generation computer-based test allows researchers to design the interactive scenario which simulates the real world, and this kind of assessment basically cannot be realized by traditional assessment methods such as the paper-and-pencil test. The researchers in the PISA assessment project also pointed out that "the inclusion of interactive problems, in which students need to explore the (simulated) environment and gather feedback on the effect of their interventions in order to obtain all the information needed to solve a problem, was only possible by asking students to use a computer to complete the assessment" (OECD, 2014a, p. 30). In conclusion, computer-based assessment opens doors to explore features of 21st century skills and new constructs not yet studied (e.g. interactive problem-solving, and collaborative problem-solving), or to study previously researched knowledge and skill domains in an innovative and new environment.

2.3.2 Assessment of static problem solving

As introduced in section 2.2.2, static problem solving means that in the problem solving progress the tasks and the information stay static and non-changeable (Greiff et al., 2013). Static problem solving is the first kind of problem solving which was measured by psychologists (e.g. Ewert, & Lambert, 1932). Because the tasks in static problem solving are non-changeable, researchers can easily build the assessment environment even with the traditional paper-and-pencil methods.

In 2003, problem-solving was measured in the PISA assessment project for the first time (OECD, 2003). The paper-and-pencil test focused on static problem solving (it was named analytical problem solving in PISA 2003; OECD, 2003, 2004). Three types of problems were included in the assessment which were (1) decision making (required the problem solver to make his/her decision among alternative options to reach the goal); (2) system analysis (the problem solver needs to understand the structure of a complex system which contains several interrelated problems); and (3) fault finding (required the problem solver to use his/her causal understanding to find out why a system is not performing as it is expected) (OECD, 2003). The problems "were presented in real-life contexts from school and work situations or personal life" (Greiff et al., 2013, p. 75). The items provided static information and the problem solvers needed to analyze this information to solve the problem (Greiff et al., 2013). The relevant information was printed on paper. The systems and problems were described by words and pictures. The problem solvers were required to read and understand the text and use the relevant information to solve the problems (OECD, 2003). Fig. 2.4 shows an example item from the test. It is a fault finding item, which requires the problem solver to use the given information to figure out what the possible reasons could be for the failure of air to come from the hose.

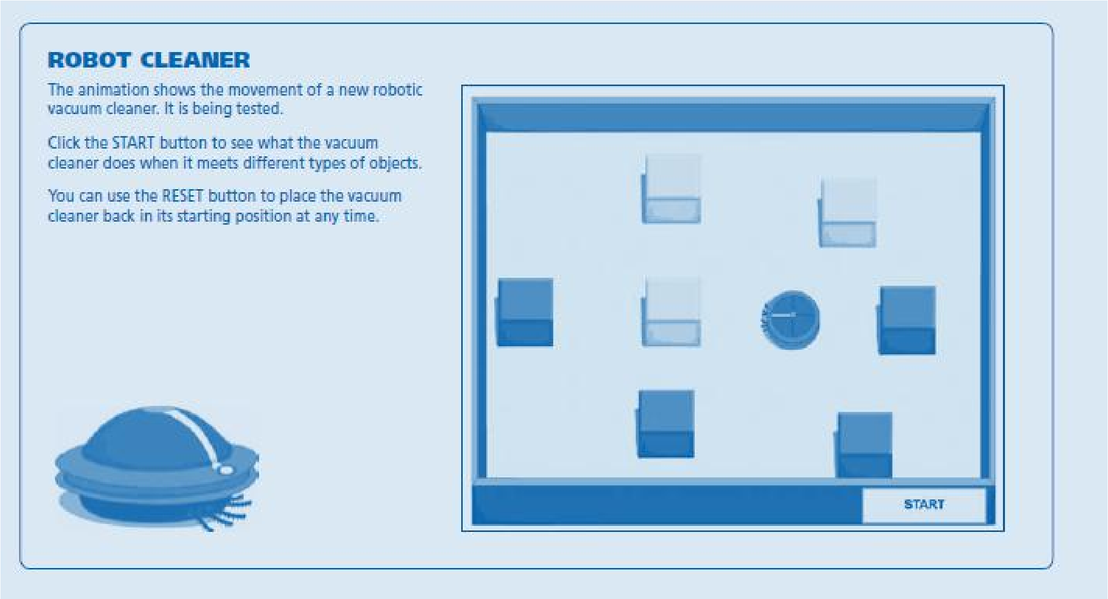


Jane had some trouble with her bicycle tyre pump yesterday. She repeatedly pulled up and pushed down on the handle of the pump, but no air came out of the hose. She wanted to find out what was wrong, so she looked in the box where the pump was kept and found a piece of paper with the following information on it.

When the handle-piston assembly is pulled up, air passes through the inner valve and fills the space between the piston and outer valve. When the handle-piston assembly is pushed down, the inner valve closes and the piston forces the air beneath the piston out through the outer valve.

Fig. 2.4 Example of a static problem-solving item in PISA 2003 (OECD 2004, cited by Greiff et al., 2013, p. 76)

After nine years, in 2012, PISA conducted the problem-solving assessment for the second time. The problems involved in the PISA 2012 problem-solving assessment were divided into two types, interactive and static. The medium of the test changed from paper-and-pencil testing to computer-based assessment. Fig. 2.5 shows a sample item for a static problem-solving item in PISA 2012. This item is named “Robot Cleaner”. It contains an animation as stimuli, which shows the behaviour of a robot vacuum cleaner in a room. Students can run the animation as many times as they need to obverse the behaviour of the cleaner to collect necessary information. This item belongs to the scope of second generation computer-based tests (see section 2.3.1). Even though the information was delivered to the students by animation and not on paper, this item can still be considered as “static” since students could not change the behaviour of the vacuum cleaner or aspects of the environment (OECD, 2014a).



ROBOT CLEANER

The animation shows the movement of a new robotic vacuum cleaner. It is being tested.

Click the START button to see what the vacuum cleaner does when it meets different types of objects.

You can use the RESET button to place the vacuum cleaner back in its starting position at any time.

Fig. 2.5 Example of a static problem-solving item in PISA 2012 (OECD, 2014a, p. 42)

With the development of problem solving research, researchers have started to consider the limitations of paper-and-pencil assessment methods. In real world environments, when dealing with problems, most problem solvers would have been likely to have an interaction with the problem, and use the interactions to gather relevant information. They would try out different options, see the responses from the system and use the information and experience to solve the problem (Greiff et al., 2013; Novick, & Bassok, 2005) However, in the static problem-solving assessment, the problem solver was not able to have the interaction with the problem like in real life. Another important point is that in paper-and-pencil assessments, as well as in the first and second generation computer-based testing, it is impossible to simulate the dynamic interaction between the problem and the problem solver as it occurs in real life. Even static problem-solving assessment has its own advantages; for example, it is easy to be designed and conduct, but still researchers were thinking about using new assessment methods and tools to overcome its limitations and problems, which resulted in the third-generation problem solving test used in the PISA 2012 assessment.

2.3.3 Assessment of interactive problem solving

The third-generation computer-based test techniques and theories opened the gate to measuring problem solving from a different perspective and applying interactivity between the test stimuli and the problem solver. Basically, there are two approaches used in interactive problem solving measures (based on Buchner, 1995):

(1) Computer-simulated microworlds similar to real world and composed of a large number of variables. Gardner, & Berry (1995) developed microworld models simulating a medical system. Moray, Lootsteen, & Pajak (1986) developed a microword about industrial production; and the famous scenario "Lohhausen" containing more than 2 000 variables must be also mentioned (Dörner, 1980; Dörner, Kreuzig, Reither, & Stäudel, 1983). In all of the cases ad hoc constructed scenarios are used to demonstrate individual differences in this approach (Greiff, & Funke, 2017). Due to its high complexity, this approach requires a long testing time.

(2) Simplistic, artificial, but still complex problems following certain construction rules, e.g. the DYNAMIS-approach (based on linear structural equation systems) designed by Funke (1992). Systematically constructed scenarios are used to demonstrate the effects of system attributes in this approach (Buchner, 1995).

From a psychometric point of view, regarding the time needed for the data collection, the reality and scalability of the problems, the validity and reliability of the test both approaches have several advantages and disadvantages (Greiff, 2012). The previous one mirrors more the problems present in real life, while the second one has favorable attributes from an assessment point of view. For example, the difficulty level of the Lohhausen problem is hard to set and its data collection lasts several hours, thus the problem is not really scalable and applicable in educational contexts, in school-based data collections, while the artificial problems in the DYNAMIS-approach are easy to scale and the administration of a single problem lasts only a few minutes. The single sub-solutions of the Lohhausen type problems are strongly connected to each other, while it is relatively easy to develop a test, having independent items with good psychometric indices in the DYNAMIS approach. The common feature of the two approaches is that both are realizable only in computer-based environment.

A sub-type of DYNAMIS approach, minimal complex systems, is proposed as an effective method of interactive problem-solving assessment (Funke, 2014; Greiff et al., 2012). The minimal complex systems contain most (or even all) of the features of a complex system (complexity, dynamics, polytely and intransparency; see Funke, 1991). But meanwhile, it contains low values for the parameters. Thus, it reduces the testing time to a minimum, especially if we compare it to the extremely difficult Microworlds.

Nowadays, the minimal complex system approach is widely accepted and employed by researchers for designing interactive problem-solving assessment (see Csapó, & Molnár, 2017; Greiff, Krkovic, & Hautamäki, 2015; Greiff, & Wüstenberg, 2014). It also has been applied in one of the most prominent large-scale assessment projects, in the PISA 2012 problem-solving assessment (OECD, 2014a).

We can distinguish two sub-approaches within the second approach: MicroFIN and MicroDYN, both using minimal complex systems (Greiff, & Funke, 2017). The MicroFIN approach uses the formalism of finite state automata (FSA) to model systems with discrete variables (Buchner, & Funke, 1993). The basic elements in MicroFIN include input signals, output signals, states and the state transitions (qualitative, e.g. from alarm mode to snooze mode, Funke, 2001) between them. A simple finite state automata is shown in Fig. 2.6. This finite state automata contains three states, z_0, z_1, and z_2. Two input signals, x_1 and x_2, can leads to state transitions and create three possible output signals, y_1, y_2, and y_3, depending on the reached state (Funke, 2001).

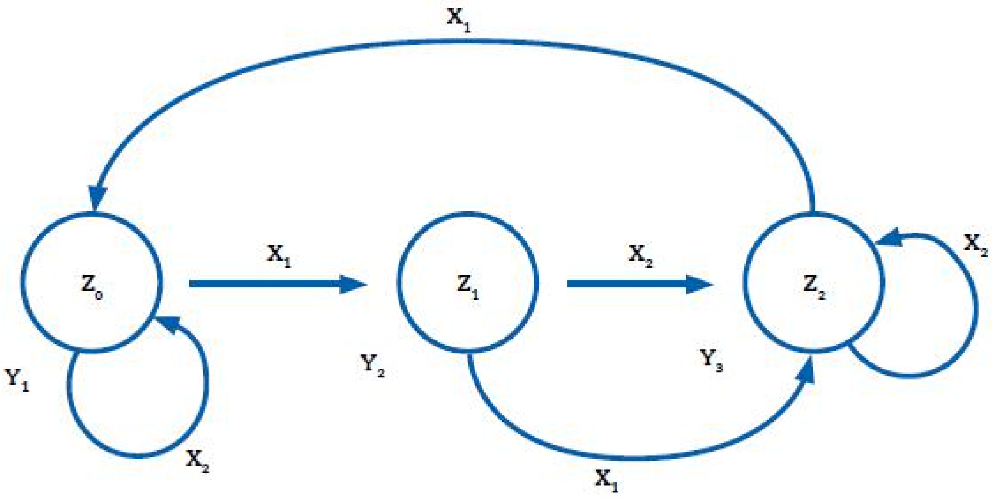
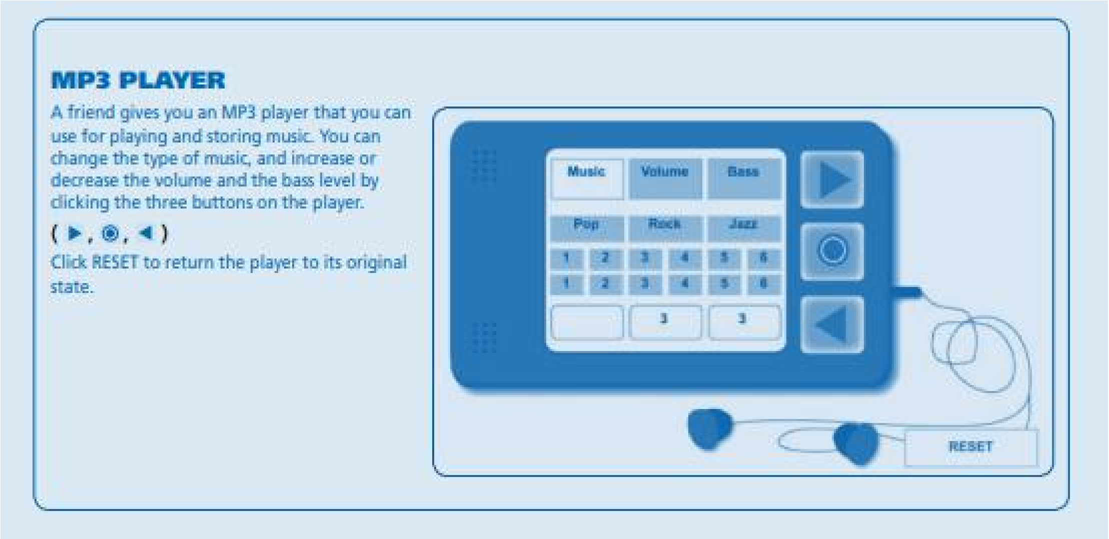


Fig. 2.6 A simple finite state automata using the MicroFIN approach (Greiff, & Funke, 2017, p. 99)

In the MicroFIN approach, the problem solver has to gather information by changing the state (0 or 1) of the input variables and finding out how these are related (state-transitions) to the output variables (knowledge acquisition). In the second half of the problems, they have to reach the given goal (knowledge application). In contrast to the MicroDYN problems, in problems developed on the MicroFIN approach numbers and quantitative values do not play a role. Exploration in the MicroFIN problem environment requires a step-by-step analysis of the state transitions. In each of the problems there are at least as many exploration steps required from the problem solvers as there are different state transitions. For instance, the “MP3 Player” task in the PISA 2012 problem-solving assessment is a typical MicroFIN problem (see Fig. 2.7). The task tells students that they have been given a MP3 player by a friend (OECD, 2014a). The system contains three clickable buttons (input – with 2 states: 0 and 1). After the buttons are clicked, the system changes (state transitions) and provides output information. Students have to interact with the system by changing the states of the input variables to find out how it works, and answer questions about the internal rules of the system, or set the system to required state (e.g. Rock, Volume 4, Bass 2) (OECD, 2014a).



MP3 PLAYER

A friend gives you an MP3 player that you can use for playing and storing music. You can change the type of music, and increase or decrease the volume and the bass level by clicking the three buttons on the player.

(▶, ◉, ◀)

Click RESET to return the player to its original state.

Fig. 2.7 Interactive problem-solving task “MP3 Player”: Stimulus information (OECD, 2014a, p. 35)

Unlike MicroFIN, MicroDYN is also based on multiple complex systems within the linear structural equation (LSE) framework (Funke, 2001). Moreover, problems in the MircoDYN approach use variables in a quantitative way. In this approach, the relations between input variables and output variables can be described by linear structural equations. Each MicroDYN task contains up to three input variables (represented by A, B, and C), which are related to up to three output variables (represented by X, Y, and Z, see Fig. 2.8; Greiff et al., 2013).

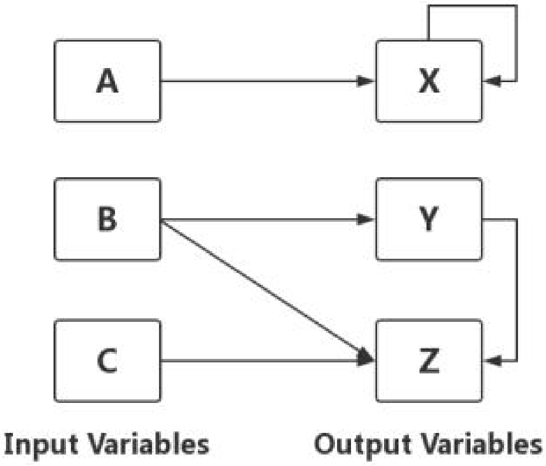
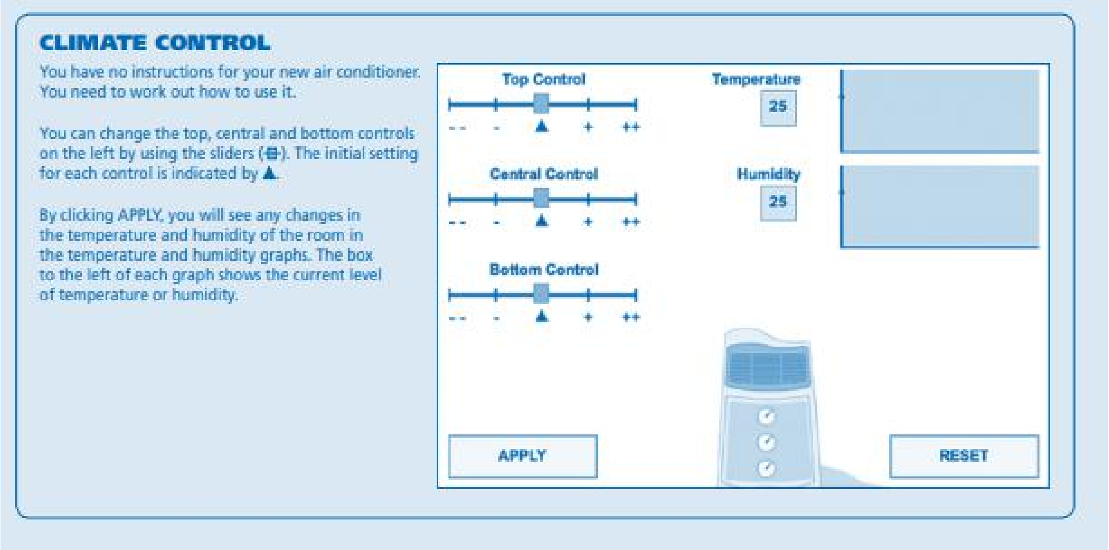


Fig. 2.8 Structure of a typical MicroDYN task with direct effects, side effect and eigendynamics (Greiff et al., 2013)

The relations between the input and output values are various. Causal relations between input variables and output variables are called direct effects, while the effects originating and ending with output variables are known as indirect effects (Greiff et al., 2013). Indirect effects can involve an output variable influencing another output variable (side effects, see Fig. 2.8: Y to Z) or influence itself (eigendynamics, see Fig. 2.8: X to X) (Greiff et al., 2013). Similarly to MicroFIN, the assessment contains two phases, knowledge acquisition and knowledge application. In the knowledge acquisition items, students have to interact with the system by changing the values of input variables, and observe the corresponding changes of output variables, so as to find out the relationships between input and output variables. In the knowledge application part, students have to solve the given problems by assigning appropriate values to the input variables, to make the output variables reach the required range (Molnár, & Csapó, 2018).

The "Climate Control" problem in the PISA 2012 assessment is based on the MicroDYN approach (see Fig. 2.9). In this task, students have to figure out how to use an air conditioner without instruction. The virtual air conditioner contains three controls (input); students can operate these controls to vary temperature and humidity levels (output), and the output information is demonstrated in both numeric and graphical forms (OECD, 2014a). In the first phrase (knowledge acquisition) of the problem, students have to find out which control variables have an effect on the output variables, more specifically, which input variable is connected to which output variable(s). In the second phase of the problem solving process, students have to operate the system by reaching the given target values of the output variables (temperature and humidity) within four steps and a given time frame (knowledge application) (OECD, 2014a).



CLIMATE CONTROL

You have no instructions for your new air conditioner. You need to work out how to use it.

You can change the top, central and bottom controls on the left by using the sliders (⊟). The initial setting for each control is indicated by ▲.

By clicking APPLY, you will see any changes in the temperature and humidity of the room in the temperature and humidity graphs. The box to the left of each graph shows the current level of temperature or humidity.

Fig. 2.9 Interactive Problem solving task "Climate Control": Stimulus information (OECD, 2014a, p. 37)

MicroDYN and MicroFIN have some common features. Both approaches allow for developing problems in the same structure consisting of knowledge acquisition and knowledge application, and sharing one principle, namely "the identification of an unknown dynamic device and the subsequent control of this device" (Greiff, & Funke, 2017, p. 100). However, "MicroDYN tasks yield the advantage of being homogenous and adding up to a narrow but reliable test set (low bandwidth, high fidelity), MicroFIN tasks are closer to real life and more heterogeneous in the skills they test (high bandwidth, low fidelity)." (Greiff, & Funke, 2017, p. 100) The feasibility, reliability, and effectiveness of these two approaches are proved by a large number of empirical studies (e.g., Csapó, & Molnár, 2017; Greiff, & Wüstenberg, 2014; Neubert, Kretzschmar, Wüstenberg, & Greiff, 2014; OECD, 2014a).

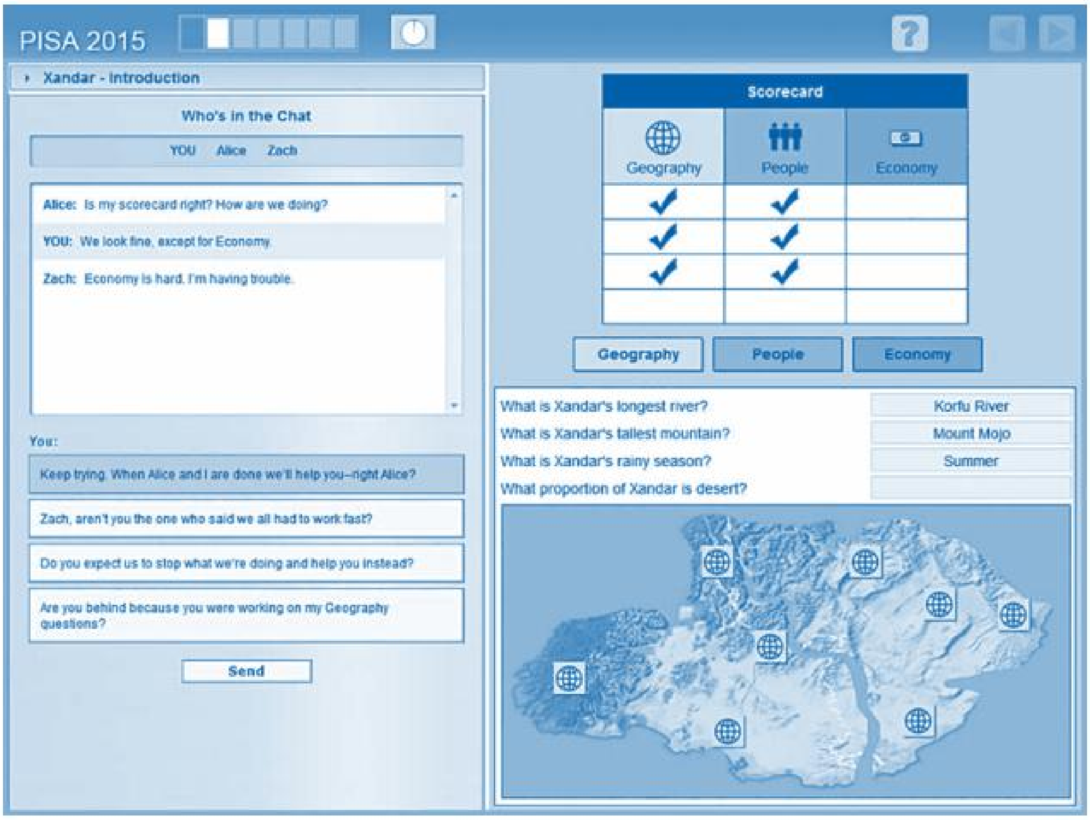
However, there are still some criticisms about MicroDYN, MicroFIN and other interactive problem-solving assessment instruments. Most of these instruments' "system behavior that are completely predictable and stable" (Dörner, & Funke, 2017, p. 1153), so it inevitably has some drawbacks when simulating the ill-defined problems which people are facing in real life. For example, Dörner and Funke (2017) pointed out that the problem-solving strategies in the MicroDYN approach are the simplistic version of the real-life problem-solving rules and principles. They believed it is hard to imagine how people apply the strategies which are the most effective in the MicroDYN approach to solve the real-life, complex, ill-defined problems. Therefore, instruments like MicroDYN and MicroFIN, even though their reliability and effectiveness have been proved by many studies, still have the great possibility to be improved.

2.3.4 Beyond individual testing of problem solving: assessment of collaborative problem solving

Collaborative problem solving, a hotly debated sub-type of problem solving, extends the problem-solving research to a new dimension. Collaborative problem solving "is a complex process that requires participants to externalize their individual problem-solving processes and to co-ordinate these contributions into a coherent sequence of events" (Hesse, Care, Buder, Sassenberg, & Griffin, 2015, cited by Care, & Griffin, 2017, p. 229). Moreover, researchers in PISA defined collaborative problem solving as "the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution." (OECD, 2017b, p. 6).

Technically, collaborative problem solving can be seen as one kind of interactive problem solving, but it contains unique interactions between problem solvers. In collaborative problem solving, problem solvers are not only required to interact with the task, but also with other problem solvers (OECD, 2013). Apart from the cognitive skills required in the individual interactive problem-solving process, collaborative problem solving also needs some social skills (Care, & Griffin, 2017) such as the ability to recognize the points of view of others in a group; to effectively share knowledge, experience, and expertise with each other; to identify what needs to be contributed to the collaboration work; and the ability to build and develop group knowledge and understanding as a team member (Greiff et al., 2014). The collaborative problem-solving assessment is also an attempt to assess students' performance and behaviour for solving ill-defined problem, because in the collaborative problem-solving assessment, "team members can neither control nor predict what other team members will do" (OECD, 2017b, p. 2).

In our modern society, social collaborative skills and teamwork skills are becoming more and more important (Autor, Levy, & Murnane, 2003). Collaboration between students also has been considered as a significant method for students to construct their knowledge (Vygotsky, 1978). Introducing collaboration into educational activities is one of the main tasks for the online education designers (Worthington, & Wu, 2015). Therefore, collaborative problem solving has been proposed to fulfill this kind of requirements. As one of the important 21st century skills, collaborative problem solving is getting more attention. For example, OECD used collaborative problem-solving assessment to replace individual problem-solving assessment in PISA 2015 (OECD, 2017b). Fig. 2.10 shows one of the PISA sample items from the field of collaborative problem solving.



Alice: Is my scorecard right? How are we doing?

YOU: We look fine, except for Economy.

Zach: Economy is hard. I'm having bouble.

Keep trying. When Alice and I are done we'll help you–right Alice?

Zach, aren't you the one who said we all had to work fast?

Do you expect us to stop what we're doing and help you instead?

Are you behind because you were working on my Geography questions?

Scorecard		
Geography	People	Economy
✓	✓	
✓	✓	
✓	✓	

What is Xandar's longest river?	Korfu River
What is Xandar's tallest mountain?	Mount Mojo
What is Xandar's rainy season?	Summer
What proportion of Xandar is desert?	

Fig. 2.10 Sample problem from the PISA 2015 computer-based collaborative problem solving assessment (OECD, 2017b, p. 62)

Compared with individual interactive problem-solving assessment, collaborative problem-solving assessment is much harder to be designed because its assessment variables, in terms of social skills, are more complex and harder to be controlled (Krkovic, Pásztor-Kovács, Molnár, & Greiff, 2014). The interactions between problem solvers are the keystone and difficulty in the collaborative problem-solving assessment design. There are two major methods to realize the interaction, these are the "Human-to-Human" and the "Human-to-Agent" approaches (Care, & Griffin, 2017; Pásztor-Kovács, 2018).

The "Human-to-Human" approach means that students interact with each other directly, and this approach can help the students to become more engaged and motivated to collaborate with their peers (Rosen, & Tager, 2013). But Rosen and Tager (2013) also pointed out that the "Human-to-Human" approach may not provide "enough opportunity to cover variations in group composition, diversity of perspectives and different team member characteristics in controlled manners" (Rosen, & Tager, 2013, p. 7), and that will significantly impact the analysis work for researchers.

In the "Human-to-Agent" approach, students will interact with each other by interacting with an agent. In most cases, the agent is the computer (Krkovic, Wüstenberg, & Greiff, 2016). The "Human-to-Agent" approach can ensure that students act as team members with different characteristics relevant to different interactive problem-solving situations (Rosen, & Tager, 2013). It can help researchers to realize stronger control, and it can provide more materials (e.g. chat record, log-file, etc.) for researchers to analyze and assess students' problem-solving process. However, Pásztor-Kovács (2018) argued that the Human-to-Agent collaborative problem-solving assessment is not able to provide test environments with realistic quality.

There are obvious differences between natural human and computer agents, as for a computer agent (based on the current technology) it is almost impossible to accurately imitate all human behaviours (Pasztor-Kovács, 2018). Care and Griffin

(2017) admitted that "it is still to be determined whether the H2A [Human-to-Agent] approach is consistent with, or different from, a construct defined by the analysis of H2H [Human-to-Human] interactions" (Care, & Griffin, 2017, p. 237). Therefore, further empirical and theoretical studies are still needed to evolve the collaborative problem-solving assessment methodology and design.

2.3.5 The challenges and potential of analyzing log-file data in problem solving environment

One of the greatest advantages computer-based assessment brings to problem-solving researchers is the availability and possibility of log-file analysis. In the context of computer-based assessment, students' operations and interactions can be recorded by the assessment system. The record is called a log-file (Zoanetti, & Griffin, 2017). These records usually contain the keystrokes and mouse events (e.g. typing, clicking, dragging, dropping, cursor movement) (Zoanetti, & Griffin, 2017), or even eye tracking, head movement, or facial expressions (Krkovic et al., 2014; Van Gog, Paas, & Van Merriënboer, 2005). In most cases, each action would be recorded with a corresponding time stamp. These timestamped interaction data have been defined as the log-file data (Arroyo, & Woolf, 2005; Zoanetti, & Griffin, 2017).

The log-file data provides new possibilities for researchers to discover and conduct new forms of data analysis in educational assessment. In the majority of the classical educational assessments, the data analyses are only based on the answer data, particularly on the final performance data of the students. Log-file analyses, thus the log-file data are able to provide much richer information. It can be used to complete, deepen, and testify the analysis results which they received from the performance data. In the interactive problem-solving environment, log-file data have been used to explore students' problem-solving behaviours (Tóth, Rölke, Goldhammer, & Barkow, 2017), exploration strategies (Molnár, & Csapó, 2018), test-taking motivation (Greiff, Wüstenberg, & Avvisati, 2015), thus help to gain a much deeper understanding of how the problem solvers interact with the problems, how they behave during the problem solving and test taking process.

However, unlike the performance data, the log-file data usually contain a certain amount of redundant and/or irrelevant data. An effective data cleaning is essential at the beginning of the log-file data analysis. Moreover, there are no certain or commonly used methods available for the log-file analyses. Based on the different research purposes, researchers should make their own decisions concerning what kind of log-file data they should collect, and what kind of log-file analyses methods they should employ (Zoanetti, & Griffin, 2017). The log-file data analysis can be considered as “an interdisciplinary challenge requiring subject matter experts, item developers, psychometricians and computer scientists to work together to extract, aggregate, model and interpret the data appropriately” (Tóth et al., 2017, p. 194).

2.4 Factors influencing problem solving achievement in educational context

According to section 2.2, the problem-solving process can be impacted by three aspects which are the problem solver, the task, and the environment (Frensch, & Funke, 1995). This study mainly focuses on how problem solvers' personal features can impact the problem-solving process. Therefore, in order to achieve this target, this study will ask all the participants to do the same tasks in a controlled environment.

The impact from the problem solver is mainly relevant with three main categories which are memory contents, information processing, and non-cognitive variables (see Fig. 2.11) (Frensch, & Funke, 1995). The following part will illustrate which factors will be included in this study from each of the three categories.

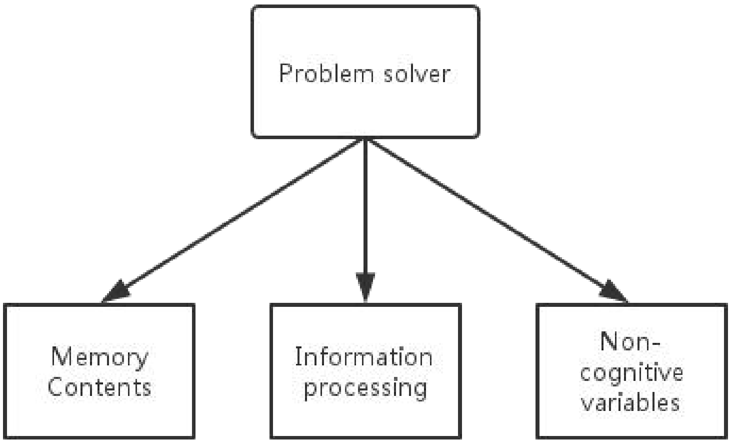


Fig. 2.11 Factors of three categories for problem solver (Frensch, & Funke, 1995)

2.4.1 Working memory

Section 2.2.1 introduced that problem-solving assessment should try to avoid the impact of examinee's knowledge storage in long-term memory. Base on Frensch and Funke's (1995) theory, in the problem solving assessment everyone should be a novice (Frensch, & Funke, 1995). Thus, this section will not focus on problem solvers' long-term memory, but on their working memory.

Working memory refers to a "brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks" (Baddeley, 1992, p. 556). Working memory is a concept which belongs to memory in the broad sense.

Baddeley (1992) proposed a model to indicate the components and structure for working memory (Fig. 2.12). Based on Baddeley's (1992) theory, working memory has three components: visuospatlal sketch pad, central executive, and phonological loop. The phonological loop and the visuospatlal sketch pad responsible for processing the acoustic and visual information, respectively. To be more detailed, "the articulatory or phonological loop was assumed to be responsible for maintaining speech-based information, [...] whereas the visuospatial sketch pad was assumed to perform a similar function in setting up and manipulating visuospatial imagery" (Baddeley, 1992, p. 557). Central executive is the core component for working memory, it is described as an attentional-controlling system, meanwhile it can also coordinate visual and acoustic information, and exchange information with the long-term memory (Baddeley, 1992).

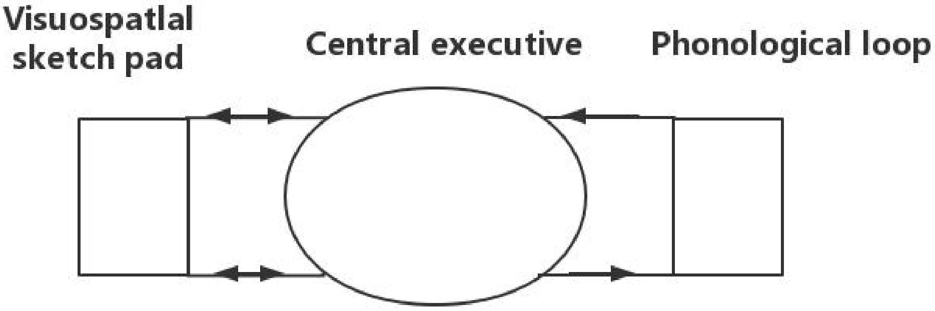


Fig. 2.12 The working memory model (Baddeley, 1992, p. 557)

Some previous studies pointed out that the capacity of working memory can influence one's reasoning process (Greiff et al., 2015; Söderqvist Bergman Nutley, Ottersen, Grill, & Klingberg, 2012). Furthermore, working memory proved to be deeply connected with intelligence (Ackerman, Beier, & Boyle, 2005; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). Kyllonen and Christal's (1990) study showed a strong positive correlation (r=.90) between working memory capacity and fluid intelligence. Moreover, the problem solver needs to use working memory to store or transform some essential information during the problem-solving process (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). According to Sweller (1988), working memory could contain the following pieces of information: (1) the problem-solving strategies which potentially be used; (2) the known variables such as the current and previous strategies the solver selected and the states which have occurred; and (3) the unknown variables which are needed in the information processing approach. The information in working memory can be transformed (e.g. from unknown to known), added, or deleted (Sweller, 1988). Therefore, the capacity of one's working memory is influencing their performance in cognitive processes which include problem solving (Fukuda, & Vogel, 2009). Passolunghi and Siegel (2001) reported that the deficit in the capacity of working memory will cause difficulties during the information processing approach and may lead to failure in problem solving.

Previous empirical studies have proved the connections between working memory and problem solving. A small-scale assessment (N=69) conducted by Andersson (2006) found that working memory has significant correlation (r>.25) with problem solving in specific domains (e.g. Mathematics). Swanson and Beebe-Frankenberger's (2004) study also reported that significant correlation can be detected between first-third grade students' working memory and their problem solving skill regarding word processing (r=.54). By contrast, Bühner, Kröner, and Ziegler's (2008) study focused on the relationships between working memory and domain-general problem solving. Their study drew a sample of N=144 undergraduate students. It measured working memory as storage in the context of processing and coordination, and problem solving as two dimensions rule knowledge (generally the same as knowledge acquisition) and knowledge application. The structural equation modeling results indicated that working memory had significant predicting power for both rule knowledge (β=.40) and knowledge application (β=.40) the two dimensions of problem solving. Thus, based on the theoretical and empirical studies, working memory is considered as one of the important influencing factors of problem solving.

2.4.2 Information processing

Nowadays, several studies have been published using the information processing approach. Inductive reasoning, combinatorial reasoning skills, and creativity have been most frequently mentioned as component skills of problem solving.

Reasoning skills. Reasoning is a kind of general thinking skill (Pellegrino, & Glaser, 1982), normally "understood as a generalized capability to acquire, apply and transfer knowledge" (Molnár et al., 2017, p. 127). It had significant influence in almost all higher-order cognitive skills and processes (Csapó, 1997; Söderqvist et al., 2012), which include knowledge acquisition and knowledge application (Bisanz, Bisanz, & Korpan, 1994; Hamers, De Koning, & Sijtsma, 2000; Molnár et al., 2017) and the general problem-solving process (Molnár et al., 2013; Tomic, 1995). In this study, two major reasoning skills, combinatorial reasoning and inductive reasoning, have been chosen for analysis because their influence on problem solving has been discussed most frequently in previous studies.

According to Adey and Csapó's (2012) definition, combinatorial reasoning is the process of creating complex constructions out of a set of given elements that satisfy the conditions explicitly given or inferred from the situation. In this process, some cognitive operations, such as combinations, arrangements, permutations, notations, and formulae will be employed (English, 2005). The capacity of creating constructions or combinations can be defined as three levels, which are (1) "use random listing procedures, without trying to find a systematic strategy", (2) "use trial and error, discovering some empirical procedures with a few elements", and (3) "discover systematic procedures of combinatorial construction" (Batanero, Godino, & Navarro-Pelayo, 1997, p. 182). Dubois (1984) classified six basic types of the combinatorial models:

"1. Ordered distributions of different objects in different containers.

2. Ordered distributions of different objects in identical containers.

3. Non-ordered distribution of different objects in different containers.

4. Non-ordered distributions of different objects in identical containers.

5. Distributions of identical objects in different containers (because the objects are identical, the order is irrelevant).

6. Distributions of identical objects in identical containers (order is irrelevant)." (Dubois, 1984, cited by Batanero, Godino, & Navarro-Pelayo, 1997, p. 183)

Combinatorial reasoning is one of the basic components of formal thinking (Batanero, Godino, & Navarro-Pelayo, 1997). The relationship between combinatorial reasoning and problem solving has been frequently discussed. English (2005) proved that combinatorial reasoning has essential meaning in several types of problem situations, such as problems requiring the systematic testing of alternative solutions. Moreover, Newell (1993) pointed out that combinatorial reasoning is applied in some key activities of the problem-solving information processing such as strategy generation and application. Its functions include, but are not limited to, helping problem solvers to discover relationships between certain elements and concepts, promoting their fluency of thinking when they are considering different strategies

(Csapó, 1999), and identifying all possible alternatives (OECD, 2014a). Moreover, even if problem solvers prefer a trial-and-error method in the interactive problem-solving environment, higher-level combinatorial reasoning skills can help them to summarize experience of failure and organize possible solutions. However, of the majority of studies discussed the relationship between combinatorial reasoning and problem solving from a theoretical point of view (see Batanero, Godino, & Navarro-Pelayo, 1997; English, 2005), and there is still a lack of empirical studies which directly describe the influence of combinatorial reasoning to problem solving in the quantitative way. This study is going to fill this gap.

Inductive reasoning is another reasoning skill which has been most frequently discussed by literature. Currently it does not have a universally accepted definition. Molnár et al. (2013) described it as the cognitive process of acquiring general regularities by generalizing single and specific observations and experiences, while Klauer (1990) defined it as the discovery of regularities that relies upon the detection of similarities and/or dissimilarities concerning the attributes or relations to or between objects. Sandberg and McCullough (2010) provided a general conclusion of the definitions of inductive reasoning, that it is the process of moving from the specific to the general.

Csapó (1997) pointed out that inductive reasoning is a basic component of thinking, and it forms a central aspect of intellectual functioning. Some studies have also discussed the role of inductive reasoning in the problem-solving environment: such as Mayer (1998) stated inductive reasoning will be applied in information processing during the process of solving general problems; Gilhooly (1982) also pointed out inductive reasoning also plays a key role in some activities in the problem-solving process like hypothesis generation, and hypothesis testing. Moreover, inductive reasoning’s influence on both knowledge acquisition and knowledge application has been analyzed and demonstrated in earlier studies (Hamers, De Koning, & Sijtsma, 2000; Molnár et al., 2013).

As introduced in section 2.2.2, interactive problem-solving tests based on the MicroDYN approach require the problem solver to discover, understand, and apply the relations between input and output variables. And the required ability largely overlap with the structure of inductive reasoning (Fig. 2.13) which was introduced by Klauer (1996). Therefore, theoretically speaking, the development level of inductive reasoning has the great possibility of influencing one's problem-solving achievement in the MicroDYN approach.

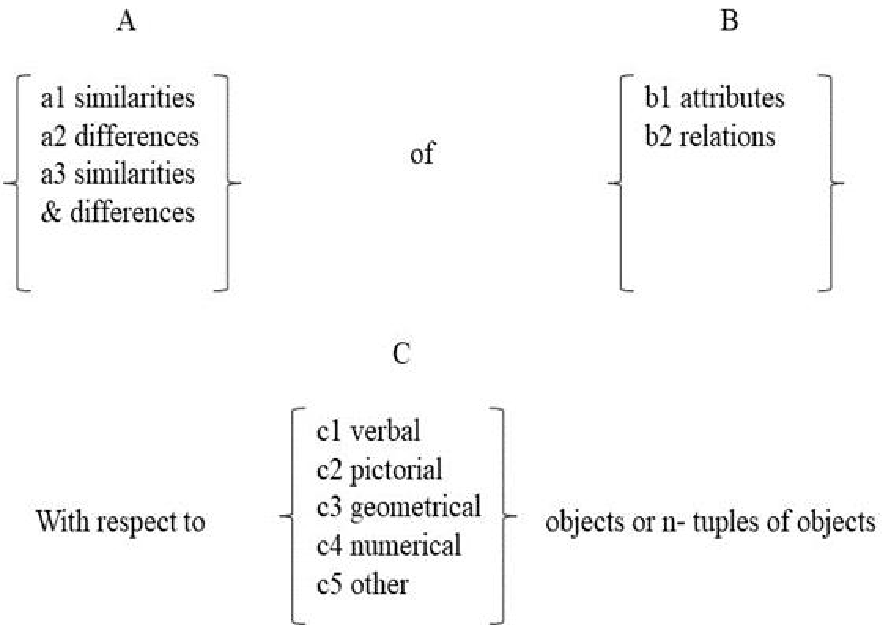

A
a1 similarities
a2 differences
a3 similarities
& differences
of
B
b1 attributes
b2 relations
C
With respect to
c1 verbal
c2 pictorial
c3 geometrical
c4 numerical
c5 other
objects or n- tuples of objects

Fig. 2.13 Structure of inductive reasoning (Klauer, 1996, p. 38)

Empirical studies also provided the evidence that inductive reasoning and problem solving are connected. Based on the results of a large-scale assessment (N=2,769), Molnár et al. (2013) showed that inductive reasoning significantly correlated with 9-17 years old students' domain-general problem solving achievement (r=.44-.52). Greiff et al. (2015) conducted a large-scale assessment project (N=2 021) in Finland to explore the connections between fluid reasoning skills and domain-general complex problem solving. The study measured fluid reasoning as a two dimensional model which consisted of deductive reasoning and scientific reasoning and included inductive thinking processes (Greiff et al., 2015). The results drawing on on structural equation modeling indicated that fluid reasoning which were partly based on inductive reasoning had significant and strong predictive effects on both knowledge acquisition (β=.51) and knowledge application (β=.55), the two dimensions of problem solving. Such studies have suggested that inductive reasoning might be one of the component skills for problem solving.

Creativity. There is no unique, universally accepted definition of creativity. According to Robinson (2009, p. 67), creativity means the ability to take "the process of having original ideas that have value". Amabile (2012, p. 2) defined creativity as “the production of ideas or outcomes that are both novel and appropriate to some goal”. Pásztor, Molnár, and Csapó (2015, p. 33) summarized the definitions of creativity discussed in some earlier studies (Piffer, 2012; Plucker, Beghetto, & Dow, 2004; Runco, & Jaeger, 2012; Simonton, 2012) and stated that most of the definitions claimed creativity to be the ability to take acts which can “result in output which is novel and has some sort of value”.

Creativity is one of the most important 21st-century skills (Binkley et al., 2012). It plays an important role in people's work, study and life in the 21st century (Pásztor, Molnár, & Csapó, 2015).

Creative ideas are often required in the problem-solving process (Pásztor, Molnár, & Csapó, 2015). Creativity can help the problem solver by creating or selecting the proper problem solving strategies. According to the second level analyses of the PISA results, OECD (2014a) published that the lack of creativity may lead the problem solver to arrive to less successful problem solutions if these are problems they are not familiar with or have never met before. Creativity can also help people to find and locate the problem in their reallife (Okuda, Runco, & Berger, 1991). To sum up, the result of former empirical studies suggest that creativity might belong to the thinking skills which has a correlation with students’ problem-solving skills and strategies (Binkley et al., 2012; Herrmann, 1995).

Some empirical studies have also discovered the connections between creativity and problem solving. Casakin (2007) studied 65 undergraduate architecture students. Significant and strong correlations were detected between these students’ general level of creativity and problem solving skills in the field of architecture design (r=.50-.77). Basadur, Wakabayashiand and Graen (1990) conducted an experiment with a group of company employees (N=156). They arranged a training session in creative thinking and observed the participants' problem-solving strategies and performance before and after the training. Results indicated that participants became better problem solvers with more active attitude after the creative thinking training program (ANOVA F=3.8, $p<.05$).

2.4.3 Non-cognitive variables

Problem solving achievement can also be impacted by some non-cognitive factors. The factors can be divided into two categories which are personality (e.g. motivation) and social (e.g. parents' education) (Frensch, & Funke, 1995) factors. There are several studies focusing on the influence of non-cognitive variables on problem-solving. The following part of the dissertation illustrates some typical cases.

Test-taking motivation. Motivation has an important influence on people's mental and practical activities (Vroom, 1964; Hackman, & Oldham, 1976). Frensch and Funke (1995) reported that motivation can "influence the way people attempt to understand and control a dynamic system" (Frensch, & Funke, 1995, p. 54), and the success of the problem solving process is also impacted by the problem solver's motivation. In most of the cases, assessments which are focusing on students' level of problem solving skills are low-stakes testing, thus the results have no or little consequence for the test-takers' future and school success (Cole, David, & Tiffany, 2008). Most of the projects including problem-solving assessment are used for accountability purposes by monitoring students' problem solving achievement to compare the efficacy of the different educational systems in the field of 21st century skills. Sometimes the students even cannot get personal feedback after the data collection.

Several studies indicated that in low-stakes testing situations students might not be motivated enough to do their best (Finn, 2015; Mislevy, 1995). Theoretically speaking, test-takers' performance in a problem-solving assessment might be lower than their true level if they have low test-taking motivation. This situation will lead to an underestimated result, damaging the validity of the assessment in a certain level

(Knekta, 2017). Therefore, we concluded that it is necessary to monitor students' test-taking motivation during the problem solving process.

Age and gender. Students' problem-solving skills develop with age. Fig. 2.14 shows the developmental curve of complex problem solving based on a large scale assessment conducted in Hungary (Molnár, Greiff, & Csapó, 2013). The development spans several years offering opportunities for the enhancement and can be described with a logistic curve. The fast-growing phase, the most sensitive period of enhancement was observed between Grades 5 and 8 (Age 12-15), thus this is the most effective time to enhance students' problem solving skills.

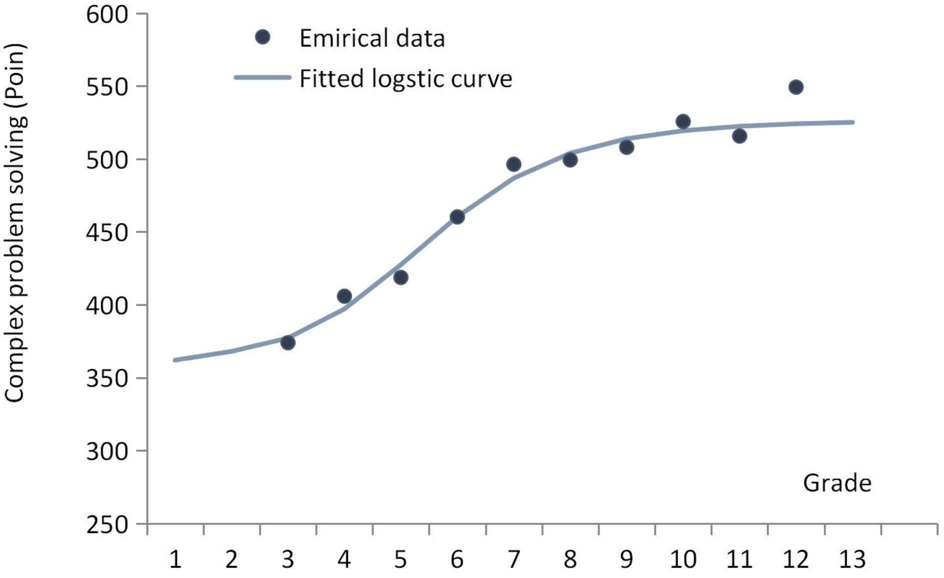


Fig. 2.14 Developmental curve of complex problem solving (Molnár, 2016)

Students with different genders have different performance in problem-solving assessments (Wüstenberg et al., 2014). According to the PISA 2012 data, the average difference between 15-year-old girls' and boys' problem solving skills is 7 score points (boys perform better) (OECD, 2014a). Similarly, some studies also pointed out males' considerable advantage regarding problem-solving (e.g. Wittmann, & Hattrup, 2004; Wittmann, & Süß, 1999). Wittmann and Hattrup's (2004) study puts forward a point of view that males may employ more efficient strategies while dealing with complex problems, thus gain the advantage in the assessments. Moreover, the PISA 2012 results showed that girls performed better on items measuring planning and executing aspects, while boys performed better on items which measuring the representing and formulating aspects.

Family situations. Family situation such as the parents' education and occupation are impacting children's educational achievements (Dubow, Boxer, & Huesmann, 2009; Sewell, & Shah, 1968). Students with "home advantages tend to enjoy stronger support for educational and career aspirations and becoming a strong learner." (Artlet, Baumert, Julius-McElvany, & Peschar, 2003, p. 60) Artelt et al. (2003) reported that students with better family background showed greater interest in some school subjects such as reading and mathematics, and were more willing to pursue a higher education degree. Moreover, students with better family background also showed higher self-confidence in school activities, and this situation remarkably benefited their learning (Artelt et al., 2003). Therefore, there is reason to believe that a better family situation could be an advantage for students regarding their development of cognitive skills, including the problem-solving skills. Some empirical results supported this assumption. For example, in the PISA 2012 problem-solving assessment, students with better socio-economic status showed remarkably higher performance in the problem-solving assessment (OECD, 2014a). Similarly, Csapó and Molnár's (2017) study indicated that students' mother's education level is significantly correlated with their performance in the knowledge acquisition phase of problem solving ($r=.08$, $p<.01$).

Learning strategies. Learning strategies are "the plans students select to achieve their goals: the ability to do so distinguishes competent learners who can regulate their learning"(Artlet et al., 2003, p.13) Students' usage and preferences of learning strategies have been proven to affect not only their learning efficiency, but also the way in which they process information (Riding, & Rayner, 2013). Artelt et al. (2003) also pointed out that when applying a specific type of learning strategy, such as memorization or elaboration strategy, corresponding information processing skills will be employed. Base on this fact, if students use different learning strategies in their daily learning activities, they might have different development level of their information processing skills, which influences their problem-solving achievement.

Moreover, some learning strategies (e.g. control strategy, Artelt et al., 2013) reflect students' conscious regulation of learning, and the use of this group of strategies has the possibility to influence students' behaviour in not only their daily learning activities, but also in experimental research (Willoughby, &Wood, 1994). As an example, Molnár et al. (2017) discussed that students' learning strategies influence their adoption of exploration strategies during the problem-solving process. Some empirical studies have found connections between the use of learning strategies and problem-solving achievement. For example, Csapó and Molnár (2017) assessed a group (N = 1,468) of Hungarian first-grade university students' problem-solving achievement and learning strategies. The results showed that students' problem-solving achievement had positive correlations with elaboration strategies (r=.07-.12, p<.05), and negative correlations with memorization strategies (r=-.07--.24, p<.05).

2.5 Students' exploration strategies in interactive problem solving in the MicroDYN approach

Wüstenberg et al. (2012) stated that creation and implementation of strategic exploration are the core actions of the problem-solving task. Effective information exploring and generating is the key to successfully solve a problem. Wittmann and Hattrup (2004) illustrated that "riskier strategies [create] a learning environment with greater opportunities to discover and master the [problem's] rules and boundaries" (p. 406). Thus, when gathering information about a complex problem, there may be differences between the exploration strategies in terms of efficacy. The MicroDYN scenarios, as the simplification and simulation of the real world problem-solving context, will also be influenced by the adoption and implementation of exploration strategies.

The isolated variation exploration strategy has been frequently discussed and mentioned in IPS research. The isolated variation strategy can also be referred to as the "Vary-One-Thing-At-A-Time-Strategy" or VOTAT (Vollmeyer, Burns, & Holyoak, 1996). To be more specific, with the VOTAT strategy "the problem solver systematically varies only one input variable, while the others remain unchanged. This way, the effect of the variable, that has just been changed, can be observed directly by monitoring the changes in the output variables" (Molnár, & Csapó, 2018, p. 2). The VOTAT strategy contains several types of strategies, which will be discussed in the following sections.

Some previous studies have indicated that students who are able to apply VOTAT are more likely to achieve higher performance in a problem-solving environment (Greiff, Molnár, Martin, Zimmermann, & Csapó, 2018), especially if the problem is a minimal complex system (such as MicroDYN) (Fischer Greiff, & Funke, 2012). Based on the findings from Lotz, Scherer, Greiff, and Sparfeldt's (2017), the efficient use of VOTAT is associated with higher levels of intelligence, and efficient exploration behaviour can lead to a better problem-solving performance. Molnár and Csapó (2018) also conducted an empirical study to explore how students' exploration strategies influence their performance in the interactive problem-solving environment. They measured a group of (n = 4,371) 3rd to 12th-grade (aged 9-18) Hungarian students' problem-solving achievement and modeled students' exploration strategies. This result has confirmed that students' exploration strategies are influencing their problem-solving performance. For example, conscious VOTAT strategy users proved to be the best problem solvers.

2.6 Cross-cultural assessment of problem solving

2.6.1 Problem-solving research in China

Domain-general interactive problem solving has been widely assessed in Europe, but not outside of it, in the USA or in China. Much of the research in both countries has focused on domain-specific and not on domain-general problem solving. In the following we summarize the main research findings in the field of problem solving in Asian context.

There are many academic articles written in Chinese that talk about the importance of problem solving in different subjects learning, such as science (Wang,

& Zhang, 2003), medicine (Zhu, 1997) and mathematics (Cai, & Nie, 2007; Li, 2010). Two Chinese researchers, Cai and Nie (2007), stated their goal of "teaching problem solving in the classroom is to develop students' problem solving skills, help them acquire ways of thinking, form habits of persistence, and build their confidence in dealing with unfamiliar situations."(Cai, & Nie, 2007, p. 417) Besides, problem solving is often linked to the constructivism education. Liu (2001) stated that introducing problem solving into natural knowledge domain training can help student to construct their own knowledge system, develop students' cognitive skills, and provide motivation for students to do self-directed learning. Chinese researchers also stated that the usage of different kinds of problems in the subjects learning would encourage students to think and help to transfer and apply in practice what they learnt at school (Zhang, 2000; Liu, 2001; Tang, & Ren, 2006). These research projects in China also included some other orientations such as the discussion about strategies for using problem solving in the classroom and the roles of teachers and students (Cai, & Nie, 2007; Liu, 2001); the different thinking models for Chinese and foreign students in domain-specific problem solving (Cai, 2004; Cai, & Hwang, 2002); or the influence of textbook design and policy making on the application of problem solving in domain learning (Cai, 2004).

There barely exists any academic article in China, which discuss research on domain-general problem solving and its development beyond the PISA publications. China is one of the participants in the PISA assessment. This project is attracting Chinese society's and researchers' interest and attention. There are some heated discussions based on PISA and its results in both the Chinese media and academic community (Yang, 2014; Hu, 2014; Xie, & Liu, 2015). As introduced above, domain-general problem solving research is in the early, starting stage in China.

On the one hand, Chinese students belong to the top performers in the PISA problem solving assessment, but, on the other hand, the achievement proved to be significantly lower than expected based on the achievement in the three main areas (mathematics, reading and science) (see the following section 2.6.2). Thus, the research area of domain-general problem solving became an essential work for

Chinese educational researchers, it can be said that domain-general problem-solving research has a great potential in China.

2.6.2 Computer-based assessment of problem solving in China and in Hungary

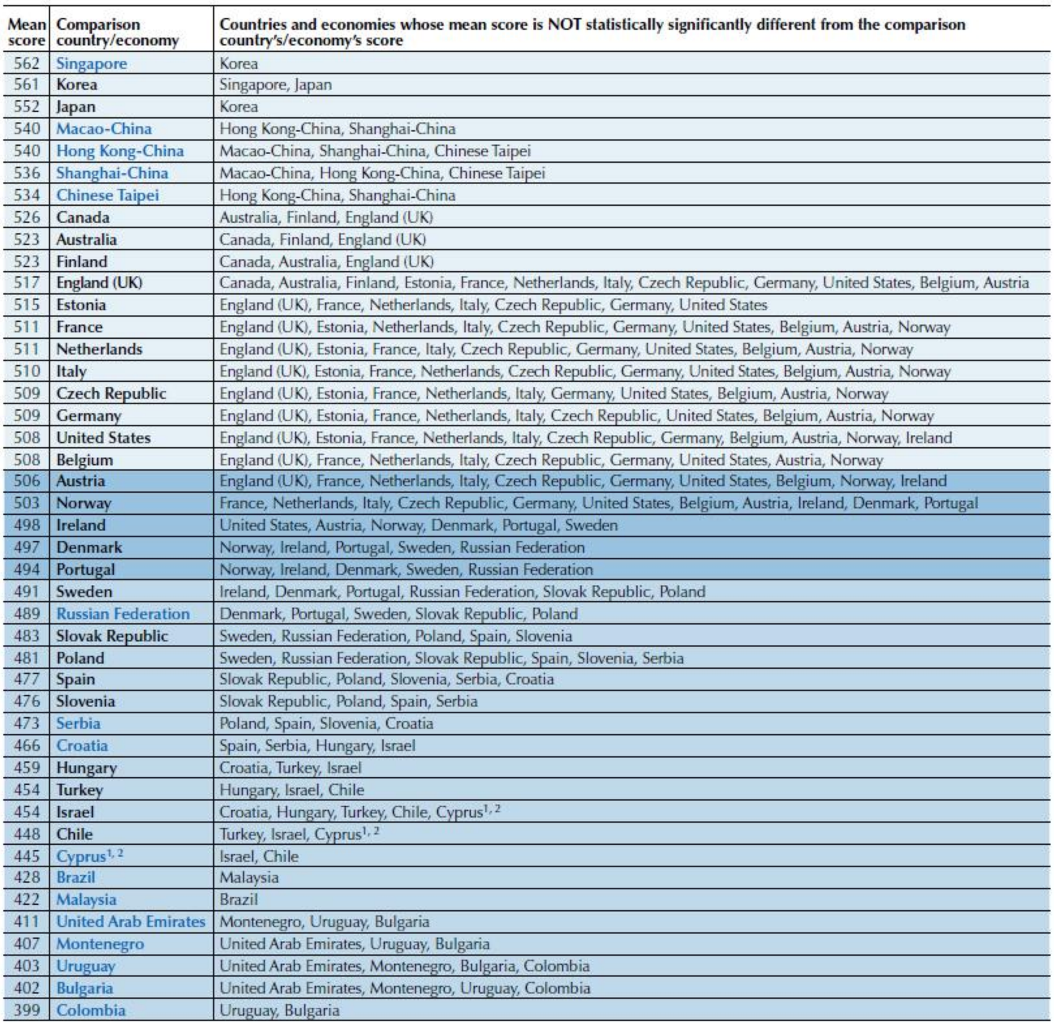
PISA measured problem solving as one of the fourth areas (mathematics, reading, science, and problem solving) in 2003, 2012 and 2015 (Greiff et al., 2013). As introduced in section 2.6.1, unlike in European countries, there have barely been any domain-general problem-solving research project conducted in China.

PISA 2012 and 2015 are the only projects that include both China and Hungary, the two countries this study focuses on, thus they can provide direct information regarding students' developmental level of problem-solving skills in these two counties. In our study we used the same construct, the same type of problem solving, in the same approach (MicroDYN) as it was basically measured in the PISA 2012 assessment (individual interactive problem solving). In the following we summarize the main PISA 2012 problem solving results regarding China and Hungary, for a top performer and for a low performer country.

China. Four cities participated in the PISA 2012 project in China: Macao, Hong Kong, Shanghai and Taipei (OECD, 2014b). Due to political and historical backgrounds, these four cities have large differences in their governmental and educational system. Therefore, in PISA, these four cities have been considered as separate entities.Among these four cities, Shanghai was representing mainland China, so the results for Shanghai students have been chosen for a detailed analysis.

Chinese students got score mark 536 in the problem-solving assessment and ranked as sixth among all the participating countries and regions. It was a very high mark and far higher than the average (OECD, 2014a). Fig. 2.12 shows the score and rank for each participating country and region. Chinese (Shanghai-China) students' mark was in front of the list (Fig. 2.15), which was better than that of the United States and European countries.

Based on students' performance, the PISA problem solving expert group distinguished six quantitatively different levels of problem solvers. Students performing at Level 5 or 6 were referred to as "top performers". The proportion of "top performers" for the Chinese students was 18.3%, which was higher than the average proportion of 11.4% (OECD, 2014a). The variation of Chinese students' problem-solving performance was not high, the standard deviation was 90, which was lower than the average (96) (OECD, 2014a). In conclusion, Chinese students' problem-solving skills had a high integral level.



Mean score	Comparison country/economy	Countries and economies whose mean score is NOT statistically significantly different from the comparison country's/economy's score
562	Singapore	Korea
561	Korea	Singapore, Japan
552	Japan	Korea
540	Macao-China	Hong Kong-China, Shanghai-China
540	Hong Kong-China	Macao-China, Shanghai-China, Chinese Taipei
536	Shanghai-China	Macao-China, Hong Kong-China, Chinese Taipei
534	Chinese Taipei	Hong Kong-China, Shanghai-China
526	Canada	Australia, Finland, England (UK)
523	Australia	Canada, Finland, England (UK)
523	Finland	Canada, Australia, England (UK)
517	England (UK)	Canada, Australia, Finland, Estonia, France, Netherlands, Italy, Czech Republic, Germany, United States, Belgium, Austria
515	Estonia	England (UK), France, Netherlands, Italy, Czech Republic, Germany, United States
511	France	England (UK), Estonia, Netherlands, Italy, Czech Republic, Germany, United States, Belgium, Austria, Norway
511	Netherlands	England (UK), Estonia, France, Italy, Czech Republic, Germany, United States, Belgium, Austria, Norway
510	Italy	England (UK), Estonia, France, Netherlands, Czech Republic, Germany, United States, Belgium, Austria, Norway
509	Czech Republic	England (UK), Estonia, France, Netherlands, Italy, Germany, United States, Belgium, Austria, Norway
509	Germany	England (UK), Estonia, France, Netherlands, Italy, Czech Republic, United States, Belgium, Austria, Norway
508	United States	England (UK), Estonia, France, Netherlands, Italy, Czech Republic, Germany, Belgium, Austria, Norway, Ireland
508	Belgium	England (UK), France, Netherlands, Italy, Czech Republic, Germany, United States, Austria, Norway
506	Austria	England (UK), France, Netherlands, Italy, Czech Republic, Germany, United States, Belgium, Norway, Ireland
503	Norway	France, Netherlands, Italy, Czech Republic, Germany, United States, Belgium, Austria, Ireland, Denmark, Portugal
498	Ireland	United States, Austria, Norway, Denmark, Portugal, Sweden
497	Denmark	Norway, Ireland, Portugal, Sweden, Russian Federation
494	Portugal	Norway, Ireland, Denmark, Sweden, Russian Federation
491	Sweden	Ireland, Denmark, Portugal, Russian Federation, Slovak Republic, Poland
489	Russian Federation	Denmark, Portugal, Sweden, Slovak Republic, Poland
483	Slovak Republic	Sweden, Russian Federation, Poland, Spain, Slovenia
481	Poland	Sweden, Russian Federation, Slovak Republic, Spain, Slovenia, Serbia
477	Spain	Slovak Republic, Poland, Slovenia, Serbia, Croatia
476	Slovenia	Slovak Republic, Poland, Spain, Serbia
473	Serbia	Poland, Spain, Slovenia, Croatia
466	Croatia	Spain, Serbia, Hungary, Israel
459	Hungary	Croatia, Turkey, Israel
454	Turkey	Hungary, Israel, Chile
454	Israel	Croatia, Hungary, Turkey, Chile, Cyprus[1, 2]
448	Chile	Turkey, Israel, Cyprus[1, 2]
445	Cyprus[1, 2]	Israel, Chile
428	Brazil	Malaysia
422	Malaysia	Brazil
411	United Arab Emirates	Montenegro, Uruguay, Bulgaria
407	Montenegro	United Arab Emirates, Uruguay, Bulgaria
403	Uruguay	United Arab Emirates, Montenegro, Bulgaria, Colombia
402	Bulgaria	United Arab Emirates, Montenegro, Uruguay, Colombia
399	Colombia	Uruguay, Bulgaria

Fig. 2.15 Students' performance in the PISA 2012 problem solving assessment (OECD, 2014a, p. 52)

Despite of the very high achievement, one important issue has been identified regarding the Chinese performance. Mathematics, reading and science are measured as fix, so called main domains in every PISA cycle. It was the case in 2012, too. In the international ranking, China was the top performer, having 1st place on each of the lists. However, this was not the case in the problem solving assessment module, where the same students were ranked for the 6th place, which was far lower than excepted based on their achievement level on the three main domains. Moreover, with this difference they had the almost largest difference between the real and the expected performance level of problem solving (Fig. 2.16).

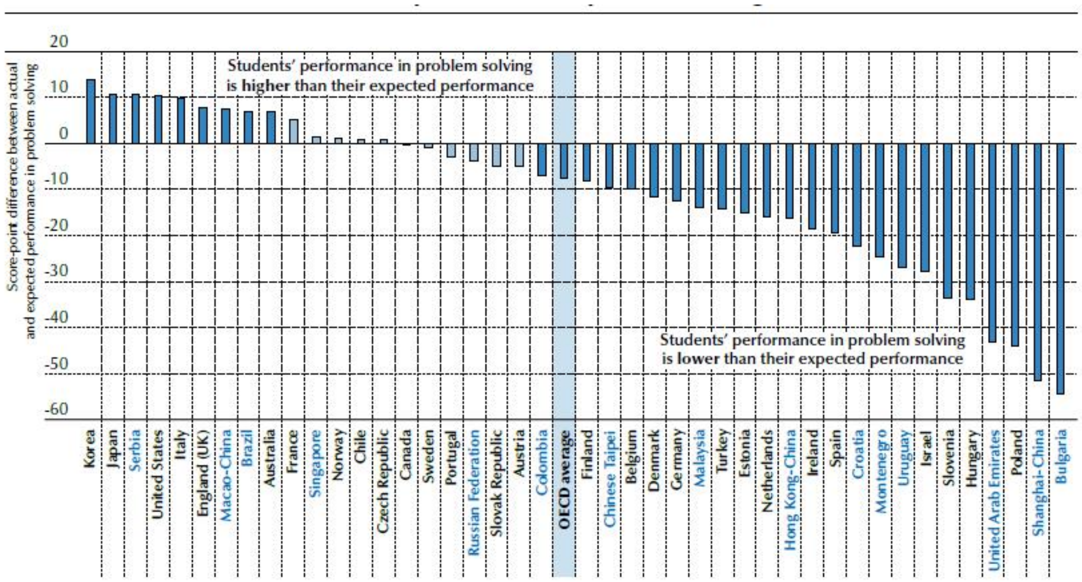


Fig. 2.16 Relative performance in problem solving (OECD, 2014a, p. 69)

Hungary. Hungarian students' problems solving performance was significantly lower than the average. It was ranked as the 33th place out of 44 (see Fig. 2.15). The percentage of the top Hungarian 15-year-old performers was 4.1%, much lower than the average. Besides, the variation in students' problem-solving performance was big, the standard deviation 104, higher than the average (OECD, 2014a). 34% of the students achieved significantly lower than expected based on their achievement on the three main domains, which resulted in significantly lower achievement on the problem solving test than expected based on the performance on the other tests (which was also significantly lower than the average OECD achievement). The difference between Hungarian male and female students' performance was lower than the average and much lower than that of the Chinese students.

To sum up, the results of PISA 2012 problem-solving assessment indicated the importance of conducting a problem solving research in both China and Hungary to detect possible reasons for lagging behind, the huge differences and to establish the scientific foundations for further training programs. The problem solving skills of the Chinese students proved to be very high; however, it was much lower than expected. The Hungarian students showed generally unsatisfactory performance in the problem solving assessment. Thus, both Chinese and Hungarian students had the necessity and possibility to improve their problem solving skills. In order to reach this goal, we need to measure and analyze the developmental level of Chinese and Hungarian students' problem-solving skills, to detect cognitive and non-cognitive factors influencing their achievement, and to map and compare the component skills of problem solving in China and in Hungary. It is the most important and foundational work which builds the bases of the present dissertation.

3. Research questions and hypotheses

The study aims to contribute to our further understanding of the cognitive construction, of the component skills of interactive problem solving (IPS), and highlight the similarities and differences that may exist between students from Hungarian and Chinese cultures.

Research on both domain-general problem-solving and computer-based assessment in educational context is still at the starting stage in China. Our study merges the two research directions and after detecting the reliability of computer-based assessment in school context, focuses on the assessment of 12-year-old students' problem solving skills in domain-general and interactive context, using third generation computer-based tests. The result of the study shows similarities and differences in the influential and predicting factors of problem solving skills in international contexts.

Particularly, at the first stage, the study aims to testify the feasibility and reliability of computer-based assessment in Chinese educational context. Due to the fact that computer-based assessment of thinking skills (included but not limited to the thinking skills involved in this study) are frequently used and have been already implemented in Hungary (e.g. Csapó, & Molnár, 2017; Csapó, Molnár, & Nagy, 2014; Molnár, Greiff, & Csapó, 2013; Molnár et al., 2017), this does not belong to the aims of the present study.

Beyond the usage of computer-based assessment in educational context in China, the study aims to detect the nature of problem solving, especially interactive problem solving in the MicroDYN approach, by analyzing the dimensionality of IPS, testifying the component skills of IPS, and exploring further cognitive and affective factors, which influence students' achievement in IPS environment. Moreover, a comparison study between Chinese and Hungarian students have been conducted to explore the similarities and differences in the cognitive construction of IPS in European and Asian educational contexts.

The study thus intends to answer eight research questions:

RQ1. *Is computer-based assessment feasible in educational context in China? Particularly, how reliable can the third-generation online test of IPS, the second generation online tests of inductive reasoning, combinatorial reasoning, and creativity be applied in the Chinese cultural and school network environment?*

Each test was translated from Hungarian and English into simplified Chinese. The IPS tests have worked well in European countries, e.g. in Germany and Hungary, showing high reliability indices (see e.g., Greiff, & Wüstenberg, 2014; Molnár, Greiff, & Csapó, 2013) and proving that the application of the tests (IPS, inductive reasoning, combinatorial reasoning, creativity) are reliable, at least in European contexts.

All of the tests were designed so that it was possible to keep the influence of the cultural differences into the minimum level (no factual knowledge with the minimum level of reading, mostly using figural items). By the test adaptation process each of the used terms were carefully selected based on the Chinese students' language use. Thus, we expect that all of the tests can be successfully implemented in China even in a computer-based format.

H1: *Computer-based assessment, as well as the tests in this study can be reliably applied in the Chinese cultural and school network environment.*

RQ2. *Is the behaviour of the tasks measuring knowledge acquisition and knowledge application as two dimensions of IPS in the MicroDYN approach independent of cultural context?*

Problems in the MicroDYN approach have two phases: knowledge acquisition and knowledge application (see section 2.3.3). These two phases have been proved as two dimensions of IPS in Hungary (Greiff et al., 2013). Based on the theories behind the MicroDYN approach, we assume that the IPS problems work similarly in China and in Hungary, independently of the cultural context, and the model of problem representation and problem solution are composed of the two dimensions of knowledge acquisition and knowledge application.

H2: *IPS is composed of two different processes, knowledge application and knowledge acquisition, independently of the cultural context. That is, a two-dimensional model was expected to show significantly better fit than a one-dimensional model with the two processes combined as one factor.*

RQ3. *Can IPS be measurement invariant across gender and nationality both in the contexts of Europe and Asia?*

Measurement invariance is inevitable in cross-national studies if we want to compare the ability level of students. If measurement invariance is not holding, it is possible that the differences between students' ability levels are influenced by group features or measurement occasion (i.e., time), that is – in the present case – they are varied due to their different cultural backgrounds and gender and not by the differences in their ability levels. During the test adaptation and translation process it can lead to measurement non-invariance if the meaning of the text changes, and this results in different reactions of the students with different backgrounds. The test adaptation and translation process in the present study have been double-checked by language experts, which should have minimized translation errors. Therefore, at this stage, we assume measurement invariance holds across nationality and gender,because the MicroDYN based IPS assessment has been proved to be measurement invariant across nationality and gender in the Hungarian and German cross-national research (Wüstenberg et al., 2014), and the PISA 2012 problem-solving assessment (part of the items were based on MicroDYN approach) worked well across countries (OECD, 2014a).

H3: *IPS can be measurement invariant between male and female, and Hungary and*

China.

RQ4. *Can developmental differences in IPS be detected between Hungarian and Chinese 12-year-old students, 3 years before the PISA age?*

According to Molnár, Greiff, & Csapó (2013), students' reasoning skills start to grown rapidly at around the age of 12 and tend to be stable at around age 15. Considering the big gap between 15-year-old Chinese and Hungarian students regarding problem solving skills identified by PISA 2012 (see section 2.6.2), we expect that this gap exists even a few years earlier, at the earlier stage of the developmental process of the reasoning skills. However, we expect a smaller gap between the Hungarian and the Chinese 12-year-old students than that is present at the PISA age.

H4: *Developmental differences in IPS can be detected between Hungarian and Chinese 12-year-old students.*

RQ5. *What kind of developmental differences can be identified across gender and nationality, between Chinese males and females, and between Hungarian males and females?*

There has been almost no IPS assessment project conducted in China beyond the PISA assessment. Thus, we can form our hypotheses only based on the PISA 2012 results. In the problem solving module of the PISA 2012 assessment, Chinese students significantly over-performed Hungarian students. Even though PISA focuses on 15-year-olds and not 12-year-old students, considering the huge gap between these two groups of students, we assume that Chinese students have a significantly higher performance than the Hungarian 12-year-old students. In PISA 2012, both Hungarian and Chinese boys performed significantly better than girls. Thus, we expected to get similar results from the comparison study. But considering that the sample size in

PISA is much larger than it is in this study, there is also certain possibility that no significant gender differences can be detected in this study.

H5: *Chinese 12-year-old students have significantly better performance than Hungarian students in the same age group; Chinese and Hungarian boys' performance is statistically higher or equal to that of the girls.*

RQ6. *Are inductive reasoning, combinatorial reasoning, and creativity component skills of problem solving? How strong is their predictive power on the problem solving achievement of Chinese and Hungarian 12-year-old students?*

Thinking skills such as inductive reasoning, combinatorial reasoning, and creativity relate to problem solving according to several research results (Bisanz, Bisanz, & Korpan, 1994; Hamers, De Koning, & Sijtsma, 2000; Herrmann, 1995; Molnár et al., 2017) (see section 2.4.2). The present study claims that these thinking skills are component skills of problem solving, and have significant predictive power on both Chinese and Hungarian students' problem solving achievement.

H6: *Inductive reasoning, combinatorial reasoning, creativity are component skills of problem solving, which have significant predictive power on the problem solving achievement by Chinese and Hungarian 12-year-old students.*

RQ7. *Which factors influence Chinese and Hungarian students' IPS achievement beyond thinking skills? How strong is the influential effect?*

Previous studies pointed out that expect for the thinking skills, some factors (e.g. working memory, learning strategies, social-economic background factors) influence students' problem solving achievements (see section 2.4.1 and 2.4.3). Based on previous studies, this study selected several important factors including working memory, gender, parents' education, learning strategies, and test-taking motivation, and expect that they will have a significant influential effect on students' problem solving achievement.

H7: *The selected non-thinking skill factors (working memory, parents' education, learning strategies, and test-taking motivation) have a significant influential effect on Chinese and Hungarian students' IPS achievement.*

RQ8. *Do Chinese and Hungarian students employ different exploration strategies during the problem-solving process? How do their exploration strategies influence their problem-solving performance?*

Molnár and Csapó's (2018) study has proved that Hungarian students' exploration strategies are influencing students' problem-solving performance. Therefore, the present study expects that the quality of students' exploration strategies explains their performance in the IPS assessment at a certain level. Very few studies have been published on students' exploration strategies in problem-solving environment in China. But according to Nisbett and Miyamoto's (2005) finding, students' cognitive styles, especially in the problem-solving environment, can be impacted by their cultural background. Considering that there is a significant cultural difference between China and Hungary, we hypothesize that Chinese and Hungarian students will employ different exploration strategies during the problem-solving process which fit better to their different cognitive styles.

H8: *Chinese and Hungarian students employ different exploration strategies during the problem-solving process. The adoption of different exploration strategy can lead to different problem solving performance.*

4. Pilot study: The feasibility of computer-based assessment of reasoning skills in China

4.1 Aims, research questions and hypotheses for the pilot study

The usage of computer-based testing in educational context is not very common in China, therefore at the first stage of the study it was necessary to detect the feasibility and applicability of computer-based assessment in China, and to make sure that the online assessment platform, eDia can be run in the Chinese network environment. Beyond testing the feasibility of computer-based assessment in China, in the pilot study we aimed to test the behaviour and psychometric indices of the newly adapted tests in the Chinese culture involved in the study. Thus, measuring the reliability of the tests was an essential work at this stage. The pilot study also provided the opportunity to run preliminary analyses and to achieve preliminary answers – at least on a methodology level – to the research questions referring to the component skills of problem solving. The databases – based on the pilot study – did not have the power to accurately answer the research questions regarding the component skills or problem solving, but it was able to confirm the feasibility of the data analysis, the feasibility of the planed research methodology and provided guidance for further study.

The pilot study aimed to answer research question RQ1 and verify hypothesis H1.
RQ1. Is computer-based assessment feasible in educational context in China? Particularly, how reliable can the third-generation online test of IPS, the second generation online tests of inductive reasoning, combinatorial reasoning, and creativity be applied in the Chinese cultural and school network environment?

H1: Computer-based assessment, as well as the tests in this study can be reliably applied in the Chinese cultural and school network environment.

4.2 Methods

4.2.1 The sample of the pilot study

As the main study aims was to explore students' developmental level of problem solving at the beginning of the developmental phase (around 12 years of age, see Molnár, Greiff, & Csapó, 2013), the participants for the pilot study were drawn from a group of 12-year-old students. 50 Chinese students (27 boys; 23 girls) attended the pilot study. All of the participants were six graders (age mean=12.28, standard deviation=.50). According to the Chinese educational system, the participants were in the last year of the primary school.

4.2.2 Instruments

4.2.2.1 Computer-based, third-generation test of IPS

The computer-based test of IPS was developed in the MicroDYN approach (see section 2.3.3). The test contained 20 items with varying difficulty levels. Complexity was defined by the number of input and output variables and the number of relations based on the Cognitive Load Theory (Sweller, 1994). "Findings show that increases in the number of relations that must be processed in parallel in reasoning tasks consistently lead to increases in task difficulty" (Beckmann, & Goode, 2017, p. 5).

The items were originally developed in Germany (Greiff et al., 2012; Wüstenberg et al., 2012). Later on, they were adapted to many languages and applied in many countries, including Hungary (see Csapó, & Molnár, 2017; Greiff, Krkovic, & Hautamäki, 2016). In 2011 Molnár (2013) adapted the items developed in the MicroDYN approach to Hungarian, later she transferred all the items (Molnár, 2014) into the eDia platform (Molnár, & Csapó, 2013).

The IPS assessment instrument used in this study is based on this adapted version. The language was translated into simplified Chinese with language experts' double checks (see Appendix A for the Hungarian-Chinese comparison of the Cat problem).

At the beginning of the test, participants were provided with video-based instructions about the usage of the user interface, including two warm-up tasks. Subsequently, students had to explore, describe and operate unfamiliar systems, they had to find out how input and output variables were interconnected, represent their knowledge in a situational model below the problem scenario (see Fig. 4.1; knowledge acquisition; Funke, 2001). In addition, in the second phase of the problem, they had to control the system by reaching given target values (knowledge application; see Greiff, Wüstenberg, & Funke, 2012; see Fig. 4.2). Students had three minutes to provide answers in each phase of the problem solving process. Fig. 4.1 and 4.2 provides the screenshot for the Cat problem.

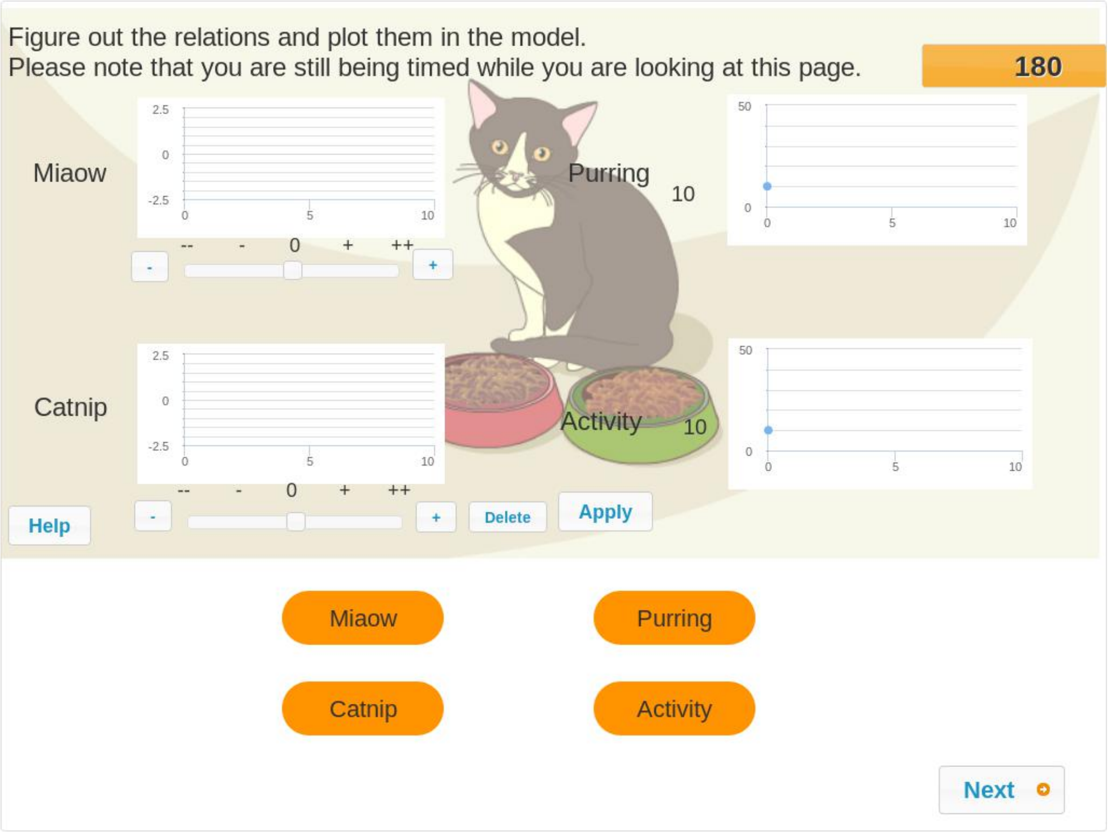


Figure out the relations and plot them in the model.
Please note that you are still being timed while you are looking at this page.

Fig. 4.1 Screenshot of the MicroDYN task Cat – First phase. [The original items were in Hungarian. The controllers of the input variables range from “- -” (value=-2) to “++“ (value=+2). The current values of the output variables are displayed numerically (e.g., current value for Purring: 10) and graphically (current value: dots) (see Greiff, Wüstenberg, Molnár, Fischer, Funke, & Csapó, 2013).]

Each set of tasks (containing one knowledge acquisition item and one knowledge application item) had one fictitious cover story, thus the contents of the problems were not based on real knowledge, and the relationships between the input and output variables were artificial. Students' content knowledge was therefore of no help to fill the gap between the insufficient information acquired from interaction and the successful solution to the problem. The items contained up to three input variables and up to three output variables.

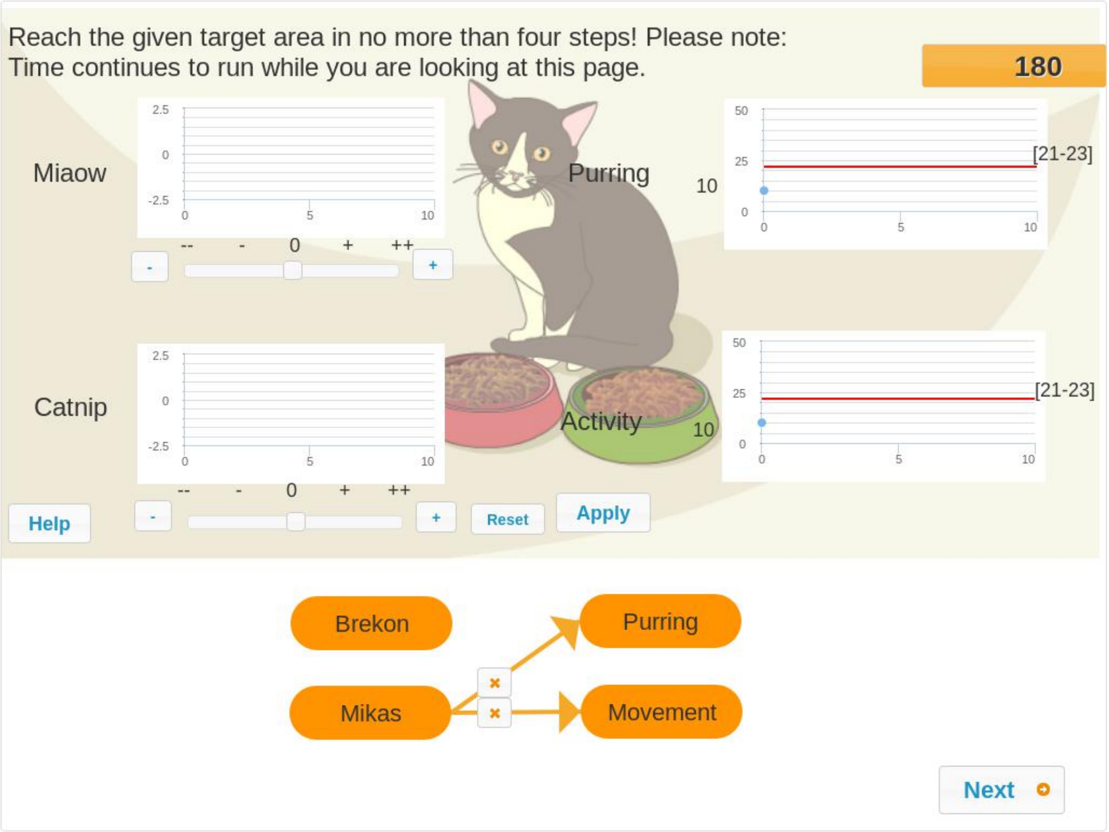

Reach the given target area in no more than four steps! Please note: Time continues to run while you are looking at this page.

Fig. 4.2 Screenshot of the MicroDYN task Cat – Second phase. [The original items were in Hungarian. The controllers of the input variables range from "- -" (value=-2) to "++" (value=+2). The current values and the target values of the output variables are displayed numerically (e.g., current value for Activity: 10, target value: 21-23) and graphically (current value: dots; target value: red line). The correct model is shown at the bottom of the figure (see Greiff, Wüstenberg, Molnár, Fischer, Funke, & Csapó, 2013).]

Each of the input variables had 5 stages: +2 (++), +1 (+), 0, -1 (-), -2 (--). To operate the system, students must change the value of the input variables by clicking on the button with + or - sign or by using the slider connected to the respective variable. The history of the values of the input variables within the same problem and within the same scenario was presented on a graph connected to each input variable. The right part of the interface contained graphs indicating the values of each of the output variables.

Beyond the input and output variables, each scenario contained a Help, Reset, Application and Next button. The Reset button set the system back to their original status, the values of each of the input variables to zero and the values of each of the output variables to their original value. The Application button was designed for testing the effect of the currently set values of the input variables on the output variables. After clicking the Application button, the output variable(s) will have corresponding change in both numerical and figural formats. Within the same phase of each of the problem scenarios, the input values remained at the level for the previous input setting. The problem solver can press Reset to set it back to zero or change them manually to any value in the available scope. The test-taker also can click the Next button move to next scenario or phase. (Molnár, & Csapó, 2018).

In the first phase, in the knowledge acquisition phase, students were asked to explore the system and find the relationships between the input and the output variables. They were able to freely operate the system by changing the values of the input variables and applying as many trials for each MicroDYN scenario as they liked within 3 minutes. During these 3 minutes, they had to draw the concept map, that is, they had to draw the arrows between the input and output variables to present the relationships they found during the exploration. In the second phase, in the knowledge application phase of the problem-solving process, participants had to check their respective system using the right concept map presented on screen. Students had to let the output variables to reach the given target values by changing the value of input variables within a given time frame. In this phase, students were only allowed to have four trials, that is, they could click on the Application button four or less than four times.

We also included IPS problems with an external manipulation-independent, internal dynamic effect, but excluded problems with multiple dependence effect in the study; also, there were no delayed or accumulating effects used in the problem environments executed. In the problems with eigendynamic (Greiff et al., 2013), output variables can influence themselves (introduced in section 2.3.3). Problem solvers cannot influence this effect directly, but they are detectable by adequate use of strategy.

The IPS tasks can be automatically scored by the eDia platform. If the test-taker successfully solves a problem within the time limitation, the system will record it as 1 score; otherwise it will be recorded as 0 score.

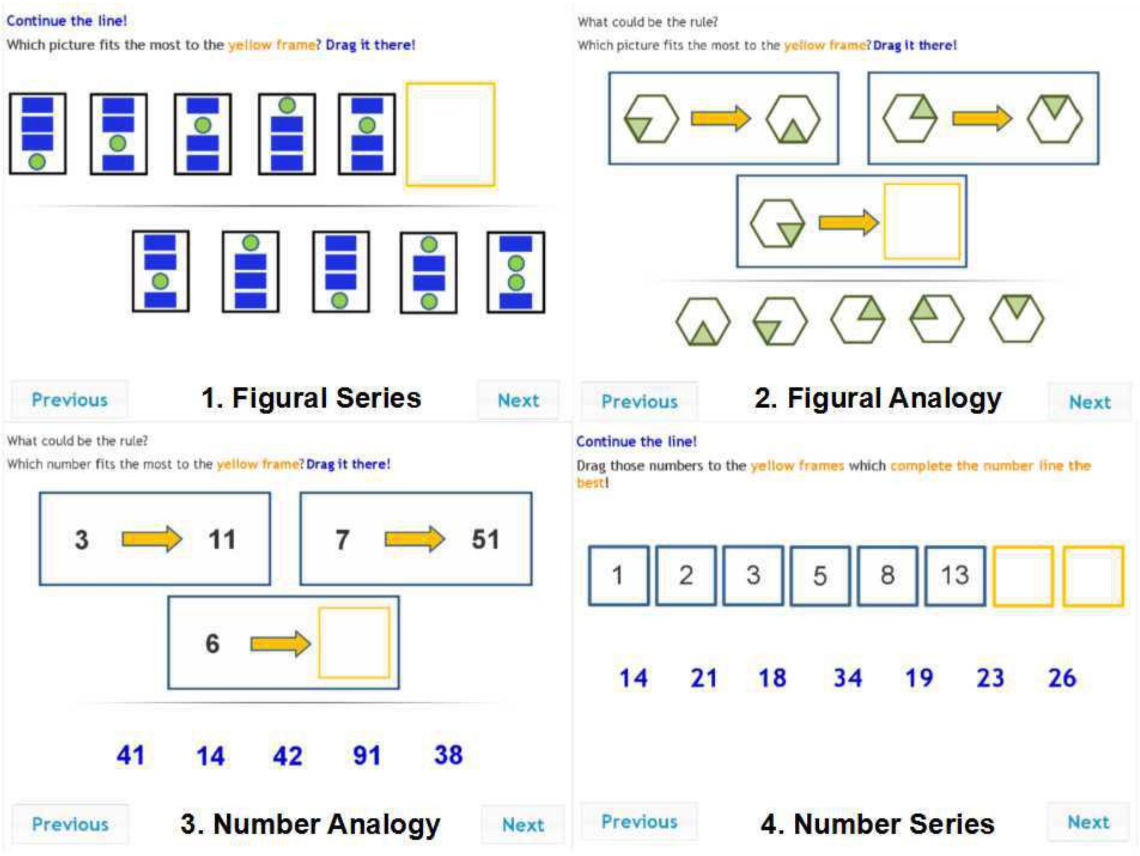
4.2.2.2 Computer-based, second-generation test of inductive reasoning

The inductive reasoning test contained 50 multiple-choice items, which consisted of four sub-constructs: figural series, figural analogy, number analogy and number series (Fig. 4.3). The development of this test has a long history.

Csapó (1997) developed a paper-and-pencil test of inductive reasoning consisting of number analogies, number series, verbal analogies and verbal series for higher graders, while Molnár (2011) developed a paper-and-pencil test of inductive reasoning for low graders including only figural items. Later on, Csapó, Molnár, & Tóth (2009) transferred these items from paper-and-pencil format to a computer-based format and scaled the items together, constructing an inductive reasoning competency scale (Molnár, & Csapó, 2011). The first computer-based data collection of these tests were carried out by the TAO platform (Testing Assisté par Ordinateur; Csapó, Molnár, & Tóth, 2009), while the refreshment of the test and test items, its colouring and the additional item development (Pásztor, Molnár, Korom, Németh, & Csapó, 2017) were connected to the eDia platform. The inductive reasoning test used in this study was developed by Pásztor, Molnár and Csapó (Pásztor, Hotulainen, Kupiainen, Molnár, & Csapó, 2018; Pásztor et al., 2017) and adapted from the eDia test version. The assessment language was translated from Hungarian to English, then from English to simplified Chinese (see Appendix B), each translation stage was double checked by language experts.

Fig. 4.3 contains 4 items, which belong to figural series, figural analogy, number analogy and number series, respectively. Students had to discover the rules from the given pictures or numbers, and use drag and drop to put their answer into the yellow square.

Students' score for the inductive reasoning test is automatically calculated by the eDia platform. The score is recorded as dichotomous variable (1: provided correct answer; 0: did not provide correct answer).



Continue the line!
Which picture fits the most to the yellow frame? Drag it there!

1. Figural Series

What could be the rule?
Which picture fits the most to the yellow frame? Drag it there!

2. Figural Analogy

What could be the rule?
Which number fits the most to the yellow frame? Drag it there!

3. Number Analogy

Continue the line!
Drag those numbers to the yellow frames which complete the number line the best!

4. Number Series

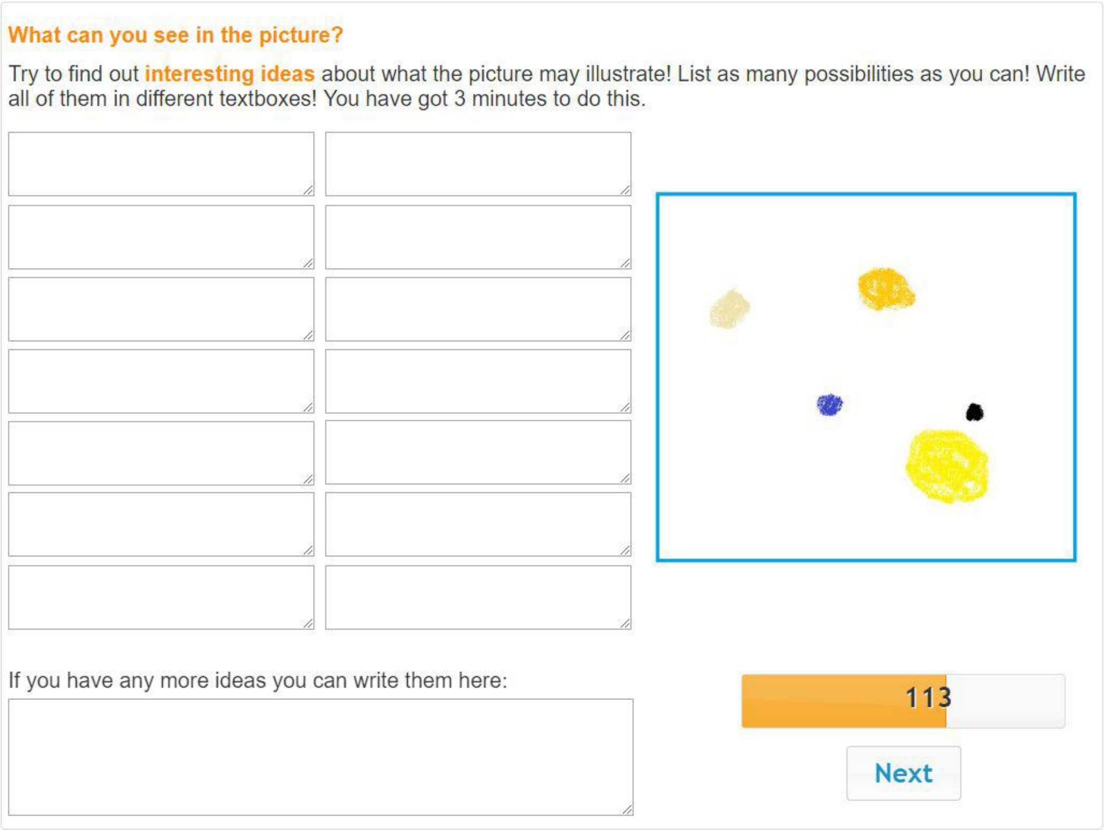
Fig. 4.3 Sample items for the inductive reasoning test

4.2.2.3 Computer-based, second-generation test of creativity

The creativity test contained 6 items in total. The creativity instrument was adapted from Pásztor, Molnár and Csapó's (2015) work. The items were open-ended type items. The test modeled students' creativity by three dimensions, which were fluency (the ability to produce numerous ideas for a given problem), flexibility (the ability to consider and see a problem from different approaches), and originality (the ability to create unique, unusual, and original ideas) (Pásztor, Molnár, & Csapó, 2015). For the verbal-verbal creativity items (Fig. 4.4), students needed to provide every single possible usage they could image for three daily necessities (match, cup, and toothbrush). There were also three picture meaning tasks (figural-verbal creativity) included in the test, in which students needed to write every meaning they could image from the different pictures (Fig. 4.5) into the textboxes. Students had three minutes to reach their answers for each item. The language was translated into simplified Chinese (see Appendix E).

The creativity test was scored based on the fluency, flexibility, and originality of students' answers (see Pásztor, Molnár, & Csapó, 2015). The scores were shown as real number (equal or above 0) with two decimals.

The original version of the test worked well in the Hungarian cultural context (see Pásztor, Molnár, & Csapó, 2015). However, there was a concern whether the test could work well in the Chinese culture since creative ideas are connected to people's daily experience at a certain level. Therefore, the validity of this assessment in Chinese cultural context had to be confirmed. The items were open-ended type items. The scoring required a lot of human work since automatic scoring was not available. The pilot test aimed to estimate the time and resources required in the scoring process, thus, to make a decision whether to include the creativity test into the comparison study.



What can you see in the picture?

Try to find out **interesting ideas** about what the picture may illustrate! List as many possibilities as you can! Write all of them in different textboxes! You have got 3 minutes to do this.

If you have any more ideas you can write them here:

Fig. 4.4 Sample item for the creativity test

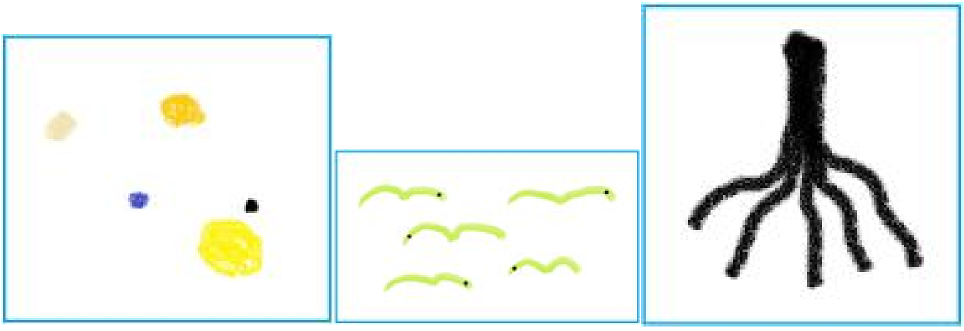


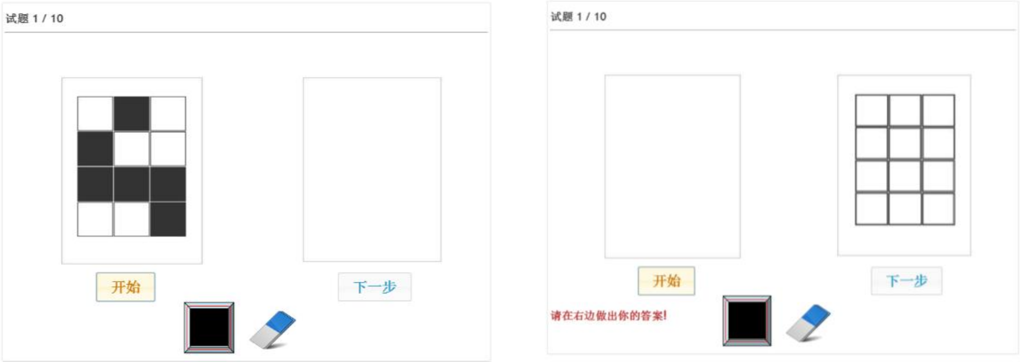
Fig. 4.5 Stimuli for the picture meaning tasks

4.2.2.4 Computer-based test of working memory

The working memory test contained 11 items with varying difficulty levels. The original idea of this working memory capacity test was created and designed in Finland (Kyttälä, Aunio, Lepola, & Hautamäki, 2014); later on it has been adapted and developed further in the eDia system in Hungary. Because the information in the

MicroDYN approach based problem-solving assessment is mainly delivered in the visual format, the working memory test mainly designed to assess visuospatial sketch pad and central executive these two working memory components (see section 2.4.1), but not the phonological loop. That is, the working memory instrument focuses on students' working memory when they are receiving visual information.

The working memory instrument has been widely used in Hungary. The language was translated from Hungarian to simplified Chinese. The test contained 2 warm-up tasks and 11 additionally tasks. Each task had two parts. In the first part, the system showed some patterns constituted by black and white squares (see Fig. 4.6). Each picture lasted for three seconds on the screen. In the second part of the task, students needed to remember and draw the pattern on the screen in the answer zone by coloring the squares black (see Appendix D). Students' score for the working memory test is automatically calculated by the eDia platform. The score for one item is identified as 1 (provided a correct answer) or 0 (did not provide a correct answer).



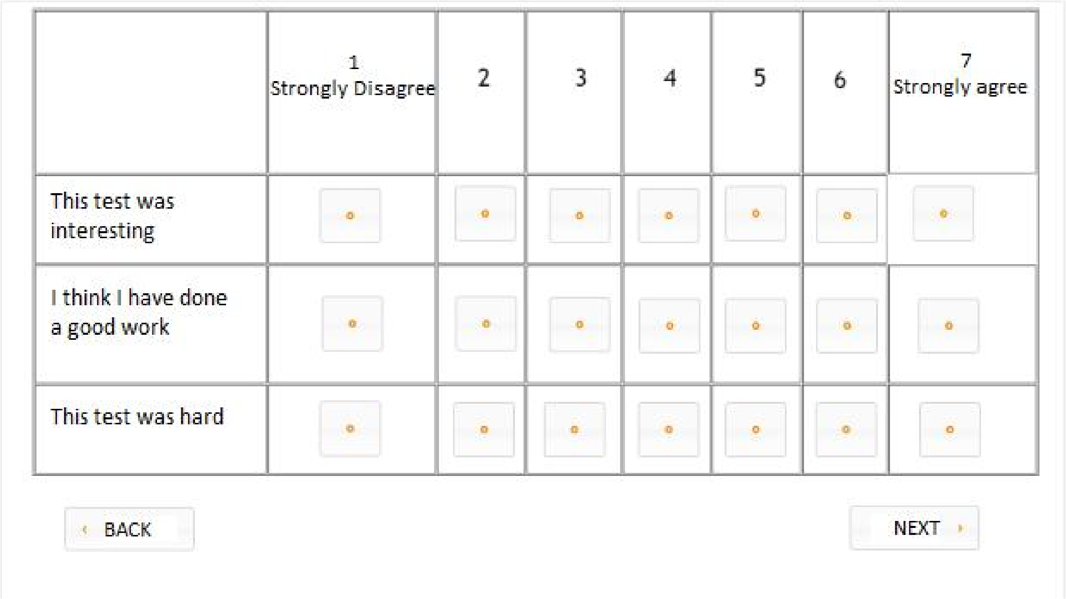
请在右边做出你的答案!

Fig. 4.6 Sample item for the working memory test

4.2.2.6 Computer-based questionnaire of test-taking motivation

Some test-taking motivation questions have been inserted into the problem-solving assessment process in order to monitor students' motivation during the whole progress. The motivation questions were first designed in Finnish then adapted into Hungarian. This study was translated from Hungarian to English, then from English to simplified

Chinese. The questions asked about students' attitudes regarding this assessment. The questions have 7 scales from 1- strongly disagree to 7- strongly agree.



	1 Strongly Disagree	2	3	4	5	6	7 Strongly agree
This test was interesting	○	○	○	○	○	○	○
I think I have done a good work	○	○	○	○	○	○	○
This test was hard	○	○	○	○	○	○	○

Fig. 4.7 Sample item for the motivation test

In addition, the motivation questions have been divided into three parts (Table 4.1). Test-taking motivation items 1-3 have been inserted between the introduction section and the first problem-solving item. They have been used to monitor students' motivation before the assessment. Items 4-6 have been inserted into the middle of the assessment, which can be used to monitor students' motivation during the assessment. Items 7-15 have been set as the last part of the assessment, which can be used to see students' motivation for the whole assessment.

Most of the questions were describing positive attitudes, while a few questions were describing negative attitudes. Table 4.1 shows the English translation for each question.

	Question	Position	Attitude
1	This test looks interesting	Before the test	Positive
2	I think I’m going to do a good work	Before the test	Positive
3	This test looks hard	Before the test	Negative
4	This test is interesting	During the test	Positive
5	I think currently I’m doing a good job	During the test	Positive
6	This test is hard	During the test	Negative
7	This test was interesting	After the test	Positive
8	I think I have done a good job	After the test	Positive
9	This test was hard	After the test	Negative
10	This test attracted my interest	After the test	Positive
11	This test attracted my attention	After the test	Positive
12	This test was entertaining	After the test	Positive
13	Participating in this test was an interesting task to do	After the test	Positive
14	I would like to do more tests like this	After the test	Positive
15	I could easily concentrate in this test	After the test	Positive

Table 4.1 Test-taking motivation questionnaire questions

4.2.2.6 Computer-based questionnaire of learning strategies

A learning strategies questionnaire has been included to invest students’ learning strategies. Educational psychologists have created several effective questionnaires to measure students’ learning strategies, such as the TLSQ (The Teaching and Learning Strategies Questionnaire) designed by the Centre for the Study of Learning & Performance at Concordia University (Abrami, Wade, Pillay, Aslan, Bures, & Bentley, 2008), and Motivated Strategies for Learning Questionnaire (MSLQ) designed by researchers from the National Center for Research to Improve Postsecondary Teaching and Learning (NCRIFFAL) and the School of Education at the University of Michigan (Pintrich, Smith, Garcia, & McKeachie, 1991). The PISA 2003 learning strategies questionnaire (Artelt et al., 2003) was selected in this study due to the following reasons: (1) it was applied on international level previously and received good and reliable results (Artelt et al., 2003), and (2) it has the official Chinese version (it was used in Taiwan), thus reduced the workload and avoided the risk of inaccurate translation.

The questionnaire lists 13 statements about different learning habits (Table 4.2) which can be grouped into three strategies: (1) elaboration strategies (strategies to link their learning content with their previous knowledge or real-life), (2) memorization strategies (strategies to remember everything without thinking), and (3) control strategies (strategies to do self-control or time-control). Students need to choose the frequency (5-scale Likert-scale: Never, Rarely, Sometimes, Often, Always) of using the listed learning strategies in their daily study.

Elaboration strategies
When I study, I try to relate new material to things I have learned in other subjects
When I study, I figure out how the information might be useful in the real world
When I study, I try to understand the material better by relating it to things I already know
When I study, I figure out how the material fits in with what I have learned
Memorization strategies
When I study, I try to memorize everything that might be covered
When I study, I memorize as much as possible
When I study, I memorize all new material so that I can recite it
When I study, I practice by saying the material to myself over and over
Control strategies
When I study, I start by figuring out what exactly I need to learn
When I study, I force myself to check to see if I remember what I have learned
When I study, I try to figure out, as I read, which concepts I still haven't really understood
When I study, I make sure that I remember the most important things
When I study, and I don't understand something, I look for additional information to clarify the point

Table 4.2 Learning strategies questionnaire questions

4.2.3 Procedures

The tests were carried out using the eDia (Electronic Diagnostic Assessment) platform in the school's ICT room in June, 2016. All the participating students sat together in one room for the assessment. Test completion was divided into three sessions, each lasting approximately 45 minutes; between each session there were 10 minute break. In session 1, students worked on the inductive reasoning test. In session 2, students had to complete the IPS test, while in session 3, they worked on the working memory test, the creativity test, and the questionnaire.

As the theoretical part introduced, combinatorial reasoning is also very much possible to be one of the component skills of problem solving (see section 2.4.2). However, the combinatorial reasoning assessment instrument was not fully prepared before the pilot study implementation. Thus, it was not included in the pilot test.

4.2.4 Data analysis

Since the sample size for the pilot study was not high, the data had no power to answer the research questions except for RQ1. However, the statistical analyses based on the pilot data still has been run, the results could be the evidence for the feasibility of the whole study.

Cronbach's alpha for each test was calculated by SPSS version 22 to indicate reliability of our computer-based assessment instruments. Structural equation modeling (SEM) was used to analyze the dimensionality of IPS and the relationships between each factor. All the models were computed by Mplus version 5 (Muthén, & Muthén, 2010). CFI (Comparative Fit Index), TLI (Tucker-Lewis Index), SRMR (Standardized Root Mean Square Residual) and RMSEA (Root Mean Square Error of Approximation) were calculated by Mplus to indicate the model fit.

In addition, correlation analysis was used to analyze the relationships between thinking skills, while independent t-test was used to analyze the relationships between thinking skills and non-thinking skill factors. The correlation analysis and t-test analyses were carried out using SPSS version 22.

Item-response theory (Rasch model) was used to analyze the difficulty the level of the items and test-taking motivation. The Rasch model was run by ConQuest (Wu, Raymond, Wilson, & Haldane, 2003).

4.3 Results

4.3.1 Descriptive statistics: Computer-based thinking skill assessment is feasible in China

The pilot test was successfully implemented. According to the interviews after the assessment, students had no problem with the operation of our computer-based assessments. Students' mean performance for each thinking skill test (problem-solving, inductive reasoning, and creativity tests) was 47.73% (SD=21.78%), 80.33% (8.63%), and 2.91 (SD=4.02), respectively. The Cronbach's alpha for these three tests ranged from .72 to.90. (Table 4.3)

Test	Mean	SD	Cronbach 's alpha
Problem Solving	47.73%	21.78%	.72
Inductive reasoning	80.33%	8.63%	.75
Creativity	2.91	4.02	.90

Table 4.3 Mean, standard deviation and Cronbach's alpha for each thinking skill test

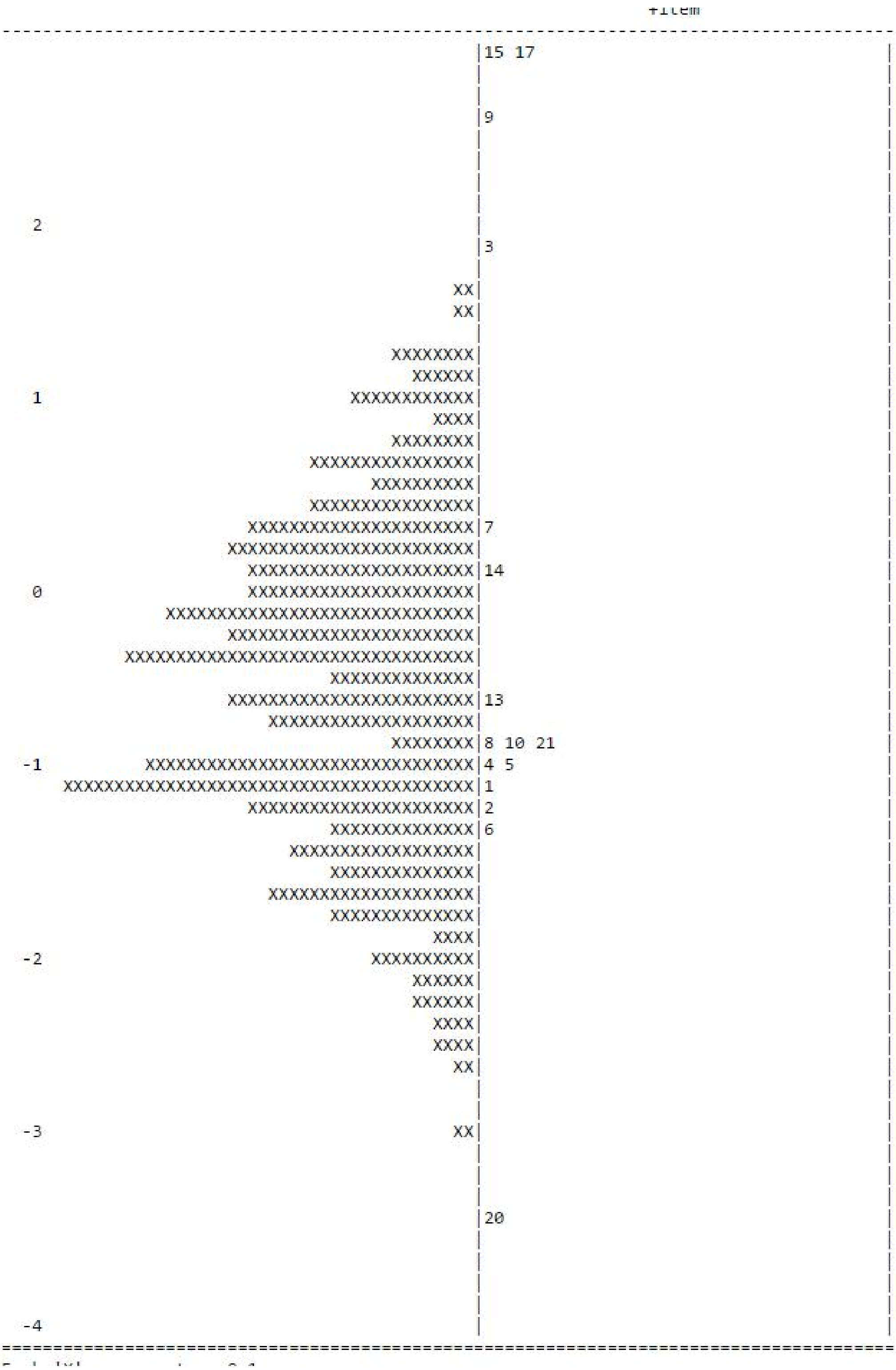
The primary aim of the pilot study was to explore the applicability of online assessments in China and to test the reliability of every thinking skill test involved in the project. Firstly, there was no problem with running the tests on the eDia online assessment platform in the Chinese network environment. Therefore, the feasibility was proven. As indicated above, the IPS and inductive reasoning tests showed acceptable internal consistencies; in the meanwhile, the internal consistency for the creativity test was satisfactory. To answer research question RQ1, the results proved that computer-based assessment of thinking skills is feasible and reliable in China, and the tests are also reliable to measure Chinese 12-year-old students' thinking skills. Thus, *hypothesis H1 was supported.*

4.3.2 The behaviour of the IPS test in China: dimensionality and difficulty

Rasch analysis was conducted to analyze whether the difficulty level of the IPS test fits to students' ability level. The results indicate that some IPS items proved to be difficult for the students (numbered with 3, 9, 15 and 17; see Fig. 4.8), while others were very easy (item numbered with 20). All of the hard items contained eigendynamics (see section 2.3.3), which increased the difficulty level of the item in a significant manner.

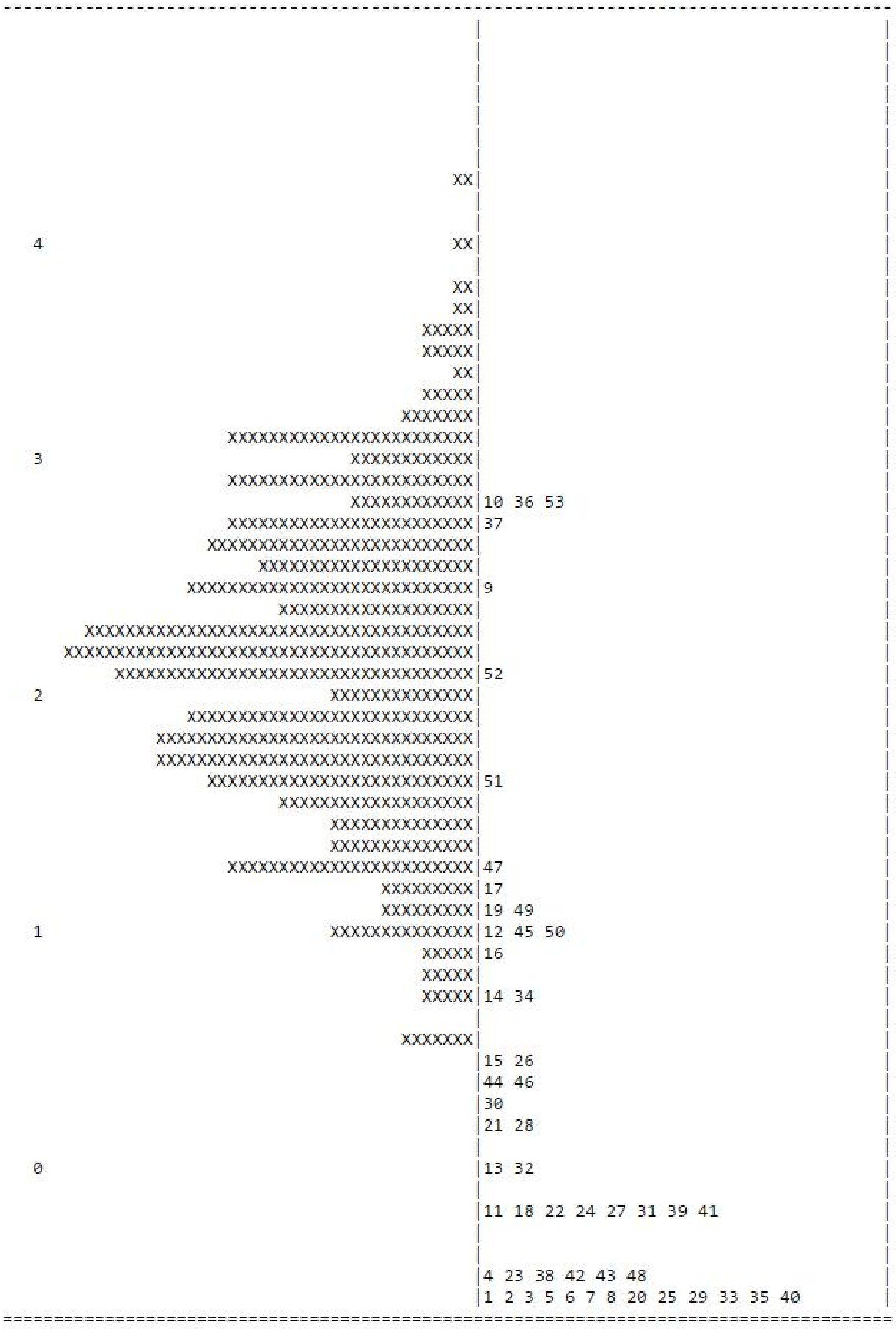
As for the inductive reasoning test, there were a great number of items which were too easy for the participating students. These items mainly belonged to the figural analogy and figural series subtests (Fig. 4.9).

Based on the literature, we tested the two-dimensional measurement model including the IPS processes knowledge acquisition and knowledge application to confirm the dimensionality of IPS in China. Beyond the two-dimensional measurement model, we also tested a one-dimensional model with both processes combined under one general factor. In the one-dimensional model, two sub-factors (knowledge acquisition and knowledge application) were considered as one general factor as problem solving; while the two-dimensional model created a latent variable as problem solving which consisted of knowledge acquisition and knowledge application these two factors.



```
                                                               +item
---------------------------------------------------------------------------------------
                                              |15 17                                  |
                                              |                                       |
                                              |                                       |
                                              |9                                      |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
   2                                          |                                       |
                                              |3                                      |
                                              |                                       |
                                            XX|                                       |
                                            XX|                                       |
                                              |                                       |
                                      XXXXXXXX|                                       |
                                        XXXXXX|                                       |
   1                              XXXXXXXXXXXX|                                       |
                                          XXXX|                                       |
                                      XXXXXXXX|                                       |
                              XXXXXXXXXXXXXXXX|                                       |
                                    XXXXXXXXXX|                                       |
                              XXXXXXXXXXXXXXXX|                                       |
                        XXXXXXXXXXXXXXXXXXXXXX|7                                      |
                      XXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                        XXXXXXXXXXXXXXXXXXXXXX|14                                     |
   0                    XXXXXXXXXXXXXXXXXXXXXX|                                       |
                XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                      XXXXXXXXXXXXXXXXXXXXXXXX|                                       |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                                XXXXXXXXXXXXXX|                                       |
                      XXXXXXXXXXXXXXXXXXXXXXXX|13                                     |
                          XXXXXXXXXXXXXXXXXXXX|                                       |
                                      XXXXXXXX|8 10 21                                |
  -1          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|4 5                                    |
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|1                                      |
                        XXXXXXXXXXXXXXXXXXXXXX|2                                      |
                                XXXXXXXXXXXXXX|6                                      |
                            XXXXXXXXXXXXXXXXXX|                                       |
                                XXXXXXXXXXXXXX|                                       |
                          XXXXXXXXXXXXXXXXXXXX|                                       |
                                XXXXXXXXXXXXXX|                                       |
                                          XXXX|                                       |
  -2                                XXXXXXXXXX|                                       |
                                        XXXXXX|                                       |
                                        XXXXXX|                                       |
                                          XXXX|                                       |
                                          XXXX|                                       |
                                            XX|                                       |
                                              |                                       |
                                              |                                       |
  -3                                        XX|                                       |
                                              |                                       |
                                              |                                       |
                                              |20                                     |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
  -4                                          |                                       |
=======================================================================================
```

Fig. 4.8 Item-person map for the IPS test



```
---------------------------------------------------------------------------------------
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                            XX|                                       |
                                              |                                       |
                                              |                                       |
   4                                        XX|                                       |
                                              |                                       |
                                            XX|                                       |
                                            XX|                                       |
                                         XXXXX|                                       |
                                         XXXXX|                                       |
                                            XX|                                       |
                                         XXXXX|                                       |
                                       XXXXXXX|                                       |
                      XXXXXXXXXXXXXXXXXXXXXXXX|                                       |
   3                              XXXXXXXXXXXX|                                       |
                      XXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                                  XXXXXXXXXXXX|10 36 53                               |
                      XXXXXXXXXXXXXXXXXXXXXXXX|37                                     |
                    XXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                         XXXXXXXXXXXXXXXXXXXXX|                                       |
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXX|9                                      |
                           XXXXXXXXXXXXXXXXXXX|                                       |
        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|52                                     |
   2                                XXXXXXXXXXXXXX|                                   |
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                    XXXXXXXXXXXXXXXXXXXXXXXXXX|51                                     |
                           XXXXXXXXXXXXXXXXXXX|                                       |
                                XXXXXXXXXXXXXX|                                       |
                                XXXXXXXXXXXXXX|                                       |
                      XXXXXXXXXXXXXXXXXXXXXXXX|47                                     |
                                     XXXXXXXXX|17                                     |
                                     XXXXXXXXX|19 49                                  |
   1                            XXXXXXXXXXXXXX|12 45 50                               |
                                         XXXXX|16                                     |
                                         XXXXX|                                       |
                                         XXXXX|14 34                                  |
                                              |                                       |
                                       XXXXXXX|                                       |
                                              |15 26                                  |
                                              |44 46                                  |
                                              |30                                     |
                                              |21 28                                  |
                                              |                                       |
   0                                          |13 32                                  |
                                              |                                       |
                                              |11 18 22 24 27 31 39 41                |
                                              |                                       |
                                              |                                       |
                                              |4 23 38 42 43 48                       |
                                              |1 2 3 5 6 7 8 20 25 29 33 35 40        |
=======================================================================================
```

Fig. 4.9 Item-person map for the inductive reasoning test

Both models showed a good model fit (Table 4.4). In order to test which model fitted the data better, we carried out the special chi-square test for difference testing computed by Mplus. The results indicate that, there was no significant difference detectable between these two models (p>.05). IPS can be explained as both one-dimensional and two-dimensional construct; however, due to the small sample size of the pilot study, we cannot make large generalizations of it. The follow-up comparison study will further analyze and confirm the dimensionality of IPS. As the two-dimensional model is favored and empirically proved in all large-scale assessments in Europe, and its Chi-square value was smaller, we decided to use it in our further analyses.

Model	Chi-square	df	RMSEA	CFI	TLI
1-dimensional	15.257	16	.00	1.00	1.01
2-dimensional	14.280	15	.00	1.00	1.01

Table 4.4 Goodness of fit indices for testing dimensionality of IPS

4.3.3 Students' test-taking motivation in IPS environment

The test-taking motivation was analyzed separately with other factors, since the motivation questionnaire items were only about students' test-taking motivation during the IPS assessment. It was meaningless and not precise to analyze students' performance in other thinking skill assessments and their test-taking motivation for the IPS assessment.

Students responded to the test-taking motivation questions using the Likert scale from ”1-strongly disagree” to ”7-strongly agree”. Assuming opinion 7 “strongly agree” for every single question as 100%, for the positive questions, the average response for the students was 83.85% (SD=16.97%)., while for the negative questions, the average response was 39.56% (SD=23.36%). The results demonstrated students' high motivation in the problem-solving assessment.

Significant correlations were found between students' performance in problem-solving assessment and their responses to the motivation questions (for both positive (r=.384, p<.05) and negative questions (r=-.340, p<.05)). The results of the analysis effectively proved the influence of test-taking motivation on the test-takers' performance in the IPS environment.

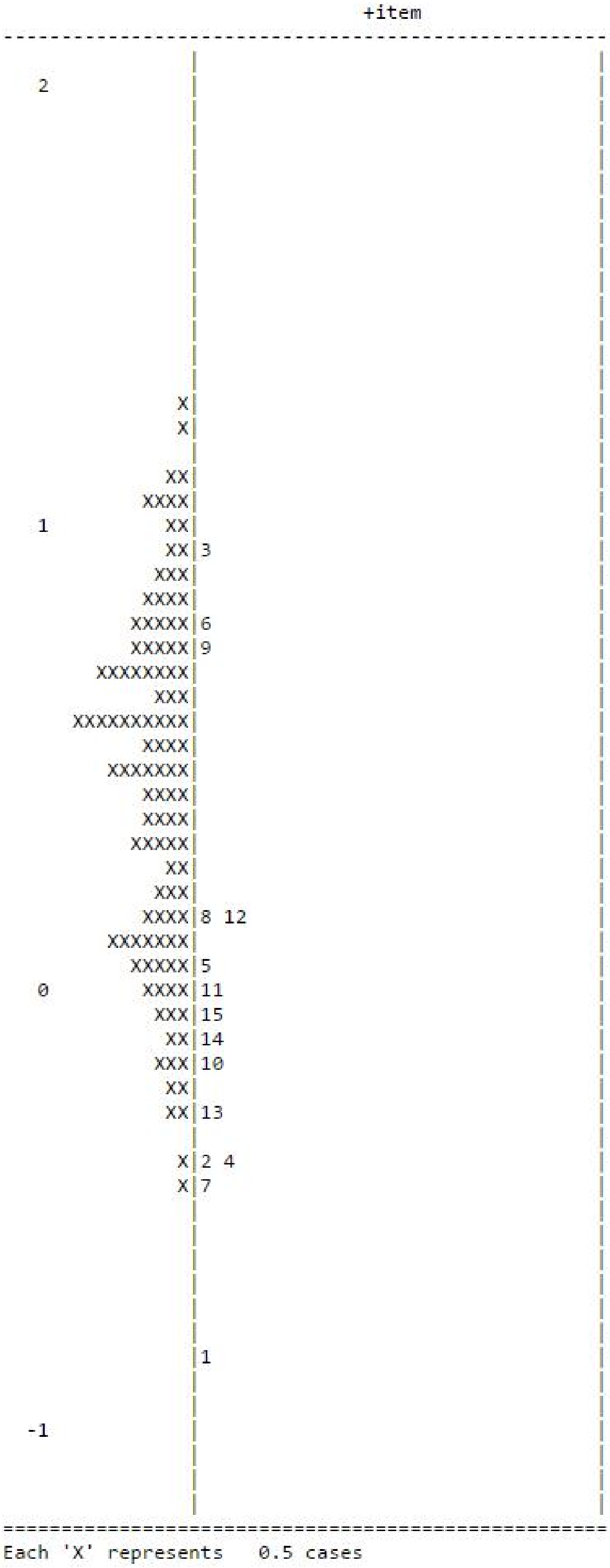
Item response theory was used to analyze students' motivation level in the interactive problem-solving environment. Firstly, the model fits for rating scale model and partial credit model based on the assessment data were calculated.

Model Type	*Total number of estimated parameters*	*Final deviance*
Rating Scale Model	21	2130.256
Partial Credit Model	80	1995.223

Table 4.5 Model fits for rating scale model and partial credit model

The analysis indicated a significant difference between these two models (p<.001); meanwhile the partial credit model showed a better model fit. Therefore, the partial credit model was chosen for the analysis. Based on the partial credit model, the item person map was calculated and drawn (see Fig. 4.10). The distribution of students (represented by "x"s) in the map indicates that the overall motivation turns to be high. Furthermore, students were much more likely to choose affirmative responses for the positive questions (question 1, 2, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15), and choose negative responses for the negative questions (question 3, 6, 9).

Students held the most affirmative attitude to question 1, followed by questions 2, 4 and 7 (for the content of each question, please see table 4.1). This means that students believed this assessment to be interesting, either before, during or after this assessment. The map also shows that students had strong confidence to do a good job at the very beginning (based on question 2), but their confidence level dropped after they actually started the assessment (based on questions 5 and 8). Similarly, from the positions of questions 3, 6 and 9 in the map, it can be seen that at first students did not think this test would be difficult to do. But as the assessment proceeded, students found this test to be a little harder than their expectation. The distribution for the rest of the items proved students' overall positive attitude and high test-taking motivation in the IPS environment.

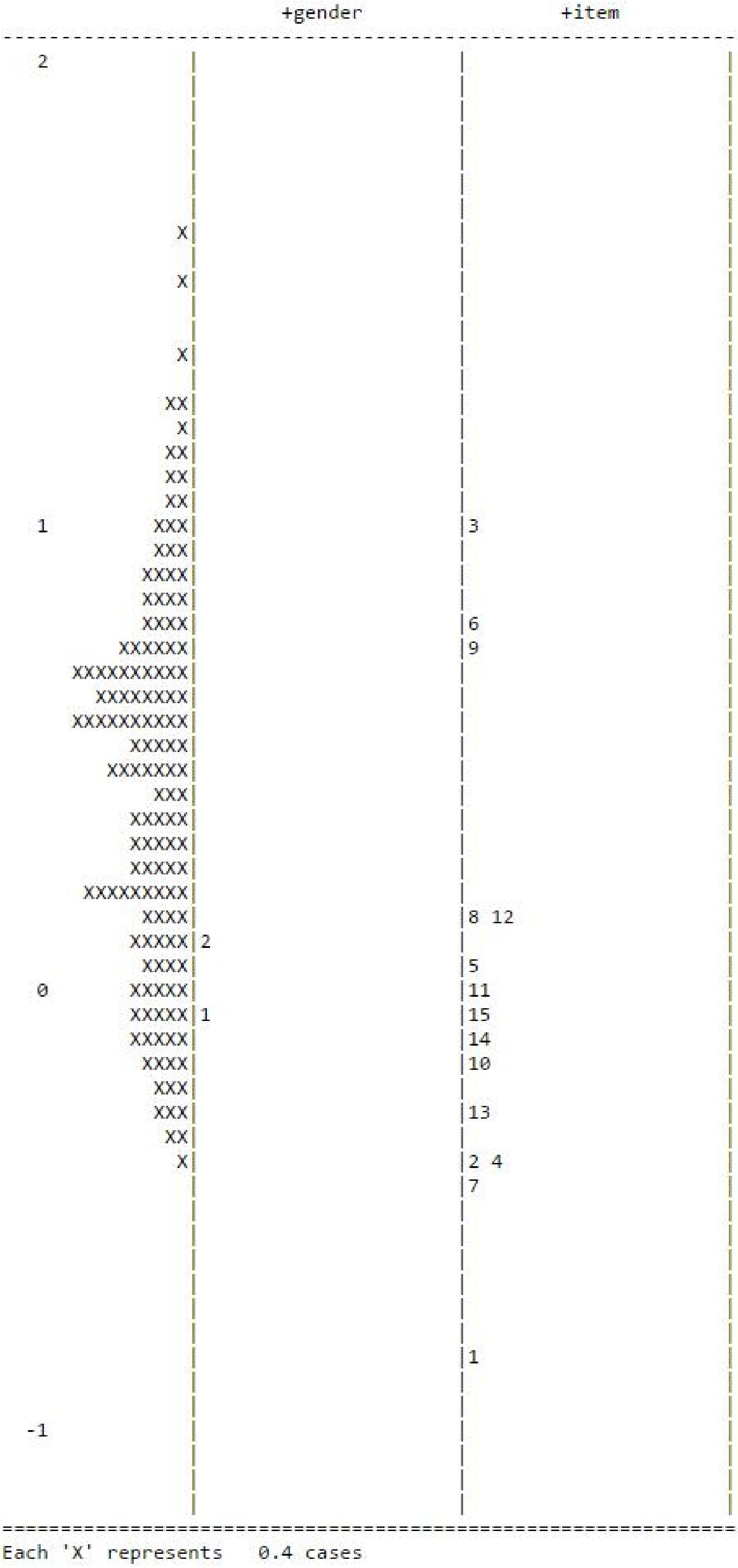


```
                               +item
----------------------------------------------------
                |                                  |
   2            |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
               X|                                  |
               X|                                  |
                |                                  |
              XX|                                  |
            XXXX|                                  |
   1          XX|                                  |
              XX|3                                 |
             XXX|                                  |
            XXXX|                                  |
           XXXXX|6                                 |
           XXXXX|9                                 |
        XXXXXXXX|                                  |
             XXX|                                  |
      XXXXXXXXXX|                                  |
            XXXX|                                  |
         XXXXXXX|                                  |
            XXXX|                                  |
            XXXX|                                  |
           XXXXX|                                  |
              XX|                                  |
             XXX|                                  |
            XXXX|8 12                              |
         XXXXXXX|                                  |
           XXXXX|5                                 |
   0        XXXX|11                                |
             XXX|15                                |
              XX|14                                |
             XXX|10                                |
              XX|                                  |
              XX|13                                |
                |                                  |
               X|2 4                               |
               X|7                                 |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |1                                 |
                |                                  |
                |                                  |
  -1            |                                  |
                |                                  |
                |                                  |
                |                                  |
====================================================
Each 'X' represents   0.5 cases
```

Fig. 4.10 Item-person map for the motivation questions

In order to compare the gender differences in test-taking motivation, a model which considered gender as a factor was built. The item-person map (Fig. 4.11) showed that boys (represented by 1) were more likely to choose the affirmative response for the questions, while girls' (represented by 2) attitude towards the questions was less positive,which means that in this assessment boys had higher motivation.

The results indicated the influence of test-taking motivation on one's performance in the IPS assessment. Chinese students had overall high motivation in the IPS environment, while boys' motivation was higher than that of girls. Students showed a strong interest in this assessment format. However the result indicated that students might have small level of misjudgment about the difficulty level of the assessment, which impacted their motivation to a certain extent. The result suggested that interactive problem solving can be an interesting and welcomed assessment in the China context, even though it was delivered by a low-stakes test format. However, a study with a larger sample size is still needed to confirm the findings.



```
                        +gender                  +item
----------------------------------------------------------------
   2            |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |                      |
               X|                       |                      |
                |                       |                      |
               X|                       |                      |
                |                       |                      |
                |                       |                      |
               X|                       |                      |
                |                       |                      |
              XX|                       |                      |
               X|                       |                      |
              XX|                       |                      |
              XX|                       |                      |
              XX|                       |                      |
   1         XXX|                       |3                     |
             XXX|                       |                      |
            XXXX|                       |                      |
            XXXX|                       |                      |
            XXXX|                       |6                     |
          XXXXXX|                       |9                     |
      XXXXXXXXXX|                       |                      |
        XXXXXXXX|                       |                      |
      XXXXXXXXXX|                       |                      |
           XXXXX|                       |                      |
         XXXXXXX|                       |                      |
             XXX|                       |                      |
           XXXXX|                       |                      |
           XXXXX|                       |                      |
           XXXXX|                       |                      |
       XXXXXXXXX|                       |                      |
            XXXX|                       |8 12                  |
           XXXXX|2                      |                      |
            XXXX|                       |5                     |
   0       XXXXX|                       |11                    |
           XXXXX|1                      |15                    |
           XXXXX|                       |14                    |
            XXXX|                       |10                    |
             XXX|                       |                      |
             XXX|                       |13                    |
              XX|                       |                      |
               X|                       |2 4                   |
                |                       |7                     |
                |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |1                     |
                |                       |                      |
                |                       |                      |
  -1            |                       |                      |
                |                       |                      |
                |                       |                      |
                |                       |                      |
================================================================
Each 'X' represents   0.4 cases
```

Fig. 4.11 Item-person map with gender difference

4.3.4 The relationships between students' performance in problem solving, inductive reasoning, creativity, working memory and learning strategies

A correlation matrix was built to indicate the relations between the developmental level of the different thinking skills. From the correlation matrix (Table 4.6) it can be seen that significant correlations were detectable between problem solving – inductive reasoning and inductive reasoning – creativity. But there was no significant correlation between problem solving and creativity. However, since the correlation analysis was based on the small sample size, some correlation relationships were hard to be detected and confirmed in the statistical means. But based on the result, at least the influence between problem solving and inductive reasoning was confirmed.

	Problem Solving	Inductive Reasoning	Creativity
Problem Solving	1.00		
Inductive Reasoning	.440**	1.00	
Creativity	n.s	.363*	1.00

Note. **Correlation is significant at the 0.01 level. *Correlation is significant at the 0.05 level.

Table 4.6 Correlation matrix between the measured thinking skills

Working memory also showed significant correlations with problem solving ($r=.522$, $p<.05$), but no significant correlation was detected between working memory and inductive reasoning; and working memory and creativity.

No significant gender difference was detected. Independent t-test showed that girls performed equally with boys in creativity ($t=1.07$, $p>.05$), problem solving ($t=.10$, $p>.05$), and inductive reasoning ($t= -.89$, $p>.05$) tests. Independent t-test also indicated that students with different parental educational level also showed statistically equal performance in the thinking skill tests ($p>.05$).

Students' regularly applied learning strategies were correlated with their developmental level of thinking skills. Students were divided into different groups by their most commonly used learning strategies (elaboration, control, and memorization strategies). Significantly different test performances were found between the groups of students having different learning style.

The independent t-test results indicated that students who preferred memorization strategies in their study had significantly lower performance in the inductive reasoning test than students who did not prefer memorization during learning (t= -2.942, p<.05). In the meanwhile, students who preferred control strategies performed significantly better in the problem solving (t=2.194, p<.05), in the inductive reasoning (t=2.806, p<.05) and in the creativity (t=2.197, p<.05) tests, too, compared to their peers. The results highlighted the importance of the regularly applied learning strategies on students' developmental level of thinking skills. Results also showed the possibility that instructors might help students to develop thinking skills by training and encouraging students to apply more effective learning strategies.

4.3.5 The influencing factors of problem solving

To test the component skills of problem solving, and the factors influencing students' problem solving skills, we built a structural equation model of problem solving and its possible component skills (Fig. 4.12). As the two-dimensional measurement model of problem solving showed a good model fit (RMSEA=.00; SRMR=.03; CFI=1.00; TLI=1.02), we applied problem solving as a latent construct in the model, which can be explained by knowledge acquisition and application (β=.496-.669; p<.001). Inductive reasoning, and working memory showed significantly positive influence (β=.453-.594, p<.05) on one's problem solving achievement in the SEM model, while working memory and inductive reasoning proved to be significantly correlated constructs (r=.260, p<.05).

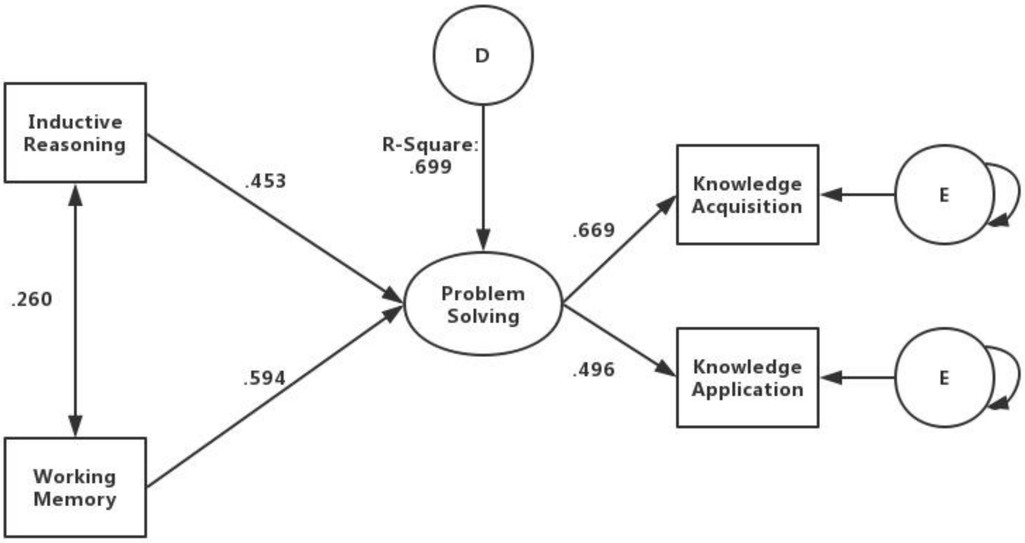

D
Inductive Reasoning
R-Square:
.699
.453
Knowledge Acquisition
E
.669
.260
Problem Solving
.594
.496
Knowledge Application
E
Working Memory

Fig. 4.12 Structural equation model of problem solving and its influencing factors

The other factors included in the pilot study did not show any statistically significant effect on problem solving in the SEM model, however, the pilot study had a very strong limitation with its small sample size. The analysis confirmed the relations between problem solving and some specific factors, and proved the feasibility of using structural equation modeling as the tool for latter analyses regarding the component skills of problem solving.

4.4 Conclusions of the pilot study

The main aim of the pilot study was to answer research question RQ1. The results indicated that computer-based assessment and the online assessment platform (eDia) are applicable in China. The tests of problem solving, inductive reasoning and creativity were reliable to measure students' thinking skills in China. The result gave a positive answer to the basic but most important question, whether the project – defining the component skills of problem solving and running comparison studies between China and Hungary – is feasible and realizable from the Chinese educational context point of view.

Furthermore, the results proved the theoretical statement: one's problem-solving achievement is influenced by some specific thinking skills and non-thinking skill factors. We preliminary confirmed the influence of some reasoning skill such as inductive reasoning, and some non-cognitive factors such as working memory and learning strategies on students' problem-solving achievement. To conclude, the findings effectively testified the feasibility of this project, and built a solid foundation for future study.

4.5 Limitations of the pilot study

The pilot study pointed out some limitations and drawbacks regarding the assessment design, and some improvements were made based on these.

Firstly, a significant system bug showed up in the pilot test. Each IPS assessment item had a three minutes time limitation. In the pilot test, there were several Chinese students who did not provide any answer in the three minutes, thus caused a bug and they could not continue the assessment. During the pilot test, the students who got this problem were provided a second chance to take the assessment. The bug was fixed after the pilot.

However, this system bug had never occurred before in Hungary. The reason behind this situation is an interesting topic to be focused on. It can be assumed that Asian students and European students use different strategies when working on the problem-solving items, thus caused this issue. Or it can be assumed that Chinese students are highly motivated, so they were willing to try every single possible solution before they actually gave their answer, and that made them use all the provided time. The following sections will try to discover what the real reason behind this is.

The second issue identified is about the scoring work regarding the creativity test. The results indicated that creativity does not have a statistically significant connection with problem-solving ability. In addition, the scoring for the creativity assessment can only be realized by manual work, which was unexpectedly time-consuming. Considering the further assessment session with a larger group of students, creativity assessment was deleted from the project. Instead of this, another important thinking skill, combinatorial reasoning (see section 2.4.2) was included in the project.

Last but not least, the Rasch model analyses pointed out that the tests were not perfectly fit to the students' ability level. Section 4.3.2 introduced that the problem-solving test was a bit difficult for the students. Meanwhile, the analyses also showed that some items in the inductive reasoning test were too easy for the participating students.

Based on the analysis, the difficulty level of the tests was modified. The analyses proved that the IPS items which contained eigendynamics were very hard for the students. Thus, in further comparison study, these items were deleted. Meanwhile, items with eigendynamics were included in the Hungarian problem-solving test.

As for the inductive reasoning test, the easy items mainly belonged to the figural analogy and figural series subtests. In comparison, items from the number series subtest had remarkably higher difficulty level. Thus, in the further comparison study, some easy items were deleted, while some harder items (mainly number series items) were added. However, as for the cross-county comparison study we needed a larger number of anchor items between the Hungarian and the Chinese versions of the tests, some easy items still remained for the further analyses in the final version of the inductive reasoning test, as anchor items.

5. Cross-national comparison study of problem solving and its influential factors in educational context

5.1 Aims, research questions and hypotheses of the comparison study

The pilot study confirmed the feasibility of computer-based assessment in China, thus built good foundation for the cross-national comparison study. Based on that, a Chinese-Hungarian cross-national comparison study of problem solving and its' influential factors was conducted. This comparison study aimed to answer the majority of the research questions. Because the creativity test was removed from the assessment (see section 4.5), corresponding change was made in research question RQ6.

RQ2. Is the behaviour of the tasks measuring knowledge acquisition and knowledge application as two dimensions of IPS in the MicroDYN approach independent of cultural context?

RQ3. Can IPS be measurement invariant across gender and nationality both in the contexts of Europe and Asia?

RQ4. Can developmental differences in IPS be detected between Hungarian and Chinese 12-year-old students, 3 years before the PISA age?

RQ5. What kind of developmental differences can be identified across gender and nationality, between Chinese males and females, and between Hungarian males and females?

RQ6. Are inductive reasoning and combinatorial reasoning component skills of problem solving? How strong is their predictive power on the problem solving achievement of Chinese and Hungarian 12-year-old students?

RQ7. Which factors influence Chinese and Hungarian students' IPS achievement beyond thinking skills? How strong is the influential effect?

The following hypotheses are expected to be tested based on the results of this comparison study. Hypothesis H6 removed the creativity part as well.

H2: IPS is composed of two different processes, knowledge application and knowledge acquisition, independently of the cultural context. That is, a two-dimensional model was expected to show significantly better fit than a one-dimensional model with the two processes combined as one factor.
H3: IPS can be measurement invariant between male and female, and Hungary and China.
H4: Developmental differences in IPS can be detected between Hungarian and Chinese 12-year-old students.
H5: Chinese 12-year-old students have significantly better performance than Hungarian students in the same age group; Chinese and Hungarian boys' performance is statistically higher or equal to that of the girls.
H6: Inductive reasoning and combinatorial reasoning are component skills of problem solving, which have significant predictive power on the problem solving achievement by Chinese and Hungarian 12-year-old students.
H7: The selected non-thinking skill factors (working memory, parents' education, learning strategies, and test-taking motivation) have a significant influential effect on Chinese and Hungarian students' IPS achievement.

5.2 Methods

5.2.1 The sample of the comparison study

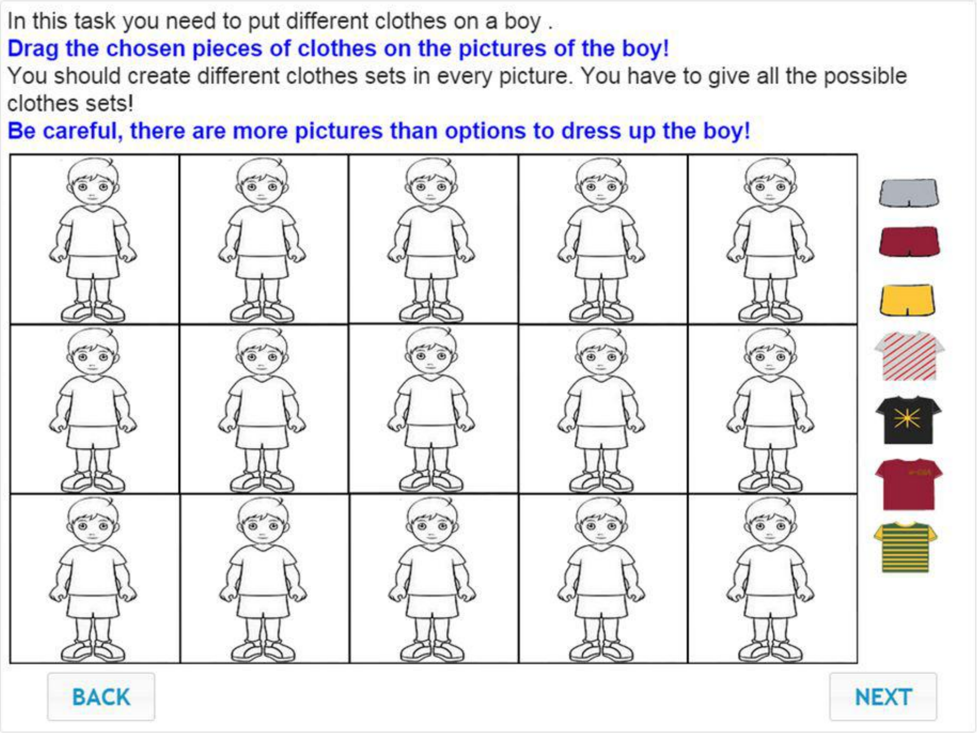
The sample was drawn from 6th grade, 12-year-old students in Hungarian and Chinese primary schools. 187 Chinese students (85 boys and 102 girls; mean age=11.93, SD=1.06) and 835 Hungarian students (382 boys and 453 girls, mean age=11.86, SD=.43) took part in the study. The sampling process emphasized the background matching, thus these two groups of students had similar social and family backgrounds (include the size of living city, type of school, parents’ educational level, etc.).

5.2.2 Instruments

Some modifications were made in the tests based on the results of the pilot study. These are the followings:

1. From the IPS test we deleted all the problems having eigendynamics, because those proved to be too difficult for the 12-year-old students, based on the pilot study. Due to time limitation, we reduced the number of the test-taking motivation items in the problem-solving test to 9 items (item 1-9, see Table 4.1).
2. We have increased the difficulty level of the inductive reasoning test for Chinese students by deleting easy items and adding harder items to the test (see section 4.5).
3. We decided to exclude the creativity test from the study (see section 4.5).
4. Based on the literature review, we decided to include combinatorial reasoning in the comparison study. The combinatorial reasoning test was originally developed and designed by Csapó (1999) as a paper-and-pencil test. Pásztor and Csapó (2014) improved and computerized it. In the comparison study we adapted Pásztor and Csapó's (2014) design and translated the assessment items into simplified Chinese for the Chinese assessment (see Appendix C). Language experts have double checked the translation. Students needed to use given elements to create combinations which satisfied the given requirement. According to the elements given in the tasks, the test can be divided into two sub-tests containing figural or verbal items (Fig. 5.1 & Fig. 5.2). For the figural items (Fig. 5.1), students were required to combine figures and pictures and create different combinations of them by means of drag-and-drop. For the verbal items (Fig. 5.2), students were required to create combinations of the given letters and/or numbers, and type their answers into the input box. Student's score for the combinatorial reasoning test was automatically calculated by the eDia platform using the formula: [(the number of

the correct combinations student answered)/(the number of all the possible combinations)]*100%.



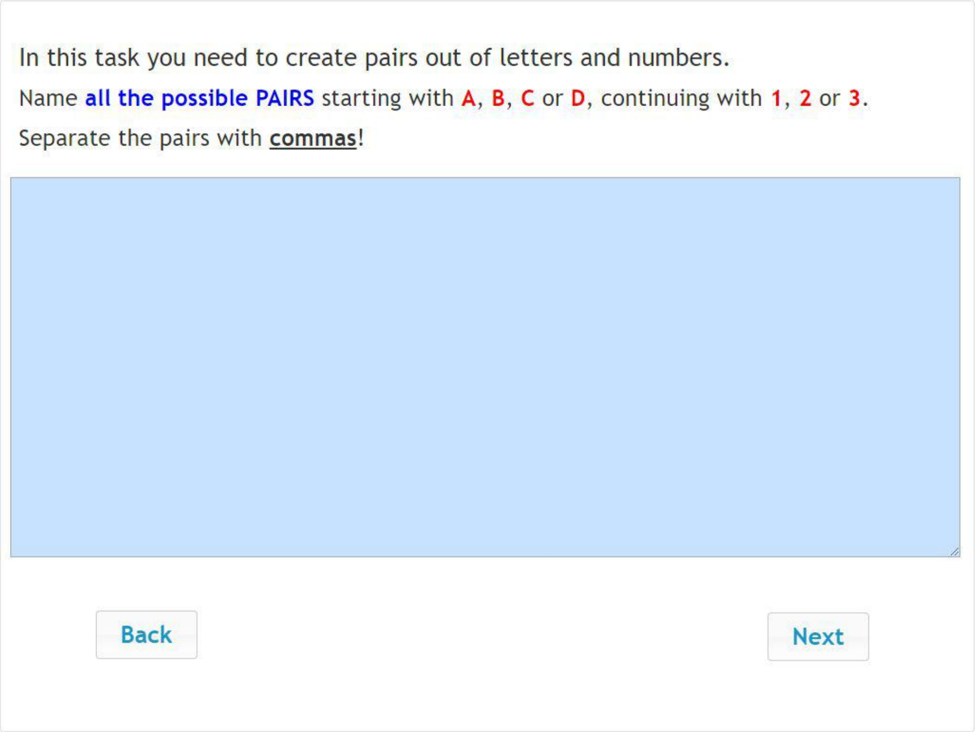
In this task you need to put different clothes on a boy .

Drag the chosen pieces of clothes on the pictures of the boy!

You should create different clothes sets in every picture. You have to give all the possible clothes sets!

Be careful, there are more pictures than options to dress up the boy!

Fig. 5.1 Sample item for the combinatorial reasoning test-figural



In this task you need to create pairs out of letters and numbers.

Name **all the possible PAIRS** starting with **A**, **B**, **C** or **D**, continuing with **1**, **2** or **3**.

Separate the pairs with **commas**!

Fig. 5.2 Sample item for the combinatorial reasoning test-verbal

The Chinese and the Hungarian test versions of the thinking skills were not completely the same. Thus, a direct comparison of the results by means of classical test theory was not possible. However, both the Chinese and the Hungarian test versions contained a certain proportion of anchoring items, thus, by means of items response theory we were able to represent Chinese and Hungarian students' ability level on a common scale.

5.2.3 Procedures

The whole assessment was carried out using the eDia (Electronic Diagnostic Assessment; Molnár, 2015) platform in the schools' ICT room in June and July, 2017. Both the online assessment system and the tests had been conducted in Hungary several times, thus the feasibility of online assessment and the behaviour, particularly the reliability of the tests in the Hungarian educational context could be proved by the former assessment projects (e.g. Molnár, 2015; Molnár et al., 2013; Csapó, & Molnár, 2017). The feasibility and reliability of online assessment via the eDia platform in the Chinese context have been established by the pilot study (see Chapter 4). The assessment took two hours in total, divided into three 40-minute sessions.

Session 1: IPS assessment (combined with test-taking motivation questionnaire)
Session 2: Inductive reasoning assessment, working memory assessment
Session 3: Combinatorial reasoning assessment, background questionnaire (see Appendix F)

All the items were administered in simplified Chinese in China and in Hungarian in Hungary. For all the tests, students' scores were automatically recorded and calculated by the eDia platform.

5.2.4 Data analyses

SPSS version 22 was used to conduct the descriptive statistics and basic analyses (e.g. correlation analysis, independent t-test, etc.). Item response theory (Rasch Model) was selected as the tool for transferring Chinese and Hungarian students' performance into a common scale. For calculating their ability level, weighted likelihood estimates (WLE) was chosen as estimator. Item response theory was also used to analyze the difficulty level of the items and the test-taking motivation. The item response theory analyses were run by ConQuest (Wu et al., 2003). Structural equation modeling (SEM; Bollen, 1989) was the main tool for data analysis in this study. It was used to test the measurement invariance, the structure of the different reasoning skills assessed, as well as the relationships between these skills. The analyses were run by the software Mplus (Muthén, & Muthén, 2010). Maximum Likelihood (ML) estimation was used in the different estimation processes. The Tucker-Lewis Index (TLI), the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) were computed for testing the fitting of the established model by Mplus and used as an indicator for the aptness of the models.

5.3 Descriptive statistics of the thinking skill tests

5.3.1 Descriptive statistics of the thinking skill tests– China

Table 5.1 shows the basic statistical information: the number of items, reliability coefficients (Cronbach's alpha), assessment score means and standard deviations for the subscales of problem solving, combinatorial reasoning and inductive reasoning. The reliability indices were satisfactory for every subscale, ranging from .79 to .94. The high internal consistencies confirmed that the assessment was reliable.

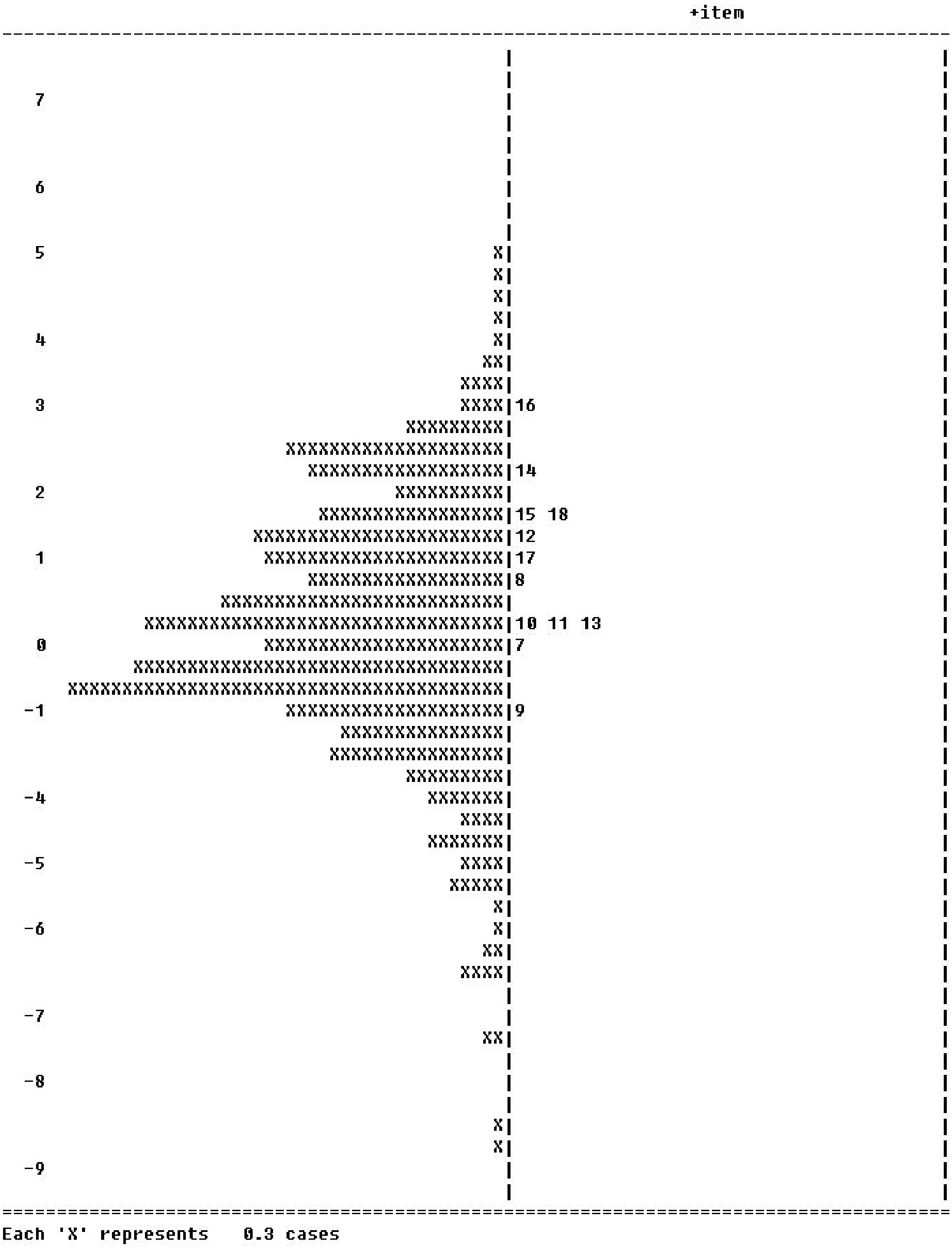
The means for the problem-solving and combinatorial reasoning tests ranged from 35% to 45%, which was a little lower than the assumed optimal value (40%-60%), but still ideal for analysis. The mean values for the inductive reasoning subscales varied widely (38%-77%), which was caused by the different levels of difficulty for each subscale. Students' performance in inductive reasoning was close to the initial assumption, and also suitable for analysis

In the pilot test, there was a problem with the piloted tests identified (see section 4.5). Both in the IPS and in the inductive reasoning tests, the students' ability level and the items' difficulty level did not meet, which could damage the accuracy of the analysis. A difficulty level adjustment was made after the pilot study. Analyses by means of Rasch model were conducted to detect the effects of the adjustment.

	Number of Items	Cronbach's alpha	Mean (%)	Standard Deviation (%)
Problem Solving	18	.90	40.20	28.30
Knowledge acquisition	9	.87	45.34	32.30
Knowledge application	9	.79	35.06	26.08
Combinatorial Reasoning	12	.93	39.87	27.34
Figural	6	.91	42.59	27.97
Verbal	6	.92	37.15	32.64
Inductive Reasoning	42	.96	60.09	25.04
Figural series	11	.85	77.30	25.77
Figural analogy	15	.94	73.26	32.28
Number analogy	8	.82	51.40	30.85
Number series	8	.93	38.40	32.33

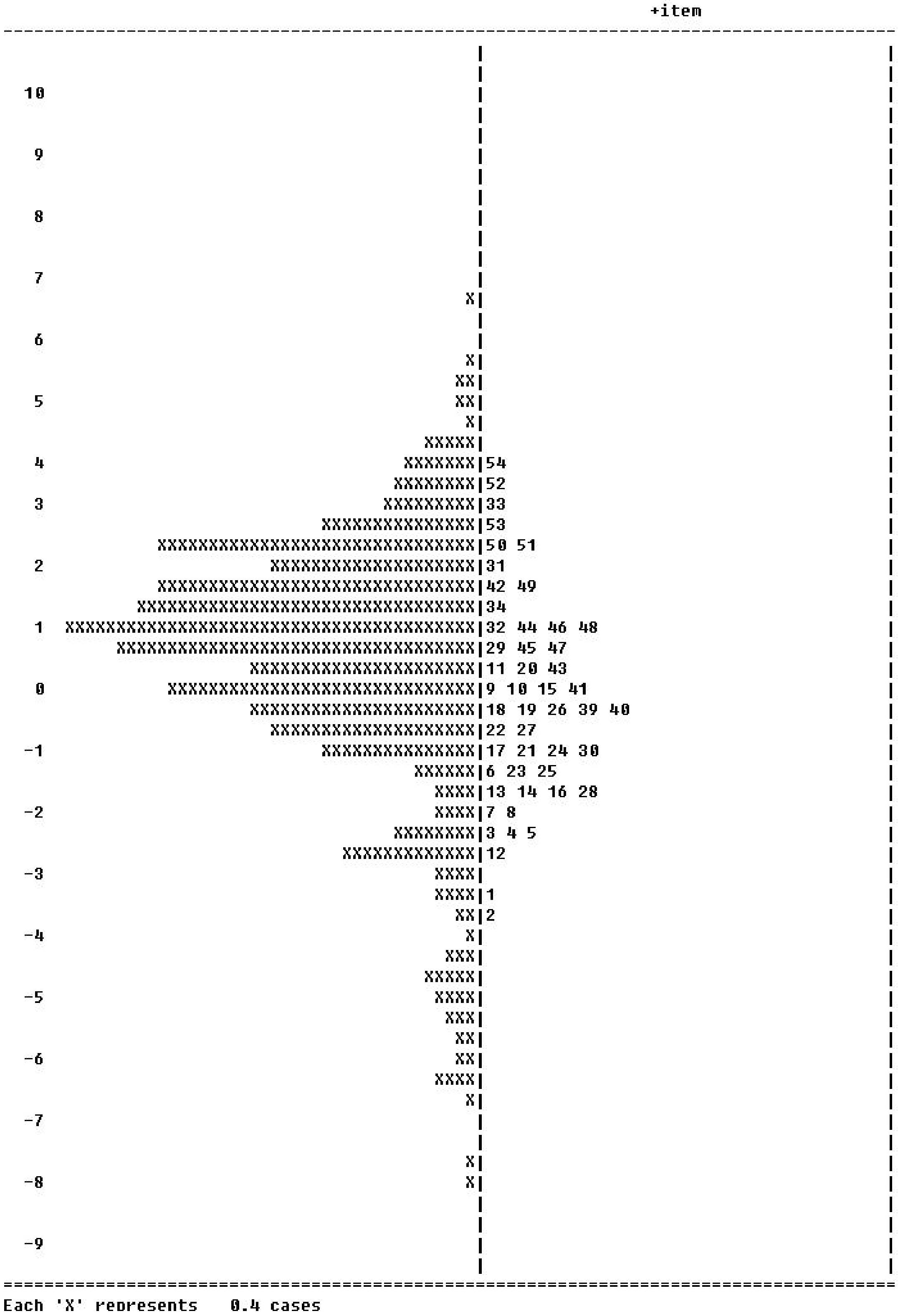
Table 5.1 Descriptive statistics for each assessed thinking skill and their subscales – Chinese sample

Based on the item-person maps, we concluded that after the adjustment, the difficulty levels of the IPS (Fig. 5.3) and inductive reasoning (Fig. 5.4) tests fitted much more to the students' ability levels, which was a great benefit for further analysis. The combinatorial reasoning test was not included in the pilot study, thus, we did not have data regarding its behaviour. According to the item-person map (Fig. 5.5), it lacked some easy items, but still generally fitted to the students' ability level.



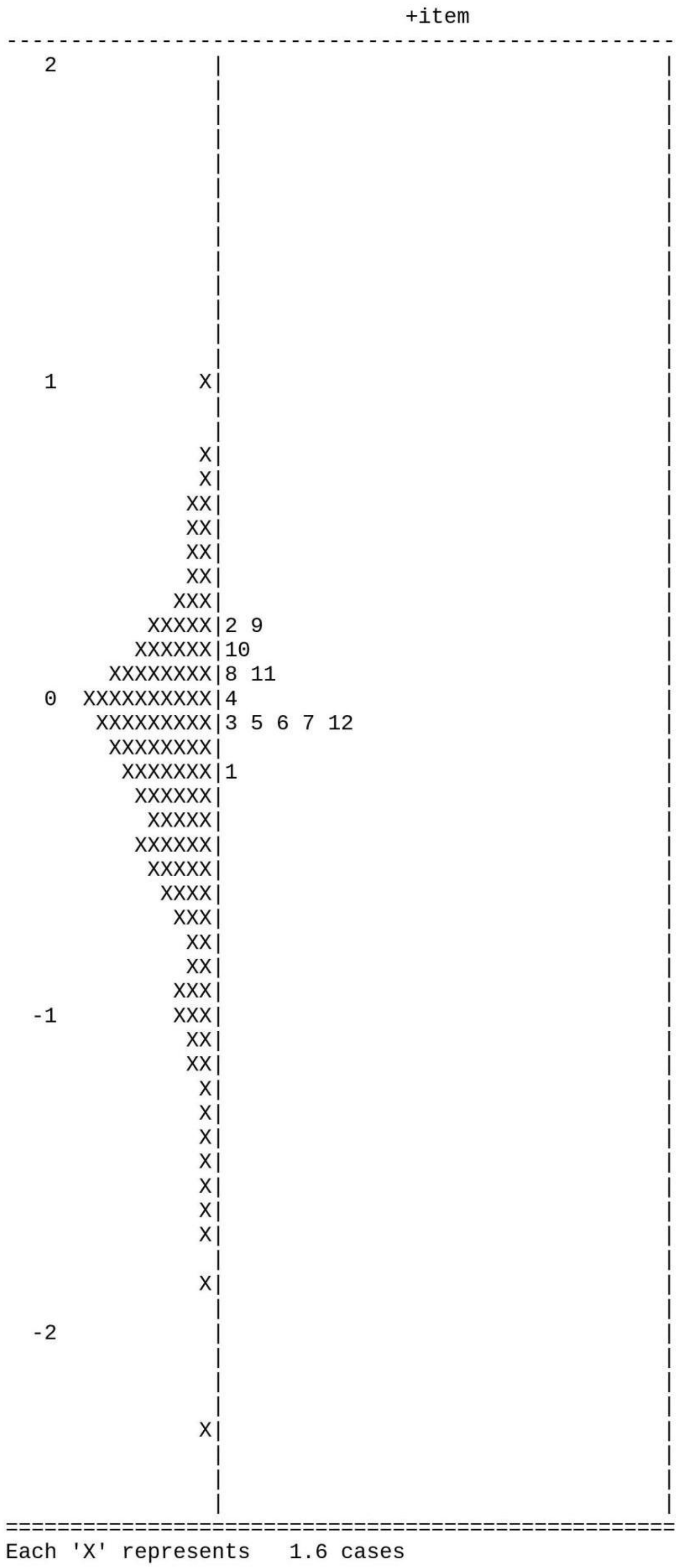
```
                                                               +item
---------------------------------------------------------------------------------------
                                              |                                       |
                                              |                                       |
   7                                          |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
   6                                          |                                       |
                                              |                                       |
                                              |                                       |
   5                                         X|                                       |
                                             X|                                       |
                                             X|                                       |
                                             X|                                       |
   4                                         X|                                       |
                                            XX|                                       |
                                          XXXX|                                       |
   3                                      XXXX|16                                     |
                                     XXXXXXXXX|                                       |
                          XXXXXXXXXXXXXXXXXXXX|                                       |
                            XXXXXXXXXXXXXXXXXX|14                                     |
   2                                XXXXXXXXXX|                                       |
                             XXXXXXXXXXXXXXXXX|15 18                                  |
                       XXXXXXXXXXXXXXXXXXXXXXX|12                                     |
   1                    XXXXXXXXXXXXXXXXXXXXXX|17                                     |
                            XXXXXXXXXXXXXXXXXX|8                                      |
                    XXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|10 11 13                               |
   0                    XXXXXXXXXXXXXXXXXXXXXX|7                                      |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
  -1                      XXXXXXXXXXXXXXXXXXXX|9                                      |
                               XXXXXXXXXXXXXXX|                                       |
                              XXXXXXXXXXXXXXXX|                                       |
                                     XXXXXXXXX|                                       |
  -4                                   XXXXXXX|                                       |
                                          XXXX|                                       |
                                       XXXXXXX|                                       |
  -5                                      XXXX|                                       |
                                         XXXXX|                                       |
                                             X|                                       |
  -6                                         X|                                       |
                                            XX|                                       |
                                          XXXX|                                       |
                                              |                                       |
  -7                                          |                                       |
                                            XX|                                       |
                                              |                                       |
  -8                                          |                                       |
                                              |                                       |
                                             X|                                       |
                                             X|                                       |
  -9                                          |                                       |
                                              |                                       |
=======================================================================================
Each 'X' represents   0.3 cases
```

Fig. 5.3 Item-person map for the IPS test – Chinese sample



```
                                                              +item
--------------------------------------------------------------------------------------
                                              |                                      |
                                              |                                      |
  10                                          |                                      |
                                              |                                      |
                                              |                                      |
   9                                          |                                      |
                                              |                                      |
                                              |                                      |
   8                                          |                                      |
                                              |                                      |
                                              |                                      |
   7                                          |                                      |
                                             X|                                      |
                                              |                                      |
   6                                          |                                      |
                                             X|                                      |
                                            XX|                                      |
   5                                        XX|                                      |
                                             X|                                      |
                                         XXXXX|                                      |
   4                                   XXXXXXX|54                                    |
                                      XXXXXXXX|52                                    |
   3                                 XXXXXXXXX|33                                    |
                               XXXXXXXXXXXXXXX|53                                    |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|50 51                                 |
   2                      XXXXXXXXXXXXXXXXXXXX|31                                    |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|42 49                                 |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|34                                    |
   1  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|32 44 46 48                           |
           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|29 45 47                              |
                        XXXXXXXXXXXXXXXXXXXXXX|11 20 43                              |
   0            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|9 10 15 41                            |
                        XXXXXXXXXXXXXXXXXXXXXX|18 19 26 39 40                        |
                          XXXXXXXXXXXXXXXXXXXX|22 27                                 |
  -1                           XXXXXXXXXXXXXXX|17 21 24 30                           |
                                        XXXXXX|6 23 25                               |
                                          XXXX|13 14 16 28                           |
  -2                                      XXXX|7 8                                   |
                                      XXXXXXXX|3 4 5                                 |
                                 XXXXXXXXXXXXX|12                                    |
  -3                                      XXXX|                                      |
                                          XXXX|1                                     |
                                            XX|2                                     |
  -4                                         X|                                      |
                                           XXX|                                      |
                                         XXXXX|                                      |
  -5                                      XXXX|                                      |
                                           XXX|                                      |
                                            XX|                                      |
  -6                                        XX|                                      |
                                          XXXX|                                      |
                                             X|                                      |
  -7                                          |                                      |
                                              |                                      |
                                             X|                                      |
  -8                                         X|                                      |
                                              |                                      |
                                              |                                      |
  -9                                          |                                      |
                                              |                                      |
======================================================================================
Each 'X' represents   0.4 cases
```

Fig. 5.4 Item-person map for the inductive reasoning test – Chinese sample



```
                               +item
----------------------------------------------------
   2            |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
   1           X|                                  |
                |                                  |
                |                                  |
               X|                                  |
               X|                                  |
              XX|                                  |
              XX|                                  |
              XX|                                  |
              XX|                                  |
             XXX|                                  |
           XXXXX|2 9                               |
          XXXXXX|10                                |
        XXXXXXXX|8 11                              |
   0  XXXXXXXXXX|4                                 |
       XXXXXXXXX|3 5 6 7 12                        |
        XXXXXXXX|                                  |
         XXXXXXX|1                                 |
          XXXXXX|                                  |
           XXXXX|                                  |
          XXXXXX|                                  |
           XXXXX|                                  |
            XXXX|                                  |
             XXX|                                  |
              XX|                                  |
              XX|                                  |
             XXX|                                  |
  -1         XXX|                                  |
              XX|                                  |
              XX|                                  |
               X|                                  |
               X|                                  |
               X|                                  |
               X|                                  |
               X|                                  |
               X|                                  |
               X|                                  |
                |                                  |
               X|                                  |
                |                                  |
  -2            |                                  |
                |                                  |
                |                                  |
               X|                                  |
                |                                  |
                |                                  |
                |                                  |
====================================================
Each 'X' represents   1.6 cases
```

Fig. 5.5 Item-person map for the combinatorial reasoning test – Chinese sample

All the three measured thinking skills have their sub-dimensions, however, at the first stage we run the analyses as if all of the measured reasoning skills could be described as a one-dimensional construct. The basis of these analyses was the test-level performance of the students. For the correlation analyses, the bivariate correlations and the partial correlation between the three measured reasoning skills were calculated (Fig. 5.6).

IPS, IR and CR proved to be strongly correlated constructs (p<.01). The partial correlations were also statistically significant, but they proved to be significantly lower as all bivariate relationships were influenced by the third construct. Combinatorial and inductive reasoning had the strongest relationship. Meanwhile, the correlation between IPS and combinatorial reasoning was slightly higher than the correlation between IPS and inductive reasoning. To sum up, the results indicated that the three measured thinking skills were highly correlated, thus Chinese students' level of IPS was significantly influenced by their ability level of combinatorial and inductive reasoning.

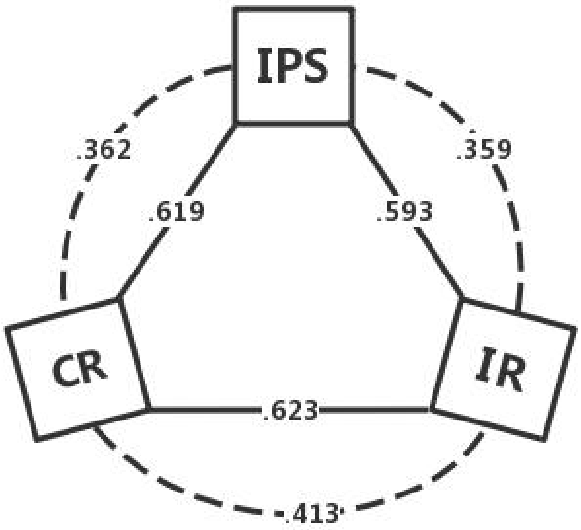


Fig. 5.6 Correlations between the thinking skills – Chinese sample (Solid lines: bivariate correlation; dashed lines: partial correlation)

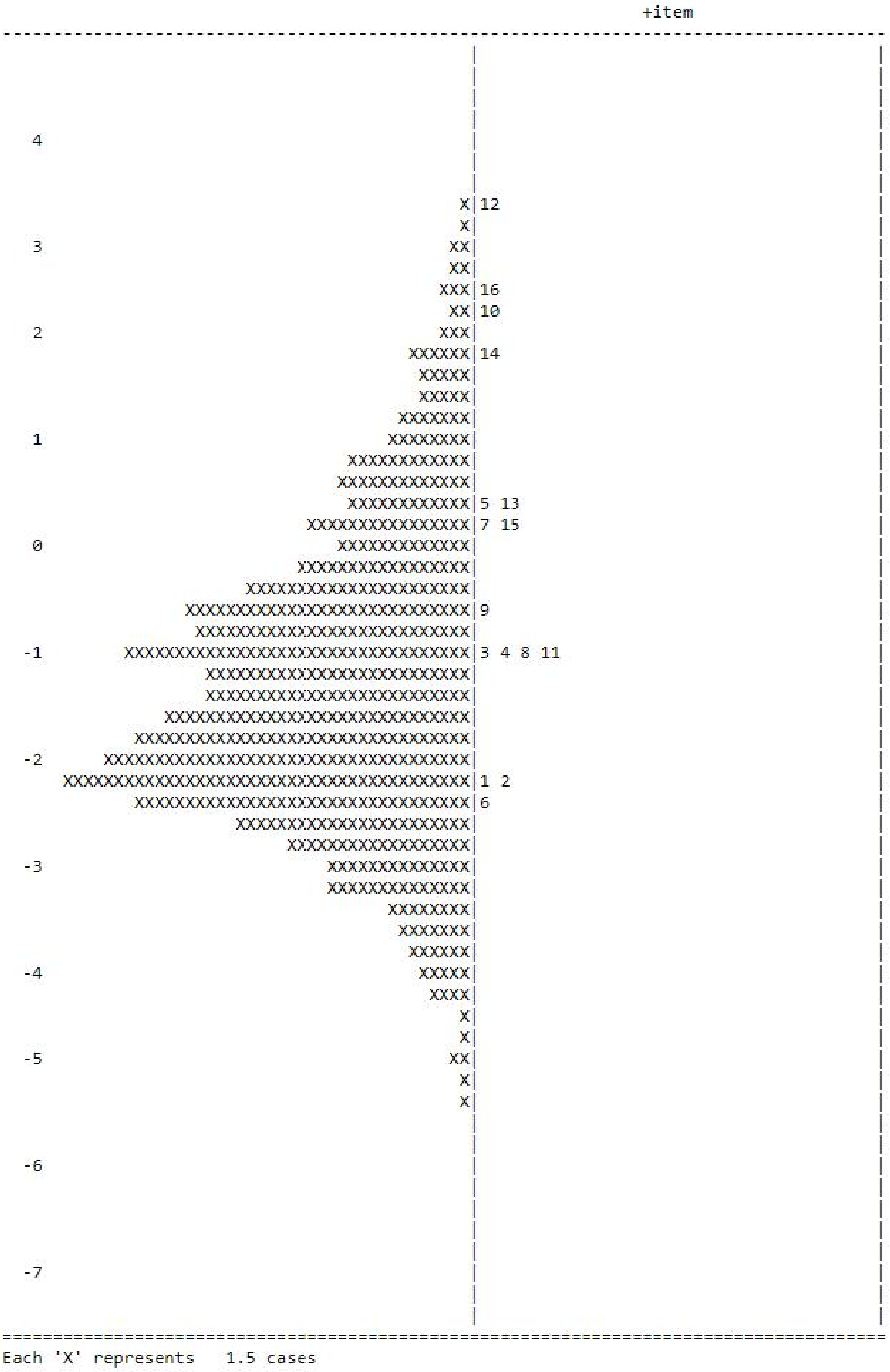
5.3.2 Descriptive statistics of thinking skill tests– Hungary

Table 5.2 indicates the basic statistical information of the three thinking skills tests for the Hungarian students. The majority of tests and sub-tests in Hungary received acceptable reliability coefficients (above .70). Only the subtest knowledge application's Cronbach's alpha was slightly lower than .70. The means for combinatorial reasoning and inductive reasoning tests and their tests were ranging from 47% to 73%, which were ideal for further analysis. The Hungarian students' problem-solving performance was slightly lower than the initial plan, but still good for analysis.

	Number of Items	Cronbach's alpha	Mean (%)	Standard Deviation (%)
Problem Solving	16	.84	33.08	22.21
Knowledge acquisition	8	.79	35.42	29.30
Knowledge application	8	.68	30.73	20.86
Combinatorial Reasoning	7	.84	55.49	22.46
Figural	3	.74	47.10	32.69
Verbal	4	.80	61.78	23.21
Inductive Reasoning	46	.92	67.80	17.85
Figural series	15	.77	73.62	19.44
Figural analogy	15	.76	77.55	19.38
Number analogy	8	.73	57.26	26.36
Number series	8	.85	63.32	23.64

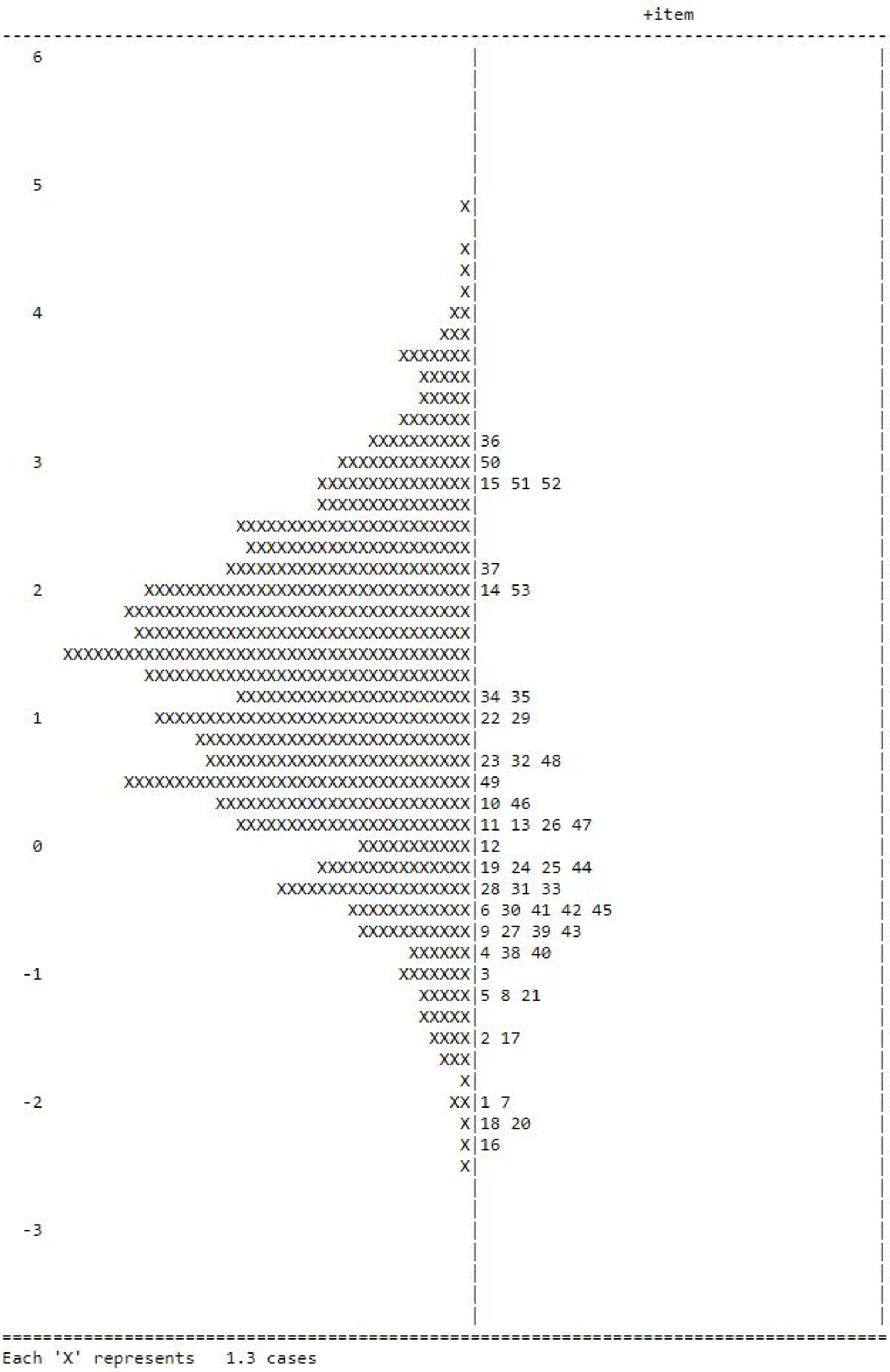
Table 5.2 Descriptive statistics for each assessed thinking skill and their subscales – Hungarian sample

The one-dimensional item-person maps indicated the fitness between students' ability level and the items' difficulty level. According to the item-person maps, the IPS test (Fig. 5.7) was slightly above while inductive reasoning test (Fig. 5.8) was slightly below the students' ability level. As the combinatorial reasoning test contained 7 items, the items could not fully cover the whole ability range of the students, but the 7 items successfully reached the targeted design. Based on their difficulty level, they mainly concentrated around the mean ability level of the modeled population (0) (Fig. 5.9). To sum up, all of the cognitive tests involved it this study fitted to the students' ability level, the tests are appropriate to measure Hungarian 12-year-old students' reasoning skills.



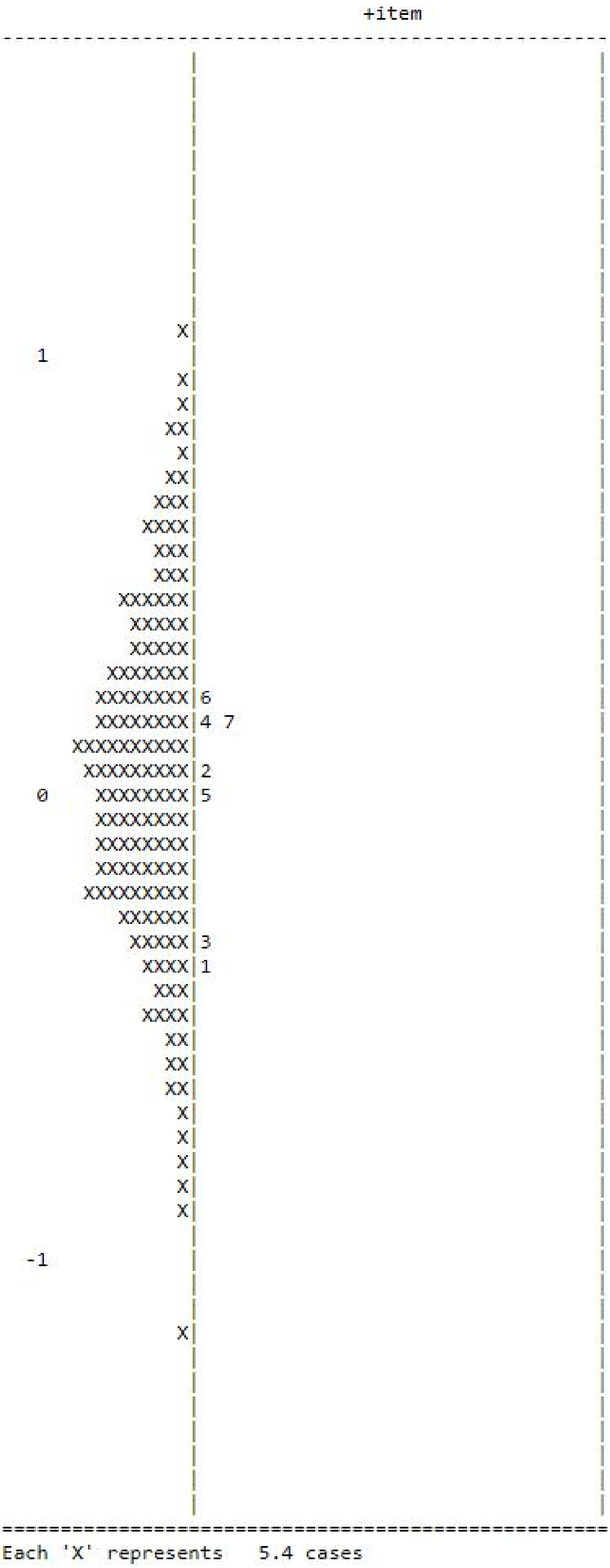
```
                                                               +item
---------------------------------------------------------------------------------------
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
   4                                          |                                       |
                                              |                                       |
                                              |                                       |
                                             X|12                                     |
                                             X|                                       |
   3                                        XX|                                       |
                                            XX|                                       |
                                           XXX|16                                     |
                                            XX|10                                     |
   2                                       XXX|                                       |
                                        XXXXXX|14                                     |
                                         XXXXX|                                       |
                                         XXXXX|                                       |
                                       XXXXXXX|                                       |
   1                                  XXXXXXXX|                                       |
                                  XXXXXXXXXXXX|                                       |
                                 XXXXXXXXXXXXX|                                       |
                                  XXXXXXXXXXXX|5 13                                   |
                              XXXXXXXXXXXXXXXX|7 15                                   |
   0                             XXXXXXXXXXXXX|                                       |
                             XXXXXXXXXXXXXXXXX|                                       |
                        XXXXXXXXXXXXXXXXXXXXXX|                                       |
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXX|9                                      |
                   XXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
  -1        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|3 4 8 11                               |
                    XXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                    XXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
  -2      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|1 2                                    |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|6                                      |
                       XXXXXXXXXXXXXXXXXXXXXXX|                                       |
                            XXXXXXXXXXXXXXXXXX|                                       |
  -3                            XXXXXXXXXXXXXX|                                       |
                                XXXXXXXXXXXXXX|                                       |
                                      XXXXXXXX|                                       |
                                       XXXXXXX|                                       |
                                        XXXXXX|                                       |
  -4                                     XXXXX|                                       |
                                          XXXX|                                       |
                                             X|                                       |
                                             X|                                       |
  -5                                        XX|                                       |
                                             X|                                       |
                                             X|                                       |
                                              |                                       |
                                              |                                       |
  -6                                          |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
  -7                                          |                                       |
                                              |                                       |
                                              |                                       |
=======================================================================================
Each 'X' represents   1.5 cases
```

Fig. 5.7 Item-person map for the IPS test – Hungarian sample



```
                                                               +item
----------------------------------------------------------------------------------------
   6                                          |                                        |
                                              |                                        |
                                              |                                        |
                                              |                                        |
                                              |                                        |
                                              |                                        |
   5                                          |                                        |
                                             X|                                        |
                                              |                                        |
                                             X|                                        |
                                             X|                                        |
                                             X|                                        |
   4                                        XX|                                        |
                                           XXX|                                        |
                                       XXXXXXX|                                        |
                                         XXXXX|                                        |
                                         XXXXX|                                        |
                                       XXXXXXX|                                        |
                                    XXXXXXXXXX|36                                      |
   3                             XXXXXXXXXXXXX|50                                      |
                               XXXXXXXXXXXXXXX|15 51 52                                |
                               XXXXXXXXXXXXXXX|                                        |
                       XXXXXXXXXXXXXXXXXXXXXXX|                                        |
                        XXXXXXXXXXXXXXXXXXXXXX|                                        |
                      XXXXXXXXXXXXXXXXXXXXXXXX|37                                      |
   2          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|14 53                                   |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
                       XXXXXXXXXXXXXXXXXXXXXXX|34 35                                   |
   1           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|22 29                                   |
                   XXXXXXXXXXXXXXXXXXXXXXXXXXX|                                        |
                    XXXXXXXXXXXXXXXXXXXXXXXXXX|23 32 48                                |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|49                                      |
                     XXXXXXXXXXXXXXXXXXXXXXXXX|10 46                                   |
                       XXXXXXXXXXXXXXXXXXXXXXX|11 13 26 47                             |
   0                               XXXXXXXXXXX|12                                      |
                               XXXXXXXXXXXXXXX|19 24 25 44                             |
                           XXXXXXXXXXXXXXXXXXX|28 31 33                                |
                                  XXXXXXXXXXXX|6 30 41 42 45                           |
                                   XXXXXXXXXXX|9 27 39 43                              |
                                        XXXXXX|4 38 40                                 |
  -1                                   XXXXXXX|3                                       |
                                         XXXXX|5 8 21                                  |
                                         XXXXX|                                        |
                                          XXXX|2 17                                    |
                                           XXX|                                        |
                                             X|                                        |
  -2                                        XX|1 7                                     |
                                             X|18 20                                   |
                                             X|16                                      |
                                             X|                                        |
                                              |                                        |
                                              |                                        |
  -3                                          |                                        |
                                              |                                        |
                                              |                                        |
                                              |                                        |
                                              |                                        |
========================================================================================
Each 'X' represents   1.3 cases
```

Fig. 5.8 Item-person map for the inductive reasoning test – Hungarian sample



```
                               +item
----------------------------------------------------
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
               X|                                  |
   1            |                                  |
               X|                                  |
               X|                                  |
              XX|                                  |
               X|                                  |
              XX|                                  |
             XXX|                                  |
            XXXX|                                  |
             XXX|                                  |
             XXX|                                  |
          XXXXXX|                                  |
           XXXXX|                                  |
           XXXXX|                                  |
         XXXXXXX|                                  |
        XXXXXXXX|6                                 |
        XXXXXXXX|4 7                               |
      XXXXXXXXXX|                                  |
       XXXXXXXXX|2                                 |
   0    XXXXXXXX|5                                 |
        XXXXXXXX|                                  |
        XXXXXXXX|                                  |
        XXXXXXXX|                                  |
       XXXXXXXXX|                                  |
          XXXXXX|                                  |
           XXXXX|3                                 |
            XXXX|1                                 |
             XXX|                                  |
            XXXX|                                  |
              XX|                                  |
              XX|                                  |
              XX|                                  |
               X|                                  |
               X|                                  |
               X|                                  |
               X|                                  |
               X|                                  |
                |                                  |
  -1            |                                  |
                |                                  |
                |                                  |
               X|                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
                |                                  |
====================================================
Each 'X' represents   5.4 cases
```

Fig. 5.9 Item-person map for the combinatorial reasoning test – Hungarian sample

The bivariate correlations (Fig. 5.10) between combinatorial and inductive reasoning were high (r_{CR_IR}=.633, $p<.01$). The relationship between IPS and combinatorial reasoning and IPS and inductive reasoning proved to be almost the same. Thus, students' problem-solving achievement was influenced by these two reasoning skills almost at the same level. The partial correlation analysis reached a similar conclusion. The strongest correlation was found between combinatorial and inductive reasoning, while both of these two reasoning skills correlated with problem solving with similar power. These analyses confirmed the relationship between these three reasoning skills in the Hungarian context.

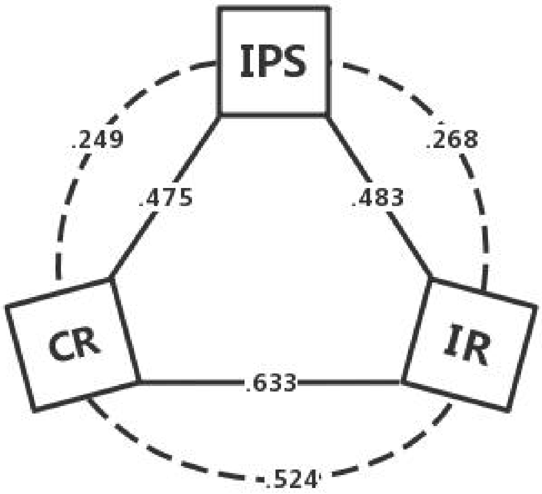


Fig. 5.10 Correlations between the thinking skills – Hungarian sample (Solid lines: bivariate correlation; dashed lines: partial correlation)

5.4 Cross-cultural comparison of problem solving, inductive reasoning and combinatorial reasoning

5.4.1 Dimensionalities of IPS, inductive reasoning and combinatorial reasoning in Chinese and Hungarian contexts

Based on the measurement instrument design, all three thinking skills contain several subscales. Multi-dimensional models were built (1) to answer research question five regarding IPS's dimensionality, and (2) to demonstrate the dimensionality of combinatorial and inductive reasoning as the preliminary work for the SEM modeling.

In the theoretical model of the IPS test, the two subscales were knowledge acquisition and knowledge application. We tested (Table 5.3) both the one-dimensional model with both processes combined under one general factor and the two-dimensional model, whether the processes knowledge acquisition and knowledge application could be empirically distinguished or not. Both models showed a good model fit for both the Chinese and Hungarian samples. In order to test which model fitted the data better, we carried out a special χ^2-difference test in Mplus, which showed that the 2-dimensional model fitted significantly better than the 1-dimensional model (CN: Chi-square=12.98, p<.01, HU: Chi-square = 78.57, p<.001) in both cultures. Subsequently, the 2-dimensional model was applied in the analyses for both countries.

To sum up, IPS should be described as a two-dimensional construct; the processes of knowledge acquisition and knowledge application could be empirically distinguished. (RQ2) Thus, *hypothesis H2 was supported*.

Nation	Model	Chi-square	df	p	CFI	TLI	RMSEA
CN	1-dimensional	103.57	65	<.01	.99	.99	.06
	2-dimensional	90.59	64	<.05	.98	.98	.05
HU	1-dimensional	200.31	65	<.01	.98	.98	.05
	2-dimensional	121.74	64	<.01	.99	.99	.03

Table 5.3 Goodness of fit indices for testing the dimensionality of IPS

The two theoretical subscales of combinatorial reasoning were combining only figural or verbal elements. For the Chinese sample, the 1-dimensional model showed a bad model fit (Table 5.4), while the 2-dimensional model fit can be considered acceptable. The difference testing indicated a significant difference

(Chi-square=152.53, p<.01) between these two models. As for the Hungarian sample, the one-dimensional model had acceptable model fit, but still significantly lower than the two-dimensional model (Chi-square=34.30, p<.01). Thus, combinatorial reasoning is much more appropriately considered as a two-dimensional construct, distinguishing the two processes that combining figural and verbal elements, in SEM modeling for both the Chinese and Hungarian samples.

Nation	Model	Chi-square	df	p	CFI	TLI	RMSEA
CN	1-dimensional	232.82	33	<.01	.76	.73	.17
	2-dimensional	80.29	32	<.01	.94	.93	.09
HU	1-dimensional	52.54	12	<.01	.91	.88	.06
	2-dimensional	18.24	11	<.01	.98	.98	.03

Table 5.4 Goodness of fit indices for testing the dimensionality of combinatorial reasoning

The IR test development contains four subscales: figural series, figural analogy, number analogy, and number series. Therefore, besides the one and four-dimensional models, we also tested two different two-dimensional models, comprising figural-numerical and series-analogies (Table 5.5). For the Chinese sample, the one-dimensional model showed an unsatisfactory model fit. Both of the two-dimensional models' model fits were acceptable, but still significantly worse than that of the four-dimensional model (p<.01). Similarly, the 4-dimensional model also had the highest model fits (p<.01) in the Hungarian sample. Therefore, inductive reasoning was built as a four-dimensional construct in the following SEM models, empirically distinguishing the four processes: figural analogies, figural series, number analogies and number series.

Nation	Model	Chi-square	df	p	CFI	TLI	RMSEA
CN	1-dimensional	204.67	64	<.01	.95	.98	.11
	2-dimensional (1)*	178.12	65	<.01	.96	.99	.09
	2-dimensional (2)*	167.09	66	<.01	.97	.99	.09
	4-dimensional	115.31	75	<.01	.99	1.00	.05
HU	1-dimensional	2072.02	285	<.01	.81	.90	.09
	2-dimensional (1)*	1696.38	286	<.01	.85	.92	.08
	2-dimensional (2)*	1888.79	284	<.01	.82	.91	.08
	4-dimensional	1441.70	287	<.01	.87	.94	.07

Note. *2-dimensional (1): figure-number; 2-dimensional (2): series-analogy

Table 5.5 Goodness of fit indices for testing the dimensionality of inductive reasoning

5.4.2 Measurement invariance across nationality and gender

Before moving on to the next step of the analysis, an essential question needed to be answered. Can IPS be measured invariant across gender and nationality both in European and Asian contexts? A measurement invariance analysis was conducted across gender and nationality,. The results are shown in Table 5.6.

The measurement invariance analysis started with identifying the baseline model which fits within the overall sample and in each subgroup (Byrne, & Stewart, 2006). Configural invariance was firstly tested by estimating the parameters of the baseline model in a multi-group model. Then strong factorial invariance was tested by constraining the factor loadings and thresholds to be equal across groups (nationality and gender). (Byrne, & Stewart, 2006) χ^2-difference test was used to compare the strong factorial invariance model and configural invariance model. Measurement invariance can be proved if no significant difference can be detected between these two models.

Group	Invariance model	χ^2	df	p	CFI	TLI	RMSEA	$\Delta\chi^2$	p
Nationality	Config. invariance	263.54	152	<.001	.99	.99	.04	122.82	<.05
	Strong factorial inv.	378.67	162	<.001	.98	.98	.05		
Gender (HU)	Config. invariance	219.46	152	<.001	.99	.99	.03	11.83	.29
	Strong factorial inv.	228.61	162	<.001	.99	.99	.03		
Gender (CN)	Config. invariance	182.95	152	<.001	.99	.99	.05	15.90	.10
	Strong factorial inv.	198.25	162	<.001	.99	.99	.05		

Table 5.6 Goodness of fit indices for testing invariance across nationality and gender of problem solving

The results based on multi-group (CN-HU) confirmatory factor analyses showed that the configural invariance model was significantly different with the strong factorial invariance model, which indicated measurement non-invariance of IPS across nationality. Thus, we checked measurement invariance by gender in the two countries. The multi-group (Boy-Girl) confirmatory factor analyses showed that the configural invariance model was equal with the strong factorial invariance model in both cultures. Thus, IPS proved to be measurement invariant across gender within the same culture. *(RQ3)* Therefore, *hypothesis H3 was partly denied.*

As the tasks required a minimal amount of reading, we assumed that the reason for non-invariance could be found in students' different cognitive styles during the problem-solving progress. This is supported by the research result that cognitive styles can be impacted by cultural background (Nisbett, & Miyamoto, 2005), and there is a significant cultural difference between China and Hungary. SEM analysis was used to testify the assumption and detect the influences problem solving received from inductive and combinatorial reasoning.

5.4.3 Students developmental level of problem solving, inductive reasoning and combinatorial reasoning in China and in Hungary

As introduced, the items delivered to Chinese and Hungarian students were not completely the same. In this section, item response theory was employed to analyze Chinese and Hungarian students' performance on the same scale. Independent t-test was applied to test the performance difference between Hungarian and Chinese students. The results are shown in Table 5.7.

Test	Nationality	Mean	SD	t	p
Problem Solving	CN	-.81	2.07	.05	.>.05
	HU	-.81	1.48		
Inductive Reasoning	CN	.79	2.20	.48	>.05
	HU	.81	1.23		
Combinatorial Reasoning	CN	-.32	.44	6.11	<.05
	HU	-.03	.62		

Table 5.7 Comparison of the ability levels of the assessed thinking skills

Chinese and Hungarian students showed statistically equal performance in the IPS and inductive reasoning tests. The results indicated that our Chinese and Hungarian sample had the same developmental level regarding interactive problem-solving and inductive reasoning. (RQ4) *Hypothesis H4 was partly denied.*

Moreover, in these two tests, the standard deviations for the Chinese sample were much higher than for the Hungarian sample, which showed that a larger internal difference existed among the Chinese students. Statistically significant difference can only be detected in the combinatorial reasoning test. Hungarian students showed a better performance than the Chinese students.

In the IPS test, Hungarian males (M=-.66, SD=1.58) showed a statistically equal performance with Chinese males (M=-.75, SD=1.40) and Chinese females (M=-.87, SD=2.02) (p>.05). Therefore, Hungarian males, Chinese males, and Chinese females had the same developmental level of problem solving at the age of 12. Only Hungarian females (M=-.93, SD=1.40) showed a significantly lower performance than the Hungarian males (t=-2.68, p<.01). (RQ5) Therefore, *Hypothesis H5 was partly supported.*

5.4.4 Similarities and differences in the relationships between problem solving, inductive reasoning and combinatorial reasoning in China and Hungary

Two SEM models were built to examine the relationships among problem solving, combinatorial reasoning and inductive reasoning in the Chinese and Hungarian contexts (Fig. 5.11 & 5.12). Based on the dimensionality testing, all three thinking skills were construed as latent variables composed of their sub-dimensions. The Chinese model will be discussed first, followed by the Hungarian model, then the similarities and differences between two countries' students' cognitive style in the problem-solving process will be discussed.

The model fits for the Chinese model were acceptable (Chi-Square=42.34, p<.001, CFI=.97, TLI=.96, RMSEA=.09, SRMR=.04). The model (Fig. 5.11) indicates that both knowledge acquisition and knowledge application were strong contributors to problem solving (β= .866-.883). The contribution from combinatorial processes applied by figural items (β=.689) to combinatorial reasoning as a latent construct was significant but weaker than that of the processes using verbal elements (β=.857). As for inductive reasoning, it was strongly supported by the processes used by building figure analogies (β=.930) and number analogies (β=.881), while the contributions from the other two dimensions were weaker but still high (β=.680-.794).

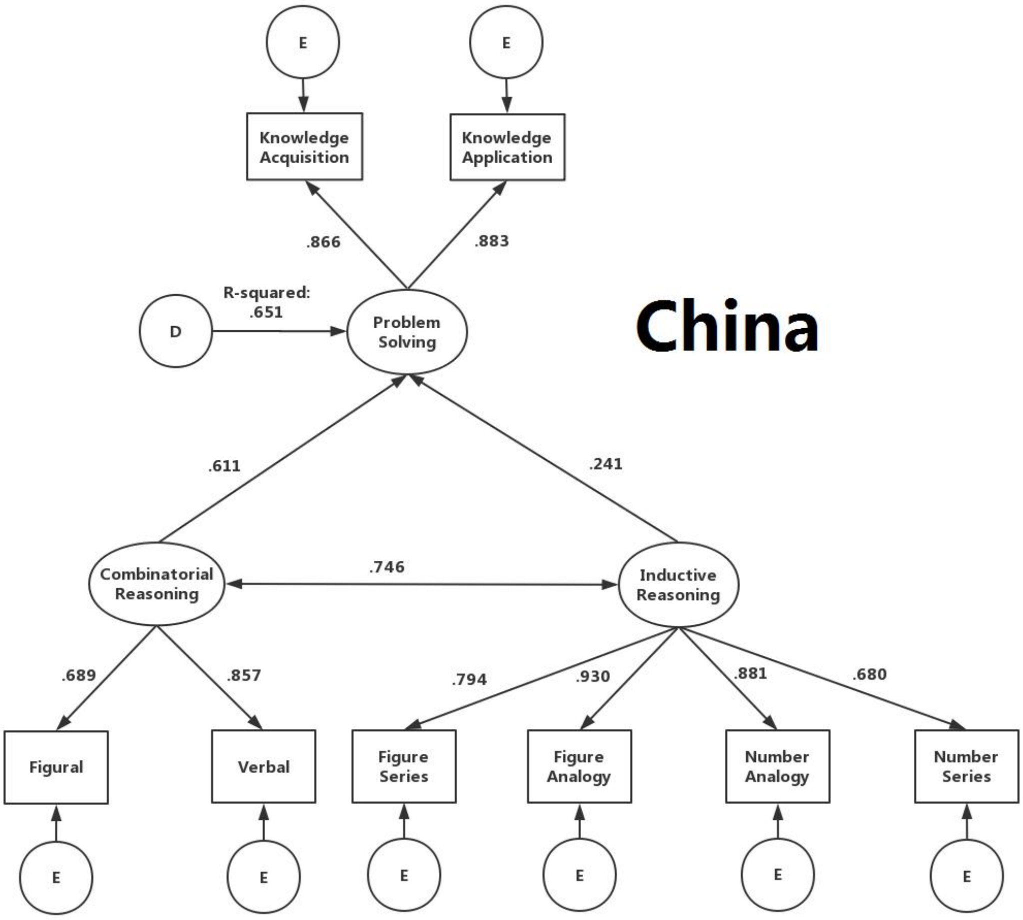

E
E
Knowledge Acquisition
Knowledge Application
.866
.883
R-squared: .651
D
Problem Solving
China
.611
.241
Combinatorial Reasoning
.746
Inductive Reasoning
.689
.857
.794
.930
.881
.680
Figural
Verbal
Figure Series
Figure Analogy
Number Analogy
Number Series
E
E
E
E
E
E

Fig. 5.11 A structural model presenting the relationships among problem solving, combinatorial reasoning and inductive reasoning – Chinese sample

Both combinatorial reasoning and inductive reasoning showed a significant predicting effect for problem solving ($p<.05$), confirming that these two reasoning skills play an important role in the problem-solving process. Moreover, the predicting effect of combinatorial reasoning ($\beta=.611$) was stronger than that of inductive reasoning ($\beta=.241$), indicating that the Chinese students solved problems by relying much more on their combinatorial reasoning skill. At the same time, combinatorial and inductive reasoning were highly correlated ($r=.746$, $p<.01$), proving that these two skills were impacting on each other in students' cognitive development.

The SEM model based on the Hungarian data achieved good model fits (Chi-square=25.83, df=17, CFI=1.00, TLI=1.00, RMSEA=.03, SRMR=.02). The model (Fig. 5.12) had the same structure as the Chinese model. Problem solving was supported by its two sub-dimensions, knowledge acquisition and application (β= .746-.764). Combinatorial and inductive reasoning also consisted of all their sub-dimensions. Similarly to the Chinese model, combinatorial reasoning received strong contributions from verbal (β= .844) and figural (β= .752) sub-dimensions, and inductive reasoning received strong contributions from its sub-dimensions (β=.724 - .820). Problem solving got significant predicting effects from combinatorial reasoning (β=.339) and inductive reasoning (β=.376), while inductive reasoning and combinatorial reasoning had a strong correlation (r=.784).

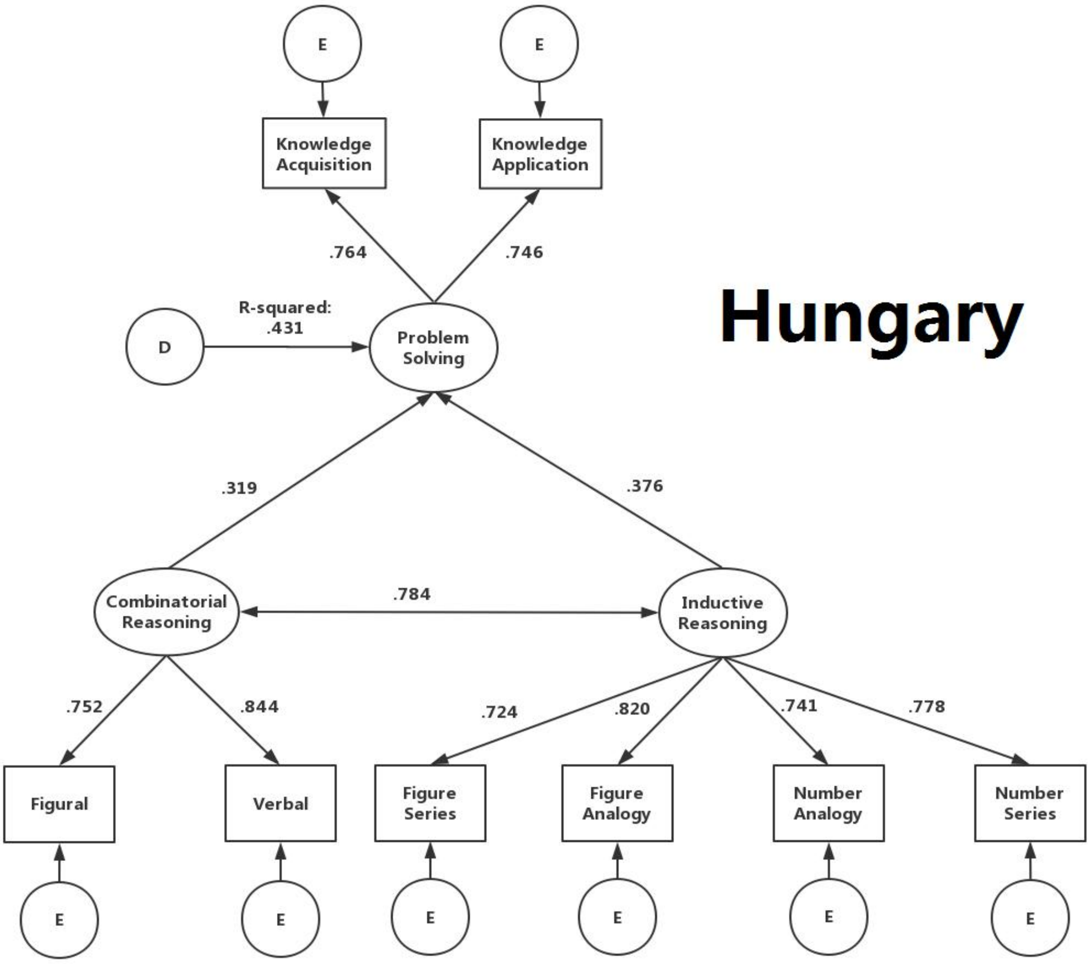

E
E
Knowledge Acquisition
Knowledge Application
.764
.746
Hungary
R-squared: .431
D
Problem Solving
.319
.376
Combinatorial Reasoning
.784
Inductive Reasoning
.752
.844
.724
.820
.741
.778
Figural
Verbal
Figure Series
Figure Analogy
Number Analogy
Number Series
E
E
E
E
E
E

Fig. 5.12 A structural model presenting the relationships among problem solving, combinatorial reasoning and inductive reasoning – Hungarian sample

These two models contained some common features. The developmental level of problem solving was significantly predicted by the developmental level of inductive and combinatorial reasoning in both models. The results confirmed that inductive and combinatorial reasoning are important component skills of problem solving, independently of the cultural context. (RQ6) Hence, *hypothesis H6 was supported.*

Inductive and combinatorial reasoning proved to be highly correlated in both models. However, there were detectable differences in the cognitive styles of Chinese and Hungarian students. Combinatorial reasoning (β=.319) and inductive reasoning (β=.376) had basically equal influence on the problem-solving achievement of the Hungarian students; while combinatorial reasoning (β=.611) played a much more important role than inductive reasoning (β=.241) in the Chinese sample. Therefore, the detected IPS assessment measurement non-invariance across nationality could be explained by students' different cognitive styles during the problem-solving progress.

5.5 Cross-cultural comparison of the relationships between problem solving performance and students' developmental level of working memory, test-taking motivation and learning strategies

5.5.1 The influential level of working memory

The working memory tests in China and Hungary contained the same items, thus the direct comparison was available regarding this factor. In the working memory test, the Chinese students (M=57.74%, SD=21.16%) showed a statistically equal performance with the Hungarian students (M=57.17%, SD=16.64%). There was no statistically significant difference between these two groups (t=.36, p>.05), which showed that the Chinese and Hungarian students had the same level of working memory at 12 years of age. The correlation analysis indicated that working memory had significant correlations with all the three thinking skills in both nations (Table 5.8).

Nation		IPS	IR	CR
CN	Working	.530	.609	.557
HU	Memory	.390	.447	.318

Note. All correlations are significant at the 0.01 level.

Table 5.8 Correlations between working memory and thinking skills

Working memory showed the strongest correlation with inductive reasoning in both China and Hungary (Table 5.8). Students' IPS and combinatorial reasoning achievement also highly correlated with their working memory. Fisher's transformation analysis indicated that the Chinese students showed significantly stronger connections between working memory and thinking skills than the Hungarian students (IPS: $z=2.19 <.05$, IR: $z=2.78$, $p<.01$, and CR: $z=3.62$, $p<.01$). But still, the result proved that working memory played a key role during students' cognitive process in both Chinese and Hungarian contexts.

5.5.2 The influential level of students' test-taking motivation in interactive problem-solving environment

Due to time limitation, the test-taking motivation part was reduced to 9 items in this assessment (item 1-9, see Table 4.1). Assuming opinion 7 "strongly agree" for every single question as 100%, for the positive questions, the average response for the Chinese students was 69.30% (SD=17.76%), which was significantly ($p<.01$) higher than that of the Hungarian students (M=55.88%, SD=13.39%). For the negative questions, the average response for the Chinese students was 47.00% (SD=21.78%), which was significantly ($<.01$) lower than that of the Hungarian students (M=54.63%, SD=25.54%). The results demonstrated that Chinese students had a generally high test-taking motivation in the IPS assessment,while the Hungarian students' motivation was lower than that of the Chinese students.

Significant correlations were detected between Chinese students' performance in IPS assessment and their responses on the motivation questions, for both positive (r=.505, p<.05) and negative questions (r= -.545, p<.05). However, there was no significant correlation between Hungarian students' test-taking motivation and their IPS performance. The analysis proved that Chinese students' level of test-taking motivation influenced their performance in the interactive problem-solving environment. Hungarian students' IPS performance was not influenced by their test-taking motivation. That could be caused by Hungarian students' generally low level test-taking motivation during the assessment.

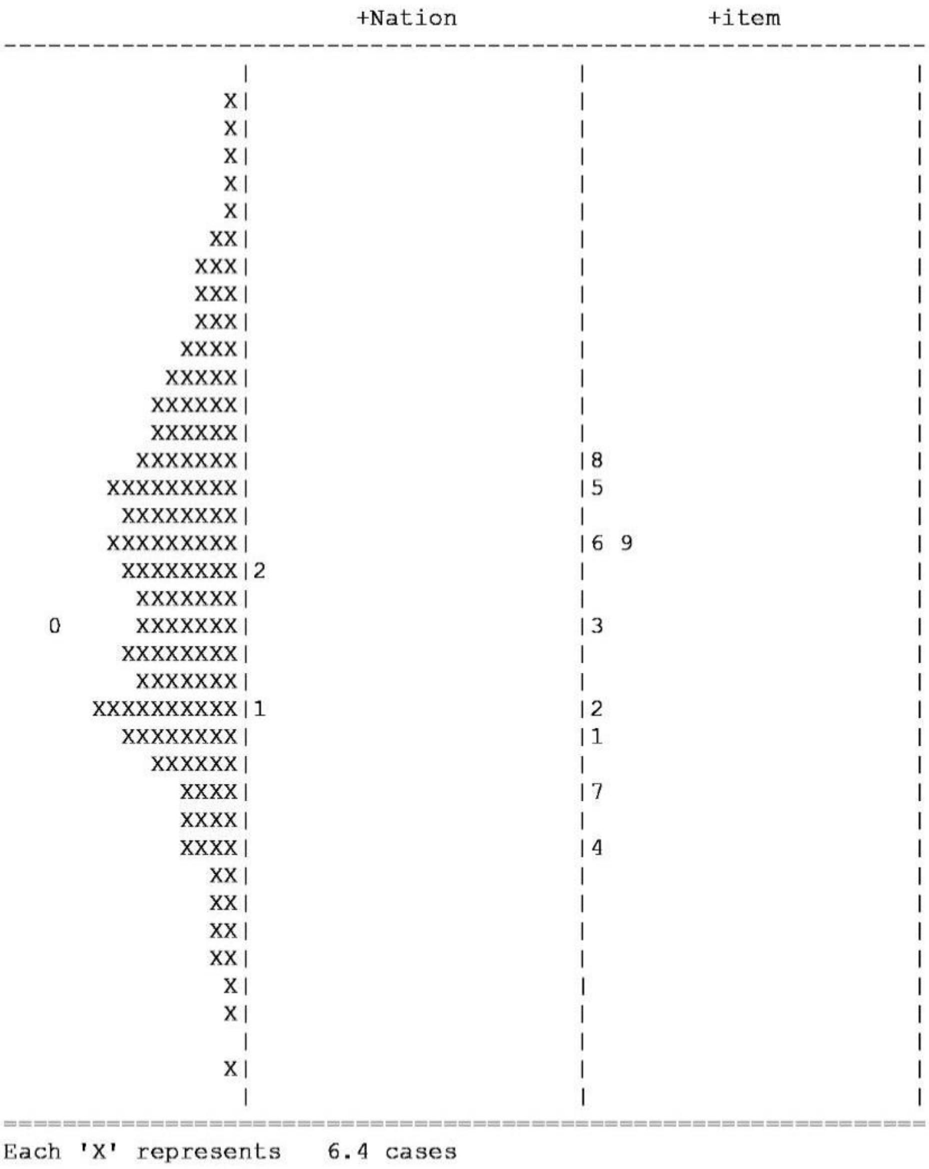
The Rasch model was used to provide further information about students' test-taking motivation. First, we tested which model fits the data better: rating scale or partial credit model (Table 5.9). Partial credit model showed better model fits (p<.01). Therefore, the Partial credit model was selected for doing the analysis.

Model Type	*Final deviance*	*Total number of estimated parameters*
Rating Scale Model	32613.215	16
Partial Credit Model	32320.453	56

Table 5.9 Model fits for rating scale model and partial credit model

Students held the most affirmative attitude to question 4 (Fig. 5.13), followed by questions 1 and 7 (see Table 4.1 for the content of items). This means that students believed this assessment is interesting, either before, during or after this assessment. Similarly to the results achieved in the pilot study, students had strong confidence to do a good job at the beginning (based on question 2), but their confidence reduced during and after the assessment (based on questions 5 and 8).

Chinese students (represented by 1) were more likely to choose the affirmative response for the questions, while Hungarian students' (represented by 2) attitude towards the questions was less positive,which confirmed that in this assessment, Chinese students had higher motivation.



```
                        +Nation                 +item
---------------------------------------------------------------
                |                      |                      |
               X|                      |                      |
               X|                      |                      |
               X|                      |                      |
               X|                      |                      |
               X|                      |                      |
              XX|                      |                      |
             XXX|                      |                      |
             XXX|                      |                      |
             XXX|                      |                      |
            XXXX|                      |                      |
           XXXXX|                      |                      |
          XXXXXX|                      |                      |
          XXXXXX|                      |                      |
         XXXXXXX|                      |8                     |
       XXXXXXXXX|                      |5                     |
        XXXXXXXX|                      |                      |
       XXXXXXXXX|                      |6 9                   |
        XXXXXXXX|2                     |                      |
         XXXXXXX|                      |                      |
   0     XXXXXXX|                      |3                     |
        XXXXXXXX|                      |                      |
         XXXXXXX|                      |                      |
      XXXXXXXXXX|1                     |2                     |
        XXXXXXXX|                      |1                     |
          XXXXXX|                      |                      |
            XXXX|                      |7                     |
            XXXX|                      |                      |
            XXXX|                      |4                     |
              XX|                      |                      |
              XX|                      |                      |
              XX|                      |                      |
              XX|                      |                      |
               X|                      |                      |
               X|                      |                      |
                |                      |                      |
               X|                      |                      |
                |                      |                      |
===============================================================
Each 'X' represents   6.4 cases
```

Fig. 5.13 Item-person map for the motivation questions

We have divided the whole sample into four groups by gender and nationality: 1. Chinese Males, 2. Chinese Females, 3. Hungarian Males, and 4. Hungarian Females. Fig. 5.14 illustrates the item/person map which showed the differences between these four groups of students in their level of test-taking motivation. Results indicated that Chinese males had the highest test-taking motivation, followed by Chinese females. Hungarian males' motivation was higher than that of the Hungarian females, but still lower than the test-taking motivation of the Chinese females.

```
                     +Nation&Gender            +item
---------------------------------------------------------------
                |                      |                      |
               X|                      |                      |
              XX|                      |                      |
               X|                      |                      |
              XX|                      |                      |
             XXX|                      |                      |
             XXX|                      |                      |
             XXX|                      |                      |
            XXXX|                      |                      |
          XXXXXX|                      |                      |
          XXXXXX|                      |                      |
          XXXXXX|                      |                      |
         XXXXXXX|                      |5 8                   |
       XXXXXXXXX|                      |                      |
        XXXXXXXX|4                     |                      |
       XXXXXXXXX|                      |9                     |
        XXXXXXXX|                      |6                     |
         XXXXXXX|3                     |3                     |
   0    XXXXXXXX|                      |                      |
        XXXXXXXX|                      |                      |
         XXXXXXX|2                     |                      |
      XXXXXXXXXX|                      |2                     |
         XXXXXXX|1                     |                      |
          XXXXXX|                      |1                     |
             XXX|                      |7                     |
             XXX|                      |                      |
           XXXXX|                      |4                     |
              XX|                      |                      |
              XX|                      |                      |
              XX|                      |                      |
              XX|                      |                      |
               X|                      |                      |
               X|                      |                      |
                |                      |                      |
               X|                      |                      |
                |                      |                      |
===============================================================
Each 'X' represents   6.5 cases
```

Fig. 5.14 Item-person map with nation & gender differences

The results indicated that the Hungarian students were not highly motivated during the test-taking process, which might have influenced their achievement. As mentioned in section 4.5, in the pilot test, there were several Chinese students who did not provide any answer in the given time-frame, during the 180 seconds, and their behaviour caused a system bug, and they could not continue the assessment after running out of time. This system bug had never occurred in Hungary. Here the findings suggested that this system bug might be caused by Chinese students' high motivational level. They were willing to try every single possible solution before they actually gave their answer, and they used all the time they had to build the solution;

by contrast, Hungarian students' test-taking motivation was comparatively low, so they did not encounter this problem. They answered all of the questions much earlier, they did not play with the scenarios too much, until running out of time.

5.5.3 The influential level of parents' educational level

The background questionnaire contains information regarding students' parental educational level. However, China and Hungary have different educational systems, thus their parents' educational level could not be compared directly. We divided the educational levels of the parents into three groups: 1) primary education; 2) secondary education, and (3) higher education. Because of the prior background matching during the sampling process, the parental educational level of the students matched generally on sample level. Only the percentage of "primary education" level for the Hungarian sample was obviously higher than that of the Chinese sample. Considering that primary education in Hungary (8 years) lasts two years longer than the one in China (6 years), the result was still in the reasonable scope (Table 5.10).

	Father's education		Mother's education	
	CN	HU	CN	HU
Primary education	3.0%	11.1%	7.3%	9.1%
Secondary education	64.2%	56.9%	70.1%	52.5%
Higher education	32.8%	32.0%	22.6%	38.3%

Table 5.10 Distribution of participants' parental educational levels

In the pilot study, the parental educational level of Chinese students did not show any influence on students' level of thinking skills, but significant influences were identified in the present study. ANOVA analysis indicated significant differences between students having different parental educational levels in all of the measured domains (see Table 5.11 and 5.12), for both Chinese and Hungarian students.

		IPS	IR	CR
Mother's Education	F(5,182)	3.61	3.36	4.08
	p	<.05	<.05	<.05
Father's Education	F(5,182)	3.06	3.251	5.10
	p	<.05	<.05	<.01

Table 5.11 Performance differences for students with different parental educational levels – Chinese sample

		IPS	IR	CR
Mother's Education	F(6,829)	5.31	5.88	7.88
	p	<.01	<.01	<.01
Father's Education	F(6,829)	5.38	8.39	6.84
	p	<.01	<.01	<.01

Table 5.12 Performance differences for students with different parental educational levels – Chinese sample

Correlation analysis showed that students with higher educated parents had comparatively high performances in all the three thinking skill tests (see Table 5.13). The mother's educational level showed a slightly higher influence on the level of thinking skills than father's in both cultures. The Fisher's transformation analysis also indicated that Chinese students' mothers' education level had significantly stronger influences on all the thinking skills than that of the Hungarian students (IPS: z=2.29, p<.01, IR: z=2.42, p<.01, CR: z=2.51, p<.01). Chinese students' father's education level also had stronger influences on inductive reasoning than that of the Hungarian students (IPS: not significant, IR: z=2.46, p<.05, CR: not significant). A previous study pointed out that in China people with high educational level have remarkably better possibility (than the world average level) to reach a higher income and social position (see Wang, & Wu, 2018). This feature led to Chinese students with higher educational level parents having advantages in many aspects (economic, social, etc.). Thus, parents' educational level played a more important role in students' development progress in the Chinese culture.

		IPS	IR	CR
CN	Mother's Education	.330	.291	.326
	Father's Education	.232	.289	.233
HU	Mother's Education	.155	.102	.129
	Father's Education	.122	.097	.125

Note. The correlations are significant at the 0.01 level

Table 5.13 Correlations between mothers' and fathers' educational level and thinking skills

The influence of parental educational level on the level of their children's thinking skills can also be detected indirectly though the usage of the different learning strategies. The following section will discuss this indirect influence in more detail.

5.5.4 The influential level of students' learning strategies

Based on the results of the learning strategies questionnaire, 53.90% of the Chinese students stated that they applied control strategies the most frequently in their daily learning activities. In comparison, 50.78% of the Hungarian students most strongly relied on memorization strategies.

Structural equation modeling was used (1) to monitor the influence of parents' educational level on students' learning strategies in Hungarian and Chinese contexts, and (2) to detect and compare the predictive power of students' learning strategies on their IPS performance in two different cultures (see Figures 5.15 and 5.16). The fit indices were acceptable for both the Chinese (χ^2=222.34, df=111, p<.01, CFI=.93, TLI=.93, RMSEA=.08, SRMR=.08) and the Hungarian (χ^2=244.01, df=110, p<.01, CFI=.96, TLI=.95, RMSEA=.04, SRMR=.04) samples.

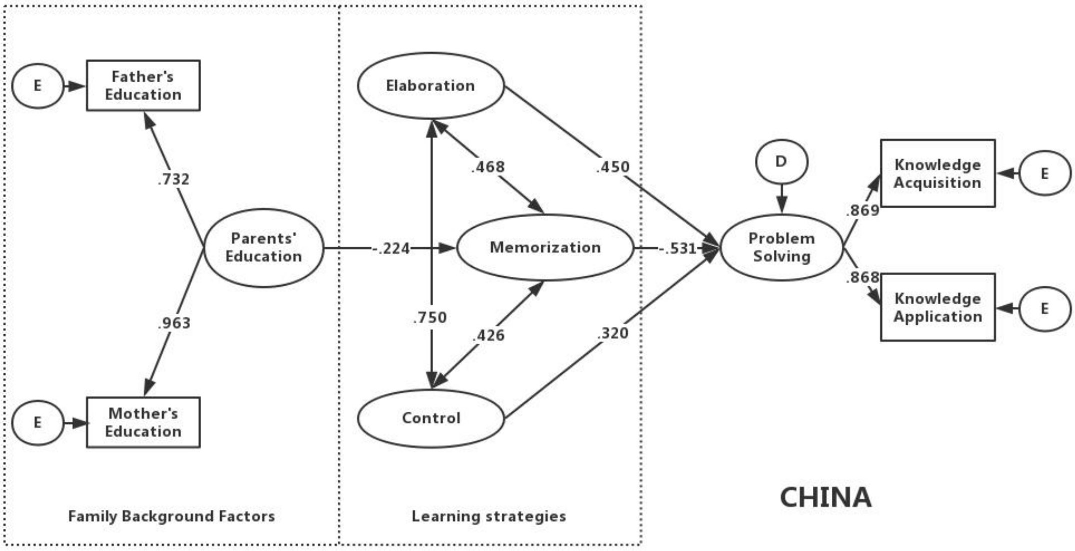

E
Father's Education
E
Mother's Education
.732
.963
Parents' Education
-.224
Elaboration
Memorization
Control
.468
.750
.426
.450
-.531
.320
D
Problem Solving
.869
.868
Knowledge Acquisition
Knowledge Application
E
E
Family Background Factors
Learning strategies
CHINA

Fig 5.15 Relationships between parents' education, learning strategies, and problem solving – Chinese sample

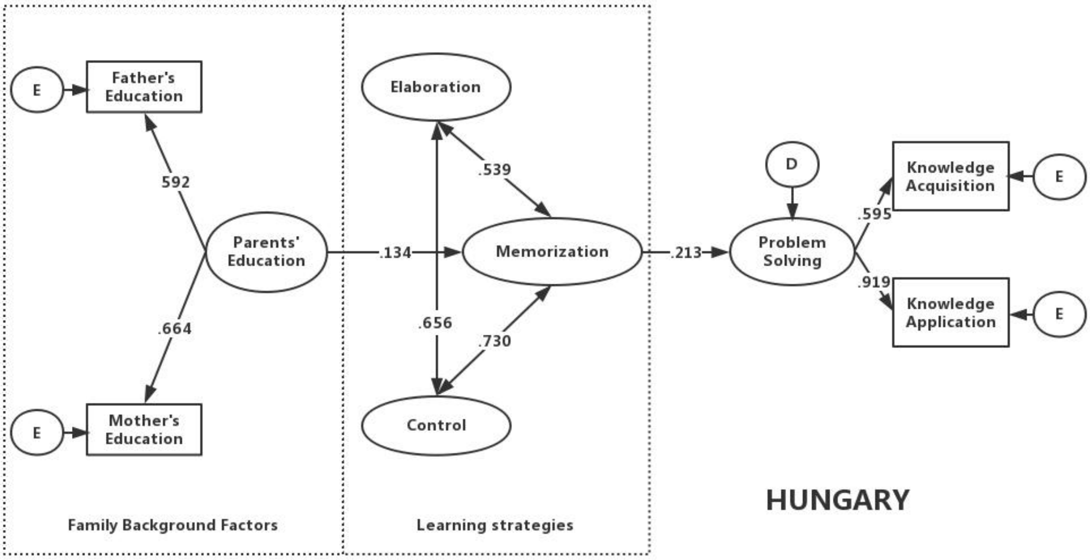

E
Father's Education
E
Mother's Education
592
.664
Parents' Education
.134
Elaboration
Memorization
Control
.539
.656
.730
.213
D
Problem Solving
.595
.919
Knowledge Acquisition
Knowledge Application
E
E
Family Background Factors
Learning strategies
HUNGARY

Fig 5.16 Relationships between parents' education, learning strategies, and problem solving – Hungarian sample

Parents' level of education significantly predicted the use of memorization strategies in both Chinese and Hungarian students. No significant predictive effect can be detected from parents' level of education to control and elaboration strategies in both culture. The effect from parents' level of education to the use of memorization strategies was negative for the Chinese students (β= -.224, p<.01) and positive for the Hungarian students (β= .134, p<.05). That is, the Chinese students with well-educated parents tended to use less memorization strategies in their learning activities, but this was not the case for the Hungarian students; the Hungarian students, even those with well-educated parents, used memorization strategies more often. Chinese students, who often learn by memorizing, proved to be less successful problem solvers (β= -.531, p<.01), while the opposite effect was observed for the Hungarian students (β= .213, p<.01).

The use of elaboration (β= .450, p<.01) and control (β= .320, p<.01) strategies significantly promoted students' problem-solving skills achievement in China. But the use of these two learning strategies had no significant effect on the IPS performance in the Hungarian group. The three learning strategies that were measured had significantly different impacts on students' IPS performance in the two groups. Based on this finding, it is proposed that the Chinese and Hungarian students used different exploration strategies during the IPS assessment process and that the different exploration strategies were impacted differently by the three learning strategies.

To sum up, all the selected non-thinking skill factors (included working memory, learning strategies, test-taking motivation, parents' education) are influencing students' problem solving achievements in both cultures. (RQ7) Even the influential powers were varied in two cultures (especially for learning strategies), but still, *hypothesis H7 was supported*.

5.6 Overall models with all factors included

Fig. 5.17 and 5.18 show all factors in one model (except for test-taking motivation) for both the Chinese and Hungarian samples. The model fits for both models were good: CN: χ^2=25.2, df=13, p<.05, CFI=.98, TLI=.96, RMSEA=.06, SRMR=.04; HU: χ^2=22.6, df=17, p<.05, CFI=1.00, TLI=.99, RMSEA=.02, SRMR=.02.

In order to detect each factor's influence on the two phases of IPS, these models have divided IPS as two entities, knowledge acquisition and knowledge application. In the Chinese model, both knowledge acquisition and knowledge application received predicting effects from thinking skills, inductive and combinatorial reasoning, while the Hungarian model also proved the importance of these two thinking skills in the two phases of IPS. Considering the previous analysis in section 5.4, the conclusion can be drawn that inductive and combinatorial reasoning are important component skills of problem solving, and they are influencing both knowledge acquisition and application. Furthermore, significant correlations between knowledge acquisition and knowledge application, and between inductive reasoning and combinatorial reasoning were detectable in both models, which confirmed the connections between these thinking skills in both cultures.

In the Chinese model, the non-thinking skill factor working memory showed strong predictive power to the developmental level of inductive and combinatorial reasoning (β= .48-59) and moderate power to knowledge acquisition (β= .16). In the Hungarian model, working memory also showed weaker but still relatively high predictive powers to inductive and combinatorial reasoning (β= .28-42). The predicting effect from working memory to knowledge acquisition in the Hungarian model (β= .16) was very close to the coefficient in the Chinese model. A remarkable difference between these two models is that working memory did not have a significant influence on knowledge application for the Chinese sample, but the influence was significant (β= .22) for the Hungarian sample. Thus, the models showed that working memory was influencing students' problem solving at a certain level, however, the influence for the Chinese students was weaker than the one for the Hungarian students.

The previous sections proved that students with different parental educational levels had different problem-solving achievement. But the models indicated that the effects were not direct. For the Chinese sample, mother's and father's educational levels showed predictive power on only one thinking skill, namely inductive reasoning (β= .20-24); for the Hungarian sample, father's educational level did not show predictive power on any thinking skills, and mother's educational level showed moderate predictive power on inductive (β= .13) and combinatorial reasoning (β= .20). Therefore, the path analysis proved that parents' education did not directly influence student's problem-solving achievement in both cultures.

The previous analysis showed that the influence of the different learning strategies was much stronger in the Chinese context than in the Hungarian one. The path analysis models further confirmed and supported this finding. In the Chinese model, almost all the three learning strategies were predicting both knowledge acquisition and application, only elaboration strategies did not show a significant predictive effect on knowledge acquisition. All the predictive powers from elaboration and control strategies were positive (β= .15-.24); while the powers from memorization strategy were negative (β= -.20--.29). In comparison, in the Hungarian model, the connection between learning strategies and thinking skills can only be found between memorization strategy and combinatorial reasoning (β= .09).

To sum up, the SEM model about the component skills of problem solving had a similar structure in both China and Hungary. Some important factors such as inductive and combinatorial reasoning were playing important roles in the developmental level of students' problem-solving skills independently of the national context. Some other factors showed also significantly different influential effect in the two models; for example, working memory did not influence Chinese students' achievement in the knowledge application phase (IPS test), and the different learning strategies had a comparatively weak influence on Hungarian students' IPS achievement (compare Fig. 5.17 and 5.18).

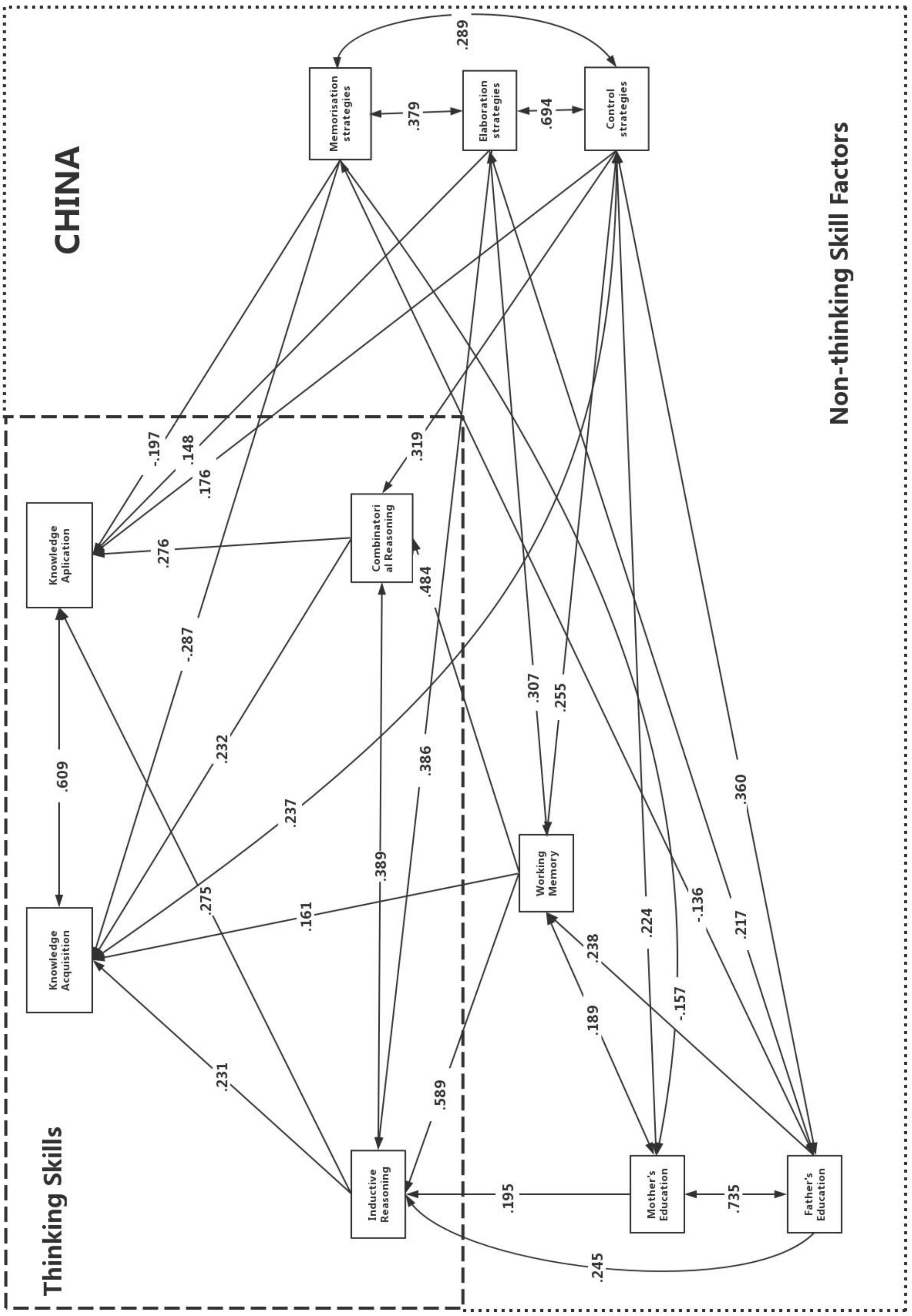

CHINA
Thinking Skills
Non-thinking Skill Factors
Memorisation strategies
Elaboration strategies
Control strategies
Knowledge Aplication
Combinatorial Reasoning
Knowledge Acquisition
Inductive Reasoning
Working Memory
Mother's Education
Father's Education
.289
.379
.694
-.197
.148
.176
.319
.276
.484
-.287
.307
.255
.609
.232
.386
.237
.360
.389
.275
.161
-.136
.224
.217
.238
.189
-.157
.231
.589
.195
.735
.245

Fig. 5.17 Relationships between each measured skill and factor – Chinese sample

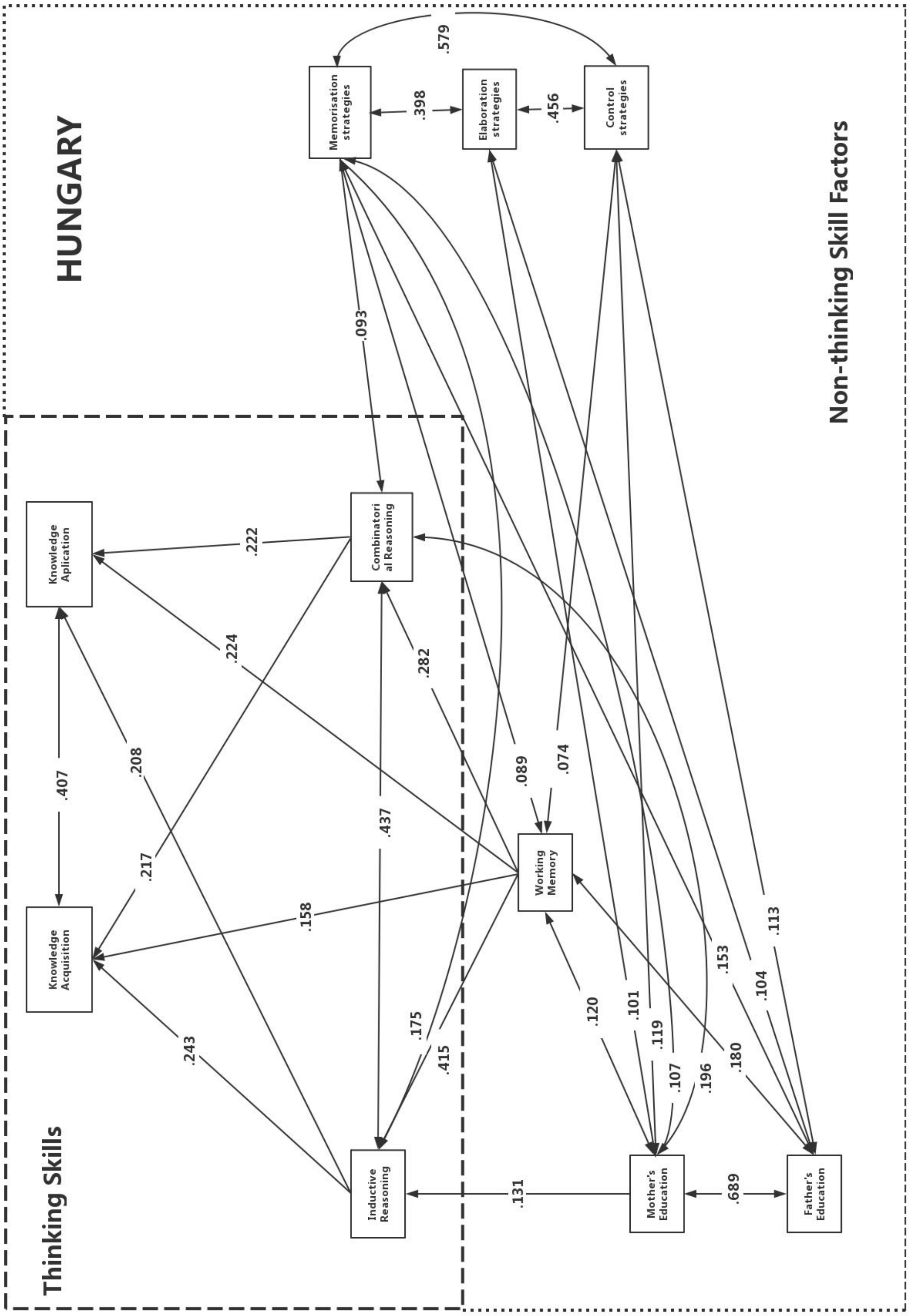

HUNGARY
Thinking Skills
Non-thinking Skill Factors
Memorisation strategies
Elaboration strategies
Control strategies
Knowledge Aplication
Knowledge Acquisition
Combinatori al Reasoning
Inductive Reasoning
Working Memory
Mother's Education
Father's Education
.579
.398
.456
.093
.222
.224
.282
.407
.208
.217
.437
.158
.243
.175
.415
.089
.074
.120
.101
.119
.107
.196
.153
.104
.113
.180
.131
.689

Fig. 5.18 Relationships between each measured skill and factor – Hungarian sample

5.7 Discussions and limitations

This chapter presented the result based on a cross-cultural comparison study, which focused on the relationships between IPS performance and students' developmental level of different reasoning skills (inductive reasoning, combinatorial reasoning) and other non-thinking factors (working memory, learning strategies, test-taking motivation, social-economic background factors). The analysis aimed to provide answers from research question RQ2 to research question RQ7.

Unlike what we expected in hypothesis H4, Chinese students and Hungarian students showed equal IPS performance at the starting age (12-year-olds) of the reasoning skills development period. The hypotheses were put forward based on the fact that Chinese students showed a significant advantage over Hungarian students in the PISA 2012 problem-solving assessment (for 15-year-olds). What the reasons behind the gap between the expected and real results could be is still unclear. Maybe it is an indicator that Chinese students have a higher development speed in the 12-15 years age range; but it is also possibly caused by some other reasons. Here we expect that the follow-up log-file data analysis would provide more information about this problem. Meanwhile H5 was partly supported, in that only Hungarian males showed better performance than females, as expected.

Chinese and Hungarian students showed some common features regarding the cognitive construction of IPS. As we expected in H2 and H6, IPS is composed of two different processes, knowledge application and knowledge, and inductive reasoning and combinatorial reasoning; these two thinking skills played significant roles in the problem-solving process. These features proved to be of common nature of problems solving, independently of the cultural context. However, measurement non-invariance of interactive problem solving was found across nationality (H3 was denied), which indicated that there were still many differences between Chinese and Hungarian students' cognitive constructions of problem solving. For instance, even both inductive and combinatorial reasoning showed a significant predicting power on problem solving, but the levels of prediction were different in these two cultures. The influences of the selected non-thinking factors (working memory, parents' education, learning strategies, and test-taking motivation) on the problem-solving achievement were confirmed (H7 was supported), but the influences also showed some differences between the two cultures, such as Chinese students' learning strategies were deeply relevant with their IPS performance, but this was not the case for the Hungarian students. These differences might have been relevant or even caused by the different exploration strategies students employed in the IPS assessment, thus the log-file analysis also aimed to find more clues to understand these differences.

To sum up, the study effectively answered the research questions, which contributed to our further understanding of the nature of problem solving in European and Asian contexts. However, some unsolved problems were detected during the analysis, which could be answered by analyzing the log-files of the study.

There are some limitations in this cross-national comparison study. The sample size for the assessment was not very high. There was also a difference between the sample sizes for the Chinese and Hungarian participants. Even though it did not cause any trouble or error during the comparison study (e.g. statistical methods were used to exclude the possibility that measurement non-invariance was caused by the different sample sizes), it still might impact the data analysis accuracy in a certain level. Moreover, the detected measurement non-invariance across nationalities, on the one hand, is a very interesting and important finding, but, on the other hand, it has the possibility to impact the accuracy of the results, especially regarding the cross-national comparison. Similarly, the different test-taking motivation between the two groups of students could also have the risk to impact comparison results. Thus, repetition of the study is required for validation in the future.

6. Cross-national comparison study of students' problem solving strategies: Log-file analysis

6.1 Background

Previous analyses in chapter 5 have indicated some similarities and differences between Chinese and Hungarian students' problem-solving skills and its cognitive structure. In order to further analyze the root reasons regarding the cognitive structural differences between these two cultures, this chapter is going to detect Chinese and Hungarian students' behaviours (both mental and physical) during the IPS environment by log-file analysis.

With the log-file analyses we focus on students' exploration behaviour, thus on the first phase in the problem-solving process, on the knowledge acquisition phase. Molnár and Csapó (2018) developed a labeling system and a mathematical model to classify whether students' exploration strategies are effective. The classification was based on the amount of information extracted during the exploration of the problem scenario. For instance, an exploration strategy was defined as theoretically effective if the problem solver was able to extract all the information needed to solve the problem (independently of the final achievement). In the present study students' actions were logged and coded according to the input behaviour model by using Molnár and Csapó's labeling procedure, then we distinguished whether students' exploration strategies were 1) effective and 2) belong to the VOTAT scope.

6.2 Aims, research question and hypothesis

The log-file analysis aimed to gain a further understanding of the similarities and differences of Chinese and Hungarian students' behaviour in the interactive problem-solving environment. Research question RQ8 is expected to be answered by the results. Moreover, hypothesis H8 is going to be verified in this chapter as well.

RQ8. Do Chinese and Hungarian students employ different exploration strategies during the problem-solving process? How do their exploration strategies influence their problem-solving performance?

H8: Chinese and Hungarian students employ different exploration strategies during the problem-solving process. The adoption of different exploration strategy can lead to different problem-solving performance.

6.3 Methods

6.3.1 Participants

The study was a follow-up analysis of the same assessment project which was introduced in the previous sections. Thus, the participants of this study were same as it was in the previous study, that is: 187 Chinese students (85 boys and 102 girls; mean age=11.93, SD=1.06) and 835 Hungarian students (382 boys and 453 girls, mean age=11.86, SD=.43).

6.3.2 Instruments

This study analyzes students' exploration strategies in the exploration phase of the problem-solving process. Thus, we focused only on the first phase of the problem-solving process, on the knowledge acquisition phase, where the problem solver had to interact directly with the problem situation by manipulating the input variables, which could have an influence on the output variables. We focused on the type of the manipulations of the input variables, on its systematism and effectiveness based on the amount of the extracted information. We selected items with the same complexity and in the same item-position for the analyses. Complexity was defined by the number of input and output variables and the number of relations.

6.3.3 Scoring

As an achievement indicator, students' responses in the knowledge acquisition phase were scored as correct ("1") if the connections between the variables were accurately indicated on the concept map, thus the drawings indicating the relationships between the input and output variables fully matched the underlying problem structure. Otherwise, the response was scored as incorrect ("0"). Beyond the achievement data the system logged every setting of the input variables, which was tested by clicking on the application button during the exploration phase of a problem. Thus, the unit of the labeling process was a trial. The sum of these trials, within the same problem environment, was called the input behaviour, which was called a strategy if it followed meaningful regularities.

"By our definition, the full input behaviour model describes what exactly was done throughout the exploration phase and what kinds of trials were employed in the problem-solving process. It consists of all the activities with the sliders and application buttons in the order they were executed during the first phase, the exploration phase of the problem-solving process. The basic input behaviour is part of the full input behaviour model by definition, when the order of the trials executed was still being taken into account, but it only consists of activities where students were able to acquire new information on the system" (Molnár, & Csapó, 2018, p. 10).

Some operations and activities were not included in the basic input behaviour model, they were deleted (e.g. redundant operations, meaningless operations, etc.) from the full input behaviour model to obtain the basic behaviour model. Finally, students' minimal input behaviour model was generated from the full input behaviour model. It focuses on those activities (without the real order of the trials), where students were able to obtain new information from the scenario. That is, the minimal input behaviour models only those manipulations which were able to provide useful and new information for the problem-solvers. If students did not apply any useful operation from the amount of the extracted information point of view (e.g. repetition), their strategy was marked as null.

The labeling, formatting and representing method of the minimal input behaviour models employed Molnár and Csapó's (2018) design, as follows:

1. "Only one single input variable was manipulated, whose relationship to the output variables was unknown (we considered a relationship unknown if its effect cannot be known from previous settings), while the other variables were set at a neutral value like zero. We labeled this trial +1.
2. One single input variable was changed, whose relationship to the output variables was unknown. The others were not at zero, but at a setting used earlier. We labeled this trial +2.
3. One single input variable was changed, whose relationship to the output variables was unknown, and the others were not at zero; however, the effect of the other input variable(s) was known from earlier settings. Even so, this combination was not attempted earlier. We labeled this trial +3." (Molnár, & Csapó, 2018, p. 8)

After labeling and recording students' exploration strategies in the problem-solving environment, we analyzed (1) whether students' exploration strategies were theoretically effective, and (2) whether students' exploration strategies were VOTAT strategies.

The effectiveness of the exploration strategies was defined and linked to the amount of information extracted. If the problem solver was able to extract all the information needed to solve the problem (no matter if he/she correctly solved the problem or not), his/her exploration strategy was defined as theoretically effective.

Base on the definition of VOTAT, if a student applied +1, +2, or +3 trial for every input variable in one task, it can be considered as if he/she was using VOTAT strategy in this task. Moreover, the study labeled students' exploration strategy in each task into three groups: (1) 0 point: no isolated variation (no isolated variation was employed); (2) 1 point: partially isolated variation (isolated variation was employed for some but not all the input variables); and (3) 2 points: fully isolated variation (isolated variation was employed for all the input variables).

6.3.4 Data analyses

Basic statistical analyses were implemented via SPSS to see Chinese and Hungarian students' similarities and differences regarding exploration strategies adoption. How exploration strategies influenced students' performance in the IPS tasks in both cultures was analyzed as well.

Latent class analysis (Collins, & Lanza, 2010) was employed in this study. LCA is a latent variable modeling approach which can be used to identify latent classes of samples who share similar observed variables (Collins, & Lanza, 2010). In this study, it was employed to build latent classes based on students' exploration strategies adoption. The LCA analysis was carried out by Mplus (Muthén, & Muthén, 2010). This study measured the quality of LCA analysis by fit indices: Akaike information criterion (AIC, lower values indicated a better model fit), Bayesian Information Criterion (BIC, lower values indicated a better model fit), adjusted Bayesian Information Criterion (aBIC, lower values indicated a better model fit), entropy (within [0,1] scale, values near one, indicating high certainty in classification), Lo-Mendell-Rubin Adjusted Likelihood Ratio (compares the model containing n latent classes with the model containing n-1 latent classes, p-value was the indicator for whether significant difference can be detected; Lo et al., 2001).

6.4 Results and discussions

6.4.1 Cross-cultural comparison of the theoretically effective strategy usage in problem-solving environment

The ratio of the effective strategy usage for Chinese students were slightly higher than that of the Hungarian students in most of the tasks (except 2-2 type items), but the differences were very small (see Table 6.1). The highest percentage of effective strategy usage was detected in an item with the simplest structure (2-1 type; around 70%). However, the percentage decreased with the increasing complexity of the problems.

Nationality	Complexity of problem (number of input and output variables)	Percentage (%)	
		The use of effective strategy	The use of non-effective strategy
CN	2-1	70.6	29.4
	2-2	65.3	34.7
	3-2	58.8	41.2
	3-3	50.3	49.7
HU	2-1	69.8	30.2
	2-2	66.9	33.1
	3-2	53.2	46.8
	3-3	48.6	51.4

Table 6.1 Percentage of the use of theoretically effective and non-effective strategy

Table 6.2 visualizes the problem-solving performance of the Chinese and Hungarian theoretically effective strategy users. Even for the simplest items, around 30% of the theoretically effective strategy users still could not achieve a high performance, and the percentage was increasing along with the items' complexity. By contrast, even if students did not apply the theoretically effective strategy (Table 6.3), they still got the chance to correctly solve the problem. The chance was actually very high (above 50%) for the simplest item (2-1 type). For the rest of the items, the chances were much lower (ranged from 7% to 23%), but still could not be ignored. The results indicated that the use of a theoretically effective strategy did not always result in high performance for both Chinese and Hungarian students. Meanwhile, the high achievement did not always represent the use of an effective problem-solving strategy.

Chinese effective strategy users had a remarkably higher percentage of correct problem solving than the Hungarian effective strategy users. In comparison, the probability for a Hungarian non-effective strategy user to receive high achievement was much higher than for a Chinese non-effective strategy user. That is, in the interactive problem-solving environment, Chinese students relied much more on the use of effective strategies. Based on the information extracted, Chinese students were more able to find the right solution to the problems, they managed to represent the information that they had obtained from the system more effectively, and they made good decisions in the problem-solving process compared to their peers in Hungary.

Nationality	Complexity of problem (number of input and output variables)	Frequency (%)	
		High achievement	Low achievement
CN	2-1	74.2	25.8
	2-2	78.0	22.0
	3-2	60.9	39.1
	3-3	61.0	39.0
HU	2-1	65.7	34.3
	2-2	43.5	56.5
	3-2	50.5	49.5
	3-3	55.1	44.9

Table 6.2 Problem-solving performance for the theoretically effective strategy users

Nationality	Complexity of problem (number of input and output variables)	Frequency (%)	
		High achievement	Low achievement
CN	2-1	54.5	45.5
	2-2	13.5	86.5
	3-2	7.8	92.2
	3-3	7.7	92.3
HU	2-1	61.1	38.9
	2-2	17.1	82.9
	3-2	22.3	77.7
	3-3	14.7	85.3

Table 6.3 Problem-solving performance for the non-theoretically effective strategy users

6.4.2 Cross-cultural comparison of the efficacy of VOTAT strategies

VOTAT strategies were more frequently employed by the Chinese students. Base on the statistics (see Table 6.4), for the problems with two input variables (2-1 and 2-2 type), more than 56% of the Chinese students employed the full VOTAT strategy. In comparison, only around 43% of the Hungarian students employed a full VOTAT strategy for the same types of problems. For the problems with three input variables (3-2 and 3-2 type), still more than 45% of the Chinese students employed a full VOTAT strategy, which was far higher than the percentage for the Hungarian students (29%-37%).

Complexity	No VOTAT (%)		Partly VOTAT (%)		Full VOTAT (%)	
of Problem	CN	HU	CN	HU	CN	HU
2-1	33.7	42.6	4.8	14.0	61.5	43.4
2-2	38.1	44.9	5.9	11.6	56.2	43.7
3-2	41.7	58.4	5.9	12.0	50.2	29.7
3-3	47.9	53.9	7.0	9.7	45.2	36.5

Table 6.4 Percentage of VOTAT strategy use

The +1+1 (or +1+1+1 for the problems with three input variables) was the only VOTAT strategy which always provided a higher probability to correctly solve the problem independently of the problem complexity in both cultures (see Fig. 6.1 and 6.2). In this study, this strategy was named as *VOTAT strategy A*. In the +1+1 (or +1+1+1) type strategy, for every operation only one single input variable was manipulated, while the other variables were set at a neutral value (at zero in the case of MicroDYN). Based on the results, this strategy was the most effective strategy in the interactive problem-solving environment built in this study.

The second most frequently employed VOTAT strategy was the +1+2 type (in case of two input variables) or the +1+2+2 type (in case of three input) strategy. This strategy was named as *VOTAT strategy B*. In this type of strategies, in the first step, only one single input variable was manipulated, while the other variables remained at zero; in the second step, the input variable which was changed in the first step remained at the previous status, while the other input variable was manipulated. This type of strategy can also provide a certain probability to correctly solve the problem. However, the users of this type of strategy also had a comparatively high possibility to provide a wrong answer. In most of the cases, the possibility to provide a wrong answer was higher or almost equal with the possibility to provide a right answer (except for problems with 2 input variables and 2 output variables in the Chinese context). Thus, the effectiveness of this kind of strategy was not very high, especially compared with the +1+1 (or +1+1+1) strategy.

The +1+3 type (in the case of two input variables) or the +1+1+2 type (in the case of three input variables) was the less effective VOTAT strategy. This strategy was named as *VOTAT strategy C*. This strategy was rarely employed by the participants. Thus, its effectiveness was hard to accurately testify. In this case, this strategy is not going to be further discussed.

By the comparison of the exploration behaviour of the Chinese and Hungarian students, we detected some remarkable differences:

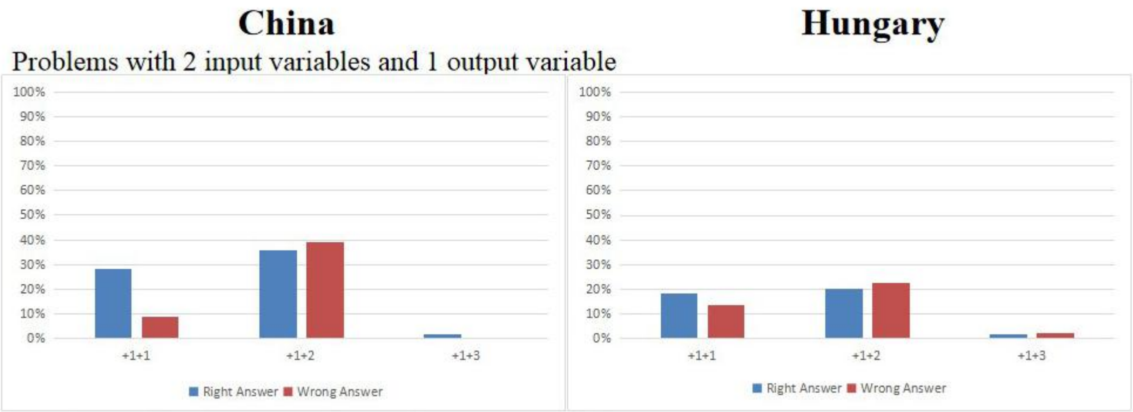
(1) VOTAT strategy A has been proved as the most effective strategy. That is, the VOTAT strategy A users had a high probability of getting the correct answer. However, in every item, the proportions for Hungarian VOTAT strategy A users with a wrong answer were higher than that of the Chinese users. All the VOTAT strategy A users extracted the same information from the system. Thus, the ability to effectively use the extracted information was the decisive factor on whether they could successfully solve the problem. Therefore, the result further proved the finding achieved in section 6.4.1, according to which, compared to Chinese students, Hungarian students had a lower ability to effectively use the extracted information.

(2) The frequency for Chinese students to employ VOTAT strategy B was higher than the frequency for the Hungarian students. VOTAT strategy B had higher effectiveness in the Chinese context. The result confirmed Chinese students' higher ability level to effectively use the extracted information. Moreover, this finding also matches the results in section 6.4.1. In IPS environment an effective strategy usage was more valuable for the Chinese problem-solvers, since they could use the extracted information more effectively.

(3) When comparing VOTAT strategy A with VOTAT strategy B, the latter one is more complex. If a student employed VOTAT strategy B, he/she needed to do a more complex analysis and calculation, which required students' combinatorial reasoning skill and mathematical computation ability. Students

have a certain possibility to employ VOTAT strategy by accidental random operations. Of course, if a student employed VOTAT strategy involuntary, but not fully understood what he/she has done, he/she was not very likely to effectively use the extracted information and solve the problem correctly. As the results indicated (Fig. 6.1 and 6.2), Chinese students had significantly higher proportion for employing VOTAT strategy B and they had a much higher possibility to achieve the correct answer by employing VOTAT strategy B. That is, there were much more Chinese students who understood how VOTAT strategy worked in the problem-solving environment during the assessment process. Moreover, applying VOTAT strategies requires students to figure out the combination of the setting of input values, and select the combination which can most effectively provide information. Therefore, the results also answered the question raised in section 5.4.4, namely why combinatorial reasoning skill had a more important meaning for Chinese students.

Problems with 2 input variables and 1 output variable



Problems with 2 input variables and 2 output variables

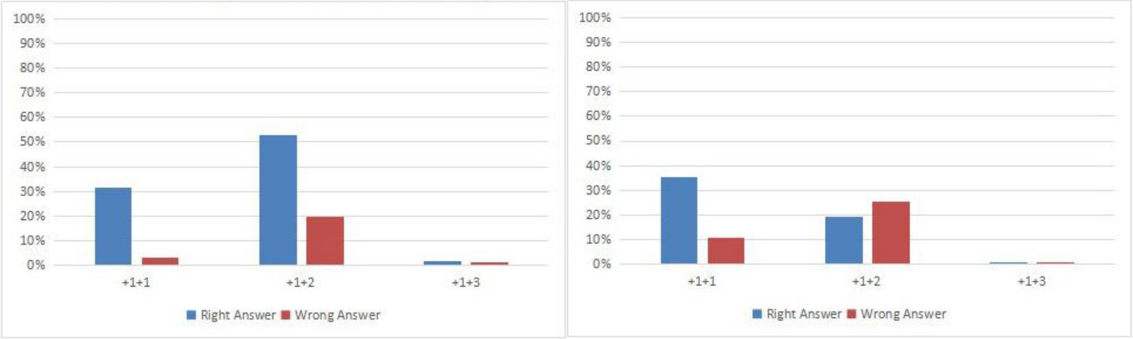


Fig. 6.1 Efficacy of the most frequently employed VOTAT strategies on problems with two input variables

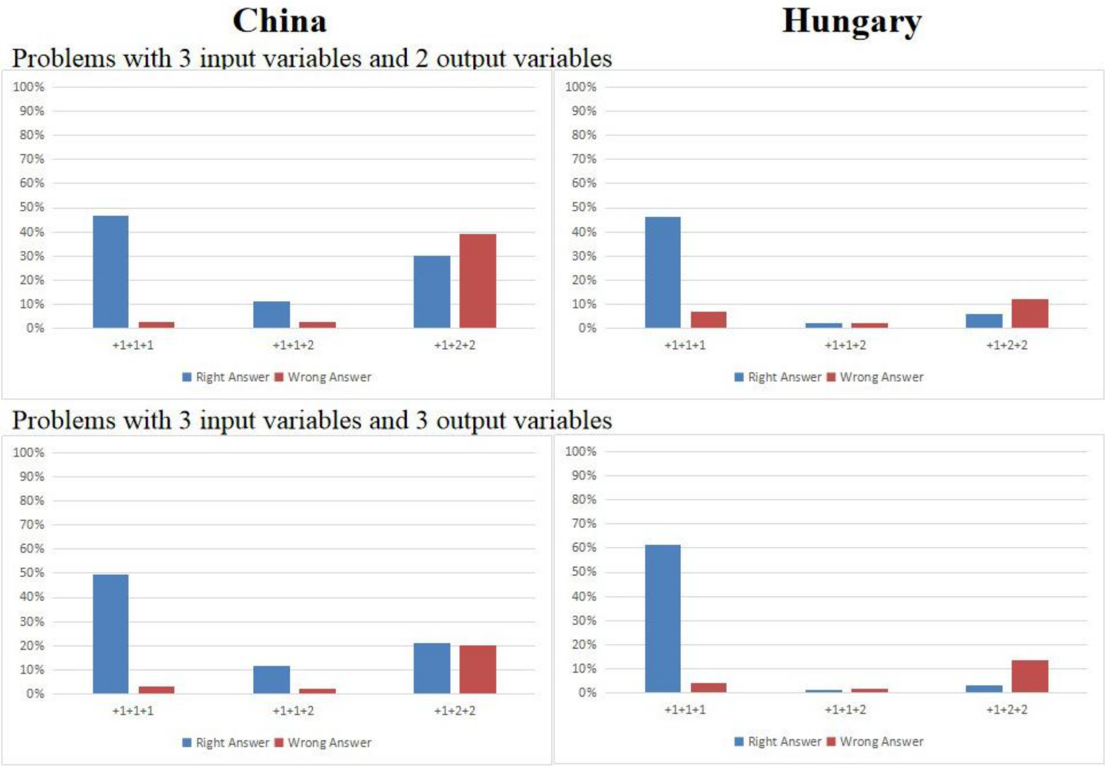


Fig. 6.2 Efficacy of the most frequently employed VOTAT strategies on problems with three input variables

6.4.3 Distinguishing qualitatively different explorer class profiles in China and in Hungary

Following Molnár and Csapó (2018, p. 13), we executed latent class analyses based on students' minimal input behaviour. "Depending on the level of optimal exploration strategy used, all the students received new categorical scores [0, 1, 2, see section 6.3.3] based on their input exploration behaviour, one for each of the CPS tasks."

The 3 latent-class model had the lowest information theory criteria fits (AIC, BIC, and aBIC) and the highest entropy in the Chinese sample. The likelihood ratio statistical test (Lo-Mendell Rubin Adjusted Likelihood Ratio Test) confirmed that the 3 latent-class model had a significantly better fit than the other models ($p<.05$) (Table 6.5). In the Hungarian context, the 3 latent-class model had the lowest information theory criteria fits; however, its entropy was lower than for the 2 latent-class model. The likelihood ratio statistical test proved that the 3 latent-class model fitted the data better (p<.001). Thus, both the Chinese and Hungarian samples should be 3 latent classes classified.

Nationality	Number of latent classes	AIC	BIC	aBIC	Entropy	L-M-R test	p
CN	2	1,480	1,561	1,482	0.969	509	<.001
	3	1,466	1,589	1,469	0.978	39	<.05
	4	1,462	1,627	1,466	0.973	30	>.05
HU	2	7,791	7,909	7,830	0.930	1,873	<.001
	3	7,550	7,730	7,609	0.817	264	<.001
	4	7,511	7,752	7,590	0.787	64	>.05

Table 6.5 Fit indices for latent class analyses

The 3 latent classes models for both the Chinese and Hungarian samples are shown in Fig. 6.3 and 6.4. Based on the students' performance, the three qualitatively different class profiles for the Chinese sample were described as follows:

(1) Class 1 – Intermediate performers on the easiest problems, but low performers on complex ones with a very slow learning effect;
(2) Class 2 – Rapid learners; and
(3) Class 3 – Proficient strategy users.

Meanwhile, the three qualitatively different class profiles for the Hungarian sample were described as follows:

(1) Class 1 – Intermediate performers on easiest problems, but low performers on complex ones with a very slow learning effect;
(2) Class 2 – Proficient strategy users; and
(3) Class 3 – Low performers.

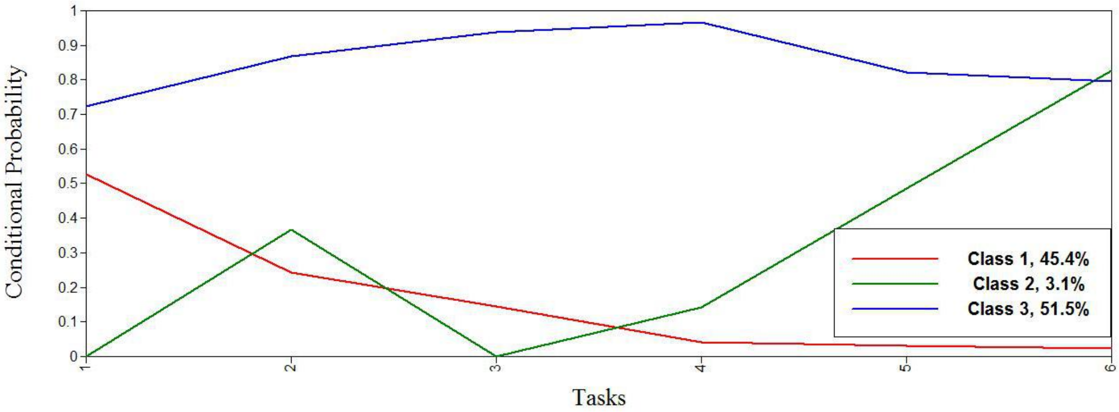


Fig. 6.3 Three qualitatively different class profiles for the Chinese sample

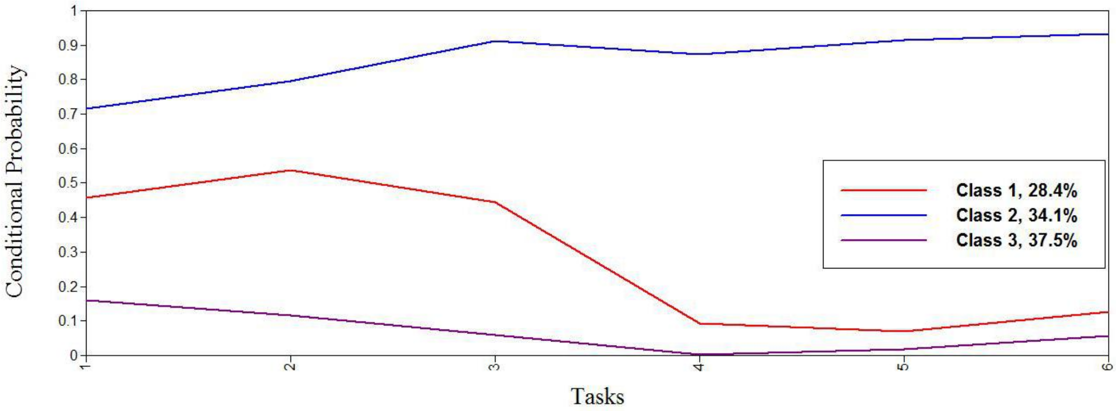


Fig. 6.4 Three qualitatively different class profiles for the Hungarian sample

The frequencies and average latent class probabilities for the latent classes are displayed in Table 6.6. 37.5% of the Hungarian students were low performers. During the IPS assessment, they rarely employed a fully or partially VOTAT strategy. 28.4% of the Hungarian students were slow learners. They employed a fully or partially VOTAT strategy in the easiest items, but they did not learn to control and understand the strategy. After the problem complexity increased, they became low performers. 34.1% of the Hungarian students were proficient strategy users and employed a VOTAT strategy from the beginning to the end.

Profiles	CN		HU	
	Frequency (%)	Average latent class probabilities	Frequency (%)	Average latent class probabilities
Low performers	-	-	37.5	.91
Intermediate performers on easiest problems, but low performers on complex ones with a very slow learning effect	45.5	.99	28.4	.85
Rapid learners	3.1	.95	-	-
Proficient strategy users	51.5	.99	34.1	.97

Table 6.6 Frequencies and average latent class probabilities

There were several remarkable advantages detectable when we compared Chinese and Hungarian students regarding the usage of the VOTAT strategy. The class “low performers” did not exist in the Chinese sample, indicating that there were almost no Chinese students who did not employ the VOTAT strategy at least once in the IPS test. 45.5% of the Chinese students were intermediate performers on the easiest problems, but low performers on complex ones with a very slow learning effect. They did not perform well on the whole test due to the slow learning effect, but at least they showed on average performance on the easy problems.

The proportion of proficient strategy users in the Chinese sample was 51.5%, which was much higher (34.1%) than the ratio for the Hungarian sample. It was an obvious and direct indicator that Chinese students had a generally better performance regarding the VOTAT strategy learning, understanding, and usage. A small group of Chinese students (3.1%) were detected as rapid learners. They did not employ fully or partially VOTAT strategy on the easy items at the beginning of the test, but they learned very rapidly and reached the top performers’ proficiency level of the VOTAT strategy usage by the end of the test. However, there was no rapid learner found in the Hungarian sample, which can be due to the relative low sample size. In a large-scale study Molnár and Csapó (2018) could distinguish 6 qualitatively different explorer class profiles regarding IPS, including the class of rapid learners (4.4% of the whole sample).

6.5 Conclusions and limitations of the log-file analyses

The analysis discovered Chinese and Hungarian students' problem-solving behaviours in the aspect of exploration strategy in IPS environment. The results from this analysis could be used to answer research question 8 (*RQ8. Do Chinese and Hungarian students employ different exploration strategies during the problem-solving process? How do their exploration strategies influence their problem-solving performance?)*

Chinese students employed theoretically effective strategies – including the most effective VOTAT strategy – more frequently than Hungarian students. Compared with the Hungarian students, the Chinese students managed to represent the information that they obtained from the system more effectively, and they made good decisions in the problem-solving process. That is, the employment of a theoretically effective, especially of the VOTAT strategy, had more value for the Chinese students, since if they collected enough information to solve the problems, they had higher chances to build the right solution. In contrast, Hungarian students' problem-solving performance received less influence from the exploration strategies they employed. However, the Hungarian student proved to be more intuitive, and, as a result, the Hungarian theoretically non-effective strategy users had a higher probability to correctly solve the problem (compared to the Chinese students). In general, there were several remarkable differences detected between Chinese and Hungarian students' exploration strategies, thus, *hypothesis H8 was supported.*

The analysis also further explored the nature of problem solving in the Hungarian and Chinese contexts. Chinese students showed a much higher performance in the PISA 2012 problem-solving assessment than Hungarian students. However, against our preliminary hypotheses, in this study, the students from these two countries showed statistically equal performance.

One of the reasons for this difference could be that we measured 12-year-old students, who are – according to Molnár et al. (2018) – still at the beginning of the IPS developmental phase, while PISA-students are in the middle of the IPS developmental phase (15-years-olds). Thus, the difference between Chinese and Hungarian students' ability of problem solving has not occurred at the starting stage.

This log-file analysis provides another possibility to understand this situation. From the log-file analysis, some information has been extracted. For instance, Chinese students employed more frequently theoretically effective exploration strategies during the problem solving process, and they benefited much more from the effective strategy usage. Chinese students had a higher proportion of rapid learners and proficient strategy users than Hungarian students. Compared with Chinese students, Hungarian students did not highly rely on the use of theoretically effective strategy in the interactive problem-solving environment.

Molnár and Csapó (2018) proved that the higher the complexity and difficulty of a problem, the lower the possibility of correctly solving it without an effective strategy usage. Therefore, Chinese students' higher proficiency level of the effective strategy usage could play a much important role in the problem-solving assessments with higher complexity and difficulty. For instance, the PISA 2012 problem-solving assessment had higher complexity and difficulty than this study (it was designed for 15-year-old students). Based on the findings in this study, there is reason to believe that Chinese students' effective strategy usage helped them to a certain level to exceed Hungarian students.

Moreover, the analysis has also answered the question raised in section 5.5.4, namely why Chinese students' problem-solving performance had a deeper connection with their learning strategies. The proficiency level of the exploration strategy in the interactive problem-solving environment is actually connected with students' daily learning strategies (Molnár et al., 2017). Thus, there is no wonder that Chinese students' problem-solving performance was more likely to be influenced by their learning strategies. Moreover, obviously, employing different exploration strategies requires different thinking skills. In section 5.4.4, we discussed that because Chinese students were more willing to and better at using VOTAT strategies, their combinatorial reasoning played a much more important role than the Hungarian students. But this was only a very basic finding. How students' thinking skills and their exploration strategies are connected in the problem-solving environment is an interesting, but huge and unexplored topic. The results indicated the possibility for further research to discover a larger picture of the connections between students' cognitive features and physical behaviour in the problem-solving environment.

The limitations for the log-file analysis were the same as in the cross-national comparison study. The sample size problem might impact the data analysis accuracy to a certain level. Even there has been no direct evidence that can prove the impact, the possibility still could not be fully eliminated.

7. Summary of the Study

7.1 Summary of the results and answering the research questions

The study was aiming to explore the nature of problem-solving in the Chinese and Hungarian contexts. The method of this study was using computer-based online instruments to measure Chinese and Hungarian students' problem-solving skills in the MicroDYN approach, as well as some other thinking skills (combinatorial reasoning, inductive reasoning, creativity) and cognitive and non-cognitive factors (working memory, test-taking motivation, learning strategies, parents' education, etc.). Regarding age, the aim population is in the first phase of the IPS developmental curve (12-year-olds). The present study aimed to detect the similarities and differences in the cognitive structures of problem solving for Chinese and Hungarian students. Log-file analyses were performed to learn more about the behaviour of the Chinese and Hungarian students in an IPS scenario.

Since computer-based assessment of thinking skills has not been commonly implemented in China, we had to test its feasibility in this educational context, and we had to test the behaviour, the psychometric indices, especially the reliability of the tests, adapted within the confines of this assessment project. A pilot study was conducted in order to prove that computer-based assessment is feasible and the selected instruments are reliable to be used in the Chinese cultural and network environment. All of the tests received good or at least acceptable internal consistencies, thus the usability of computer-based testing and the reliability of the instruments were proven. *(RQ1)*

The cross-national comparison study with a larger sample size was conducted in China and Hungary. The tests delivered to Chinese and Hungarian students were not completely the same, but a certain number of anchor items made the comparison of the results possible (by means of item response theory). The results indicated that in both Chinese and Hungarian contexts, IPS showed better model fits in the confirmatory factor analysis with the two-dimensional model consisted of knowledge acquisition and knowledge application. Thus, the IPS tasks proved to assess knowledge acquisition and knowledge application with the MicroDYN approach, independently of cultural context. *(RQ2)*

Based on the measurement invariance analyses, the IPS test proved to be measurement invariant across gender within the same culture. However, measurement non-invariance of the IPS test was found across nationality. That is, the tasks in the IPS test measure not completely the same construct in the different cultures. *(RQ3)*

Analysis by means of item response theory indicated that there was no significant difference between Chinese and Hungarian students' problem-solving skills. Thus, Chinese and Hungarian students had the same developmental level of problem solving at the beginning of their IPS developmental phase. *(RQ4)*

A comparison study between Hungarian males and females and Chinese males and females was conducted. Theresults showed that Hungarian males, Chinese males, and Chinese females had equal achievement in the IPS assessment. However, Hungarian females had the lowest achievement, which was significantly lower than for Hungarian males. That is, among these four groups of students, only the difference between Hungarian males and females was proved. *(RQ5)*

Inductive reasoning and combinatorial reasoning skills showed a significant predictive power on IPS achievement in both Chinese and Hungarian contexts. Thus, these two thinking skills were proved to be important component skills of problem solving. However, in the IPS environment, Hungarian students relied on inductive reasoning and combinatorial reasoning skills at an almost equal level, but Chinese students relied much more on their combinatorial reasoning skill. The remarkably different cognitive styles between Chinese and Hungarian students were one of the important reasons for the detected measurement non-invariance across nationality. *(RQ6)*

Working memory and parental education showed significant influence on both Chinese and Hungarian students' IPS achievement. All the three learning strategies (elaboration, memorization, control) were significantly influencing Chinese students' IPS achievement. In contrast, Hungarian students' problem-solving achievement was rarely connected with their learning strategies. Only memorization had a moderate level of predictive power on problem solving. *(RQ7)*

According to the log-file analyses, the Chinese students employed theoretically effective (including VOTAT) strategies more frequently and they proved to be better strategy users. They had a better ability to give a meaning to the extracted information and to transfer the usage of theoretically effective strategies to successful problem solving. Moreover, the results proved that Chinese students had a higher learning effect regarding the use of VOTAT strategies in the interactive problem-solving environment. In comparison, Hungarian students were less likely to rely on the usage of effective strategies in IPS environment, and that could be a significant disadvantage which impedes Hungarian students' problem-solving skills development. *(RQ8)*

7.2 Discussions of the findings

This study contributed to our further understanding of the nature of problem solving, and explored the differences between students with Chinese and Hungarian cultural backgrounds. Currently, enhancing students' ability to solve problems has become one of the main targets in school education. With no doubt, a deeper understanding of problem solving in an international context has an important meaning.

China (mainland) was one of the top performers in the PISA 2012 problem-solving assessment. However, their performance was still far lower than the expected level, and sadly there has only been very little research of domain-general problem-solving skills conducted in China. Hungary could not achieve a satisfactory performance in the PISA 2012 problem-solving assessment. Thus, it was necessary to conduct a research regarding students' problem-solving skills in these two countries.

This study has successfully discovered several pieces of valuable information. Firstly, it has proved the reliability and feasibility of conducting the computer-based cognitive assessment in the Chinese culture and network environment. China has a generally high quality education system (proved by the PISA results). However, there is still criticism saying that the educational methodology and technology in China is conservative. Computer-based cognitive assessment has not been commonly used by Chinese psychologists. This study proved that, in the Chinese context, computer-based assessment in the domain cognitive sciences can still show its high efficiency. The findings could be the base for future cognitive research to be implemented in China.

This study further confirmed the structure of problem solving. In the assessment, participants were able to demonstrate their capacity for knowledge acquisition and application separately. The modeling results supported the hypothesis proposed by previous studies, namely that problem solving could be formulated as a two-dimensional measurement model (e.g. Bühner, Kröner, & Ziegler, 2008; Wüstenberg et al., 2012), independently of the cultural context. However, the two-dimensional model is still a simplified representation of real-life problem-solving progress. As Greiff et al. (2013) have illustrated, the complexities of naturalistic environments sometimes are much more extensive than the scenario simulated by the assessment instrument; and the model that is extracted from the collected data, inevitably, can only abstract and approximate the real situation. Nevertheless, the results of this study have built a foundation and provided possibilities for future research. In general, knowledge acquisition emphasizes understanding and representing the problem, while knowledge application emphasizes finding solutions (Greiff et al., 2013). Obviously, the processes of knowledge acquisition and knowledge application consist of complex mental and practical activities, which indicate possibilities for identifying lower-level dimensions within these two processes. Future research should focus on defining these sub-level component processes and further complete the construction of the problem-solving model.

This study proved that combinatorial reasoning and inductive reasoning have significant predictive effects on the problem-solving process for both Chinese and Hungarian students. Moreover, the results confirmed that combinatorial reasoning and inductive reasoning are strongly correlated, indicating that the development of problem-solving and other relevant reasoning skills are coordinated and not isolated. Currently, enhancing students' ability to solve problems has become one of the main targets in school education, and this can be realized by explicit training (Molnár, 2011) or by improving teaching methods (Shayer, & Adey, 2002). The findings of this study suggest that the problem-solving training program should be accompanied by training in specific reasoning skills. Furthermore, certain school subjects have the capacity to promote reasoning skills development (e.g. mathematics education: Primi, Ferrão, & Almeida, 2010; Xin, & Zhang, 2009; science education: Pásztor, & Csapó, 2014; Kambeyo, & Wu, 2018) and thus further contribute to problem-solving ability development. The results suggested that schools can improve instruction methods in these subjects by paying more attention to reasoning skills enhancement. Similarly, some non-thinking skill factors which have the power to influence students' problem-solving achievement (e.g. test-taking motivation, working memory) are also worth to be focused on by educational designers and psychologists.

Chinese and Hungarian students showed a statistically equal IPS performance in this study. One of the strongest advantages of Chinese students is still to be identified, namely that Chinese students were more willing to use theoretically effective exploration strategies, and they were better proficient strategy users. This study believes that this is an important reason why Chinese students could show a good performance in the PISA 2012 assessment. China is a country with the typical East Asian culture. Due to the pressure of huge populations, "hard working" is a tag for the East Asian students. In China, higher educational level, most of time, means higher income and social position. "Returns to education" is an important concept in sociology, which "calculated for investments undertaken as a part of initial education, and account for the main costs and benefits associated with this investment decision" (OECD, 2009, p. 152). It can also be described as the increase in the income from an additional year of education for an individual who makes the investment decision on education (Borjas, & Van Ours, 2010). The returns to education in China have sharply increased in the past decades, now it was much higher than the world average level (Wang, & Wu, 2018). This means that education plays an extremely important role in the Chinese culture. In this kind of environment, Chinese students are used to working harder, concentrate more, and had a higher motivation in the school activities. Thus, their attitude and habits inevitably influenced their behaviours in this study's assessment, as well as the PISA assessment. The finding that Chinese students' learning strategies had a much deeper connection with their problem-solving performance, could be seen as the evidence for this. They had higher motivation to explore the system (see section 5.5.2) and were more willing to employ and learn the effective strategies (see section 6.4), thus gained a certain advantage in the assessment. Although this kind of social environment is a unique feature in the East Asian countries, it could not and should not (it is not good for students' metal or physically healthy, Quach, Epstein, Riley, Falconier, & Fang, 2015) be copied by the other countries. However, it can still give a hint to Hungarian educators that there is a possibility to promote students' problem-solving performance by training students' daily learning habits and attitudes, and enhance their motivation in school activities.

7.3 Limitations

Apart from the limitations regarding sample size, measurement non-invariance, and different test-taking motivation which were mentioned in the previous sections (section 5.7 & 6.5), it may also cause concern about the generalizability of the findings that this study chose China and Hungary as the representatives for Asian and European culture, but. Some studies (e.g. Wüstenberg et al., 2014) have pointed out that students from different nations (even from similar cultural backgrounds, e.g. Hungary-Germany) could possibly have different levels of development in problem-solving performance, while the relationships between the components within problem-solving skills could also vary. A future study will continue to include more countries to discover such differences in the cognitive structures for problem solving that exist between students with Asian and European cultural backgrounds, and thus address the generalizability limitation of the present study.

7.4 Originality

I certify that the content of this dissertation is the product of my own work. This dissertation has not been previously submitted for a degree or at any other educational institution.

Publications related to the dissertation

Kambeyo, L., & Wu, H. (2018). A gondolkodási képességek számítógép-alapú mérésének lehetőségei Namíbiában és Kínában. In G. Hunyady, B. Csapó, G. Pusztay, & J. Szivák (Eds.), *Az oktatás korproblémái* (pp. 179–189). ELTE Eötvös Kiadó, Budapest.

Kambeyo, L., & Wu, H. (2018). Online assessment of students' inductive reasoning skills abilities in Oshana region, Namibia. *International Journal of Educational Sciences, 21*(1–3), 1–12.

Molnár, G., & Wu, H. (2018, August). *Cross-national gender differences in complex problem solving: A comparison study between China and Hungary.* Paper presented at 9th Biennial Conference of EARLI SIG1, Helsinki, Finland.

Wang, F., & Wu, H. (2018). Returns to education in rural and urban China: an empirical study. *Asian Journal of Social Science Studies, 3*(4), 18–28.

Wang, F., & Wu, H. (2018, November). *Returns to education in rural and urban China: an empirical study.* Paper presented at XVIII. National Conference on Education, Budapest, Hungary.

Worthington, T., & Wu, H. (2015). Time-shifted learning: merging synchronous and asynchronous techniques for e-learning. In *2015 10th International Conference on Computer Science & Education (ICCSE)* (pp. 434–437). Piscataway: IEEE.

Wu, H. (2017, May). *Measuring and analyzing the factors which impact student's problem solving skills.* Paper presented at 6th Interdisciplinary Doctoral Conference, Pecs, Hungary.

Wu, H., Kambeyo, L., & Molnár G. (2017, April). *The feasibility and reliability of computer-based assessment of inductive reasoning in Namibia and China: A comparison study.* Poster session presented at 15th Conference on Educational Assessment, Szeged, Hungary.

Wu, H., & Molnár, G. (2016, April). *The component skills of problem solving: A Chinese and Hungarian comparison study.* Paper presented at 14th Conference on Educational Assessment, Szeged, Hungary.

Wu, H., & Molnár, G. (2017). Measuring Chinese students' test-taking motivation in interactive problem solving environment. In *London International Conference on Education (LICE-2017)* (pp. 562–565). London: Infonomics Society.

Wu, H., & Molnár, G. (2017, April). *Chinese students' component skills of problem solving: A pilot study*. Paper presented at 15th Conference on Educational Assessment, Szeged, Hungary.

Wu, H., & Molnár, G. (2017, August). *Computer-based assessment of thinking skills in China: A pilot study.* Poster session presented at 17th biennial conference of the European Association for Research on Learning and Instruction (EARLI 2017). Tampere, Finland.

Wu, H., & Molnár, G. (2018). Computer-based assessment of Chinese students' component skills of problem solving: a pilot study. *International Journal of Information and Education Technology, 8*(5), 381–386.

Wu, H., & Molnár, G. (2018). Interactive problem solving: assessment and relations to combinatorial and inductive reasoning. *Journal of Psychological and Educational Research, 26*(1), 90–105.

Wu, H., & Molnár, G. (2018, April). *The role of inductive reasoning and combinatorial reasoning in complex problem solving: a comparison study between China and Hungary.* Paper presented at 16th Conference on Educational Assessment, Szeged, Hungary.

Wu, H., & Molnár, G. (2018, August). *Inductive reasoning and combinatorial reasoning as component skills of complex problem solving: Analyses in European and Asian context*. Paper presented at 9th Biennial Conference of EARLI SIG1, Helsinki, Finland.

Wu, H., & Molnár, G. (2019, April). *Cross-national differences in students' exploration strategies in a computer-simulated interactive problem-solving environment: logfile analyses*. Paper presented at Symposium at 17th Conference on Educational Assessment. Szeged, Hungary.

Wu, H., & Molnár, G. (2019, April). *The influence of learning strategies on* *problem-solving performance: analyses in the European and Asian contexts*. Paper presented at 17th Conference on Educational Assessment. Szeged, Hungary.

Wu, H., & Molnár, G. (2019, August). *The impact of learning strategies on problem-solving performance: a cross-national comparison study*. Paper presented at 18th Biennial EARLI Conference. Aachen, Germany.

Wu, H., Saleh, A. R., & Molnár, G. (2018, April). *Comparing Chinese and Indonesian students' level of thinking skills: a pilot study.* Paper presented at 16th Conference on Educational Assessment, Szeged, Hungary.

References

Abrami, P. C., Wade, A., Pillay, V., Aslan, O., Bures, E., & Bentley, C. (2008). Encouraging self-regulated learning through electronic portfolios. *Canadian Journal on Learning and Technology, 34*(3), 93–117.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs?. *Psychological Bulletin, 131*(1), 30.

Amabile, T. M. (2012). Componential theory of creativity. *Harvard Business School, 12*(96), 1–10.

Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 21*(9), 1201–1216.

Adey, P., & Csapó, B. (2012). Developing and assessing scientific reasoning. In B. Csapó & G. Szabó (Eds.), *Framework for diagnostic assessment of science* (pp. 17–53.) Budapest: Nemzeti Tankönyvkiadó.

Allaire, J. C., & Marsiske, M. (2002). Well-and ill-defined measures of everyday cognition: relationship to older adults' intellectual ability and functional status. *Psychology and aging, 17*(1), 101-115.

Anzai, K., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review, 86*, 124–140.

Arifin, S., Putri, R. I. I., Hartono, Y., & Susanti, E. (2017). Developing ill-defined problem-solving for the context of "South Sumatera". *Journal of physics: Conference series, 943*(1), 1–9.

Arroyo, I., & Woolf, B. P. (2005, May). Inferring learning and attitudes from a Bayesian Network of log file data. In P*roceedings of the 2005 Conference on Artificial Intelligence in Education* (pp. 33–40). Amsterdam: IAED.

Artelt, C., Baumert, J., Julius-McElvani, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning.* Results from PISA 2000. Paris: OECD.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics, 118(4),* 1279–1333.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baker, E. L., & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior, 15*(3–4), 269–282.

Batanero, C., Godino, J. D., & Navarro-Pelayo, V. (1997). Combinatorial reasoning and its assessment. In I. Gal, & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 239–252). IOS Press.

Basadur, M., Wakabayashi, M., & Graen, G. B. (1990). Individual problem-solving styles and attitudes toward divergent thinking before and after training. *Creativity Research Journal, 3*(1), 22–32.

Beckmann, J., & Goode, N. (2017). Missing the wood for the wrong trees: On the difficulty of defining the complexity of complex problem solving scenarios. *Journal of Intelligence, 5*(2), 15.

Bennett, R. E., Persky, H., Weiss, A. R., & Jenkins, F. (2007). *Problem solving in technology-rich environments. A report from the NAEP technology-based assessment project, research and development series*. NCES 2007-466. National Center for Education Statistics.

Bhaskar, R., & Simon, H. A. (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive Science, 1*(2), 193–215.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin & E. Care (Eds.). *Assessment and teaching of 21st century skills* (pp. 17–66). Springer Netherlands.

Bisanz, J., Bisanz, G., & Korpan, C. A. (1994). Inductive reasoning. In R. Sternberg (Ed.), *Thinking and problem solving* (pp. 181–213). San Diego: Academic Press.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Borjas, G. J., & Van Ours, J. C. (2010). *Labor economics*. Boston: McGraw-Hill/Irwin.

Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 27－63). NJ: Erlbaum, Hillsdale.

Bühner, M., Kröner, S., & Ziegler, M. (2008). Working memory, visual-spatial -intelligence and their relationship to problem-solving. *Intelligence, 36*(6), 672–680.

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

Cai, J. (2004). Why do U.S. and Chinese students think differently in mathematical problem solving? Exploring the impact of early algebra learning and teachers' beliefs. *Journal of Mathematical Behavior*, *23*(2), 135–167.

Cai, J., & Hwang, S. (2002). Generalized and generative thinking in U.S. and Chinese students' mathematical problem solving and problem posing. *Journal of Mathematical Behavior*. *21*(4), 401–421.

Cai, J., & Nie, B. (2007). Problem solving in Chinese mathematics education: Research and practice. *ZDM, 39*(5–6), 459–473.

Casakin, H. P. (2007). Factors of metaphors in design problem-solving: Implications for design creativity. *International Journal of Design, 1*(2), 21–33.

Care, E., & Griffin, P. (2017). Assessment of collaborative problem-solving processes. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 227–243). Paris: OECD Publishing.

Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology, 33*(4), 609–624.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis. With applications in the social, behavioral, and health sciences.* New York: Wiley.

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

Csapó, B. (1999). Improving thinking through the content of teaching. In J. H. M. Hamers, J. E. H. van Luit, & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 37–62). Lisse: Swets & Zeitlinger.

Csapó, B., Ainley, J., Bennett, R. E., Latour, T., & Law, N. (2012). Technological issues for computer-based assessment. In P. Griffin & E. Care (Eds.). (2014). *Assessment and teaching of 21st century skills* (pp. 143–230). Springer Netherlands.

Csapó, B., & Funke, J. (2017). Epilogue. In B. Csapó and J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 265–268). Paris: OECD Publishing.

Csapó, B., Lőrincz, A., & Molnár, G., (2012): Innovative assessment technologies in educational games designed for young students. In D. Eseryel & D. Ifenthaler (Eds.), *Assessment in game-based learning: foundations, innovations, and perspectives* (pp. 235–254). Springer. New York.

Csapó, B., & Molnár, G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology, 8*, 2022.

Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology, 106*(3), 639.

Csapó, B., Molnár, G., & Tóth, K. (2009). Comparing paper-and-pencil and online assessment of reasoning skills: A pilot study for introducing TAO in large-scale assessment in Hungary. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 113–118). Luxemburg: Office for Official Publications of the European Communities.

Davis, G. A. (1973). *Psychology of problem solving: Theory and practice*. New York: Basic Books.

Dede, C. (2010). Comparing frameworks for 21st century skills. *21st Century skills: Rethinking How Students Learn, 20*, 51–76.

Dörner, D. (1980) On the difficulties people have in dealing with complexity. *Simulation & Games*, *11*, 87 - 106.

Dörner, D., & Funke, J. (2017). Complex problem solving: what it is and what it is not. *Frontiers in psychology, 8*, 1153.

Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Hans Huber, Bern.

Dossey, J. A. (1993). *Can students do Mathematical problem solving? Results from constructed-response questions in NAEP's 1992 Mathematics assessment.* US Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402–9328.

Dossey, J. A., McCrone, S. A., & O'Sullivan, C. (2006). *Problem solving in the PISA and TIMSS 2003 assessments (NCES 2007-049).* U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Dubois, J. G. (1984). Une systematique des configurations combinatoires simples. *Educational Studies in Mathematics, 15* (1), 37-57.

Dubow, E. F., Boxer, P., & Huesmann, L. R. (2009). Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations. *Merrill-Palmer Quarterly (Wayne State University. Press), 55*(3), 224.

English, L. D. (2005). Combinatorics and the development of children ' s combinatorial reasoning. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 121–141). New York: Springer.

Ewert, P. H., & Lambert, J. F. (1932). Part II: The effect of verbal instructions upon the formation of a concept. *Journal of General Psychology, 6*, 400–411.

Finn, B. (2015). Measuring motivation in low-atakes assessments. *ETS Research Report Series, 2015*(2), 1–17.

Fiore, S. M., & Schooler, J. W. (1998). Right hemisphere contributions to creative problem solving: Converging evidence for divergent thinking. *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*, 349–371.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving, 4*(1), 19–42.

Fleischer, J., Buchwald, F., Leutner, D., Wirth, J., & Rumann, S. (2017). Analytical problem solving: Potentials and manifestations. *Educational Research and Innovation*, 33–46

Fraillon, J., Ainley, J., Gebhardt, E., & Schulz, W. (2013). *Measuring computer and information literacy across countries.* Paper presented at IEA International Research Conference, Singapore, June 26–28, 2013.

Frensch, P. A., & Funke, J. (1995). *Complex problem solving: The European perspective.* New York: Psychology Press.

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *The Journal of Neuroscience, 29*(27), 8726–8733.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). New York: Psychology Press.

Funke, J. (1992). Dealing with dynamic systems: Research strategy, diagnostic approach and experimental results. *German Journal of Psychology, 16*, 24–43.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning*, *7*(1), 69–89.

Funke, J. (2014). Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in Psychology, 5*, 739.

Gardner, P. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied cognitive psychology, 9*(7), 55 – 79.

Greiff, S., Fischer, A., Stadler, M., & Wüstenberg, S. (2014). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning, 21*(3), 1–27.

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A multitrait-multimethod study of assessment instruments for complex problem solving. *Intelligence, 41*(5), 579–596.

Greiff, S., & Funke, J. (2017). Interactive problem solving: Exploring the potential of minimal complex systems. In B. Csapó and J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 93–105). Paris: OECD Publishing.

Greiff, S., Holt, D., & Funke, J. (2013). Perspectives on problem solving in cognitive research and educational assessment: analytical, interactive, and collaborative problem solving. *Journal of Problem Solving (The), 5*, 71–91.

Greiff, S., Kretzschmar, A., Müller, J., Spinath, B., & Martin, R. (2014). Computer-based assessment of Complex Problem Solving in educational contexts and how it is influenced by students' level of Information and Communication Technology literacy. *Journal of Educational Psychology, 106*(3), 666–680.

Greiff, S., Krkovic, K., & Hautamäki, J. (2015). The prediction of problem-solving assessed via microworlds. *European Journal of Psychological Assessment, 32*, 298–306.

Greiff, S., Molnár, G., Martin, R., Zimmermann, J., & Csapó, B. (2018). Students' exploration strategies in computer-simulated complex problem environments: A latent class approach. *Computers & Education, 126*, 248–263.

Greiff, S., & Wüstenberg, S. (2014). Assessment with microworlds: Factor structure, invariance, and latent mean comparison of the MicroDYN test. *European Journal of Psychological Assessment, 30*, 1–11.

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education, 91*, 92–105.

Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review, 13,* 74–83.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement, 36*(3), 189–213.

Greiff, S., Wüstenberg, S., Goetz, T., Vainikainen, M. P., Hautamäki, J., & Bornstein, M. H. (2015). A longitudinal study of higher-order thinking skills: working memory and fluid reasoning in childhood enhance complex problem solving in adolescence. *Frontiers in psychology, 6*, 1060.

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts-Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology, 105*(2), 364–379.

Gilhooly, K. J. (1982). *Thinking: Directed, undirected and creative.* London: Academic Press.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance, 16*(2), 250–279.

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

Hayes, J. (1980). *The complete problem solver*. Philadelphia: Franklin Institute Press.

Herrmann, N. (1995). Creative problem solving. *IEEE Potentials Magazine*, 4–9.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37–56). Springer Netherlands.

Hong, N. S. (1998). *The relationship between well-structured and ill-structured problem solving in multimedia simulation.* Doctoral dissertation, Pennsylvania State University.

Hu, L. (2014). Wo men de jiao yu neng geng zi xin dian ma?. *Ren Min Jiao Yu, 3*, 6-6. (In Chinese)

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829–6833.

Kambeyo, L., & Wu, H. (2018). Online assessment of students' inductive reasoning skills abilities in Oshana Region, Namibia. *International Journal of Educational Sciences, 21*(1–3), 1–12.

Kay, D. S. (1991). Computer interaction: Debugging the problems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 317–340). Hillsdale, NJ: Lawrence Erlbaum Associates.

Klauer, K. J. (1990). Paradigmatic teaching of inductive thinking. *Learning and Instruction, 2*, 23–45.

Klauer, K. J. (1996). Teaching inductive reasoning: Some theory and three experimental studies. *Learning and Instruction, 6*(1), 37-57.

Knekta, E. (2017). *Motivational aspects of test-taking: measuring test-taking motivation in Swedish national test contexts.* Doctoral dissertation, Umeå universitet.

Krkovic, K., Pásztor-Kovács, A., Molnár, G., & Greiff, S. (2014). New technologies in psychological assessment: The example of computer-based collaborative problem solving assessment. *International Journal of e-Assessment, 4*(1), 28–41.

Krkovic, K., Wüstenberg, S., & Greiff, S. (2016). Assessing collaborative behavior in Students. *European Journal of Psychological Assessment, 32*(1), 52–60.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence, 14*(4), 389–433.

Kyttälä, M., Aunio, P., Lepola, J., & Hautamäki, J. (2014). The role of the working memory and language skills in the prediction of word problem solving in 4-to 7-year-old children. *Educational Psychology, 34*(6), 674–696.

Lau, S., & Cheung, P. C. (2010). Creativity assessment: Comparability of the electronic and paper-and-pencil versions of the Wallach-Kogan Creativity Tests. *Thinking Skills and Creativity, 5*(3), 101–107.

Li, Q. (2010). Shu xue wen ti jie jue de biao zheng ji yuan ren zhi kai fa. *Jiao Shu Yu Ren, 30*, 34-34. (In Chinese)

Liu, D. (2001). Wen ti shi xue xi: Yi tiao ji zhong ti xian jian gou zhu yi si xiang de jiao xue gai ge si lu. *Jiao Yu Li Lun Yu Shi Jian*, 21(5), 53–56. (In Chinese)

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88* (3), 767–778.

Lotz, C., Scherer, R., Greiff, S., & Sparfeldt, J. R. (2017). Intelligence in action – Effective strategic behaviors while solving complex problems. *Intelligence, 64*, 98–112.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26* (1-2), 49–63.

Mislevy, R. J. (1995). Test theory and language-learning assessment. *Language Testing, 12*(3), 341–369.

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

Molnár, G. (2012). A problémamegoldó gondolkodás fejlődése: az intelligencia és szocioökonómiai háttér befolyásoló hatása 3 – 11. évfolyamon. *Magyar Pedagógia, 112*(1), 41–58.

Molnár, G. (2014). Development of dynamic problem solving from grade 3 to 9. In *12th Conference on Educational Assessment: Book of Abstract* (pp. 103). Szeged.

Molnár, G. (2015). A képességmérés dilemmái: A diagnosztikus mérések (eDia) szerepe és helye a magyar közoktatásban. *Géniusz Műhely Kiadványok*, *2*, 16–29.

Molnár, G. (2016). *Technology-based assessment in education: Assessing the development of problem solving*. DSc Thesis. Retrieved from http://real-d.mtak.hu/920/7/dc_968_14_doktori_mu.pdf

Molnár, G., & Csapó, B. (2011). Az 1–11 évfolyamot átfogó induktív gondolkodás kompetenciaskála készítése a valószínűségi tesztelmélet alkalmazásával. *Magyar Pedagógia, 111*(2), 127–140.

Molnár, G. & Csapó, B. (2013). Az eDia online diagnosztikus mérési rendszer. In *11th Conference on Educational Assessment: Book of Abstract* (pp. 82). Szeged.

Molnár, G., & Csapó, B. (2017). Exploration and learning strategies in an interactive problem-solving environment at the beginning of higher education studies. In: J. C. Spender, T. Gavrilova, & G. Schiuma (Eds.), *Knowledge management in the 21st century: Resilience, creativity and co-creation. Proceedings IFKAD2017* (pp. 283–292). St Petersburg University, St. Petersburg.

Molnár, G., & Csapó, B. (2018). The efficacy and development of students' problem-solving strategies during compulsory schooling: Logfile analyses. *Frontiers in Psychology, 9*, 302.

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, 9, 35–45.

Molnár, G., Greiff, S., Wüstenberg, S. & Fischer, A. (2017). Empirical study of computer based assessment of domain-general dynamic problem solving skills. In B. Csapó and J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 123–143). Paris: OECD Publishing.

Moray, N., Lootsteen, P., & Pajak, J. (1986). Acquisition of process control skills. *IEEE Transactions on Systems, Man, and Cybernetics, 16*(4), 497–504.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Neubert, J. C., Kretzschmar, A., Wüstenberg, S., & Greiff, S. (2014). Extending the Assessment of Complex Problem Solving to Finite State Automata: Embracing Heterogeneity. *European Journal of Psychological Assessment, 31*(3), 181–194.

Newell, A. (1993). *Reasoning, problem solving, and decision processes: The problem space as a fundamental category.* Boston: MIT Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends in Cognitive Sciences, 9*(10), 467–473.

Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 321–49). New York: Cambridge University Press.

Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity-facets of a cognitive ability construct. *Personality and Individual Differences, 29*(6), 1017–1045.

OECD (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills.* Paris: OECD Publishing.

OECD (2004). *Problem solving for tomorrow's world: First measures of cross-curricular competencies from PISA 2003.* Paris: OECD Publishing.

OECD. (2009). *Education at a glance 2009 OECD indicators: OECD indicators*. OECD Publishing.

OECD (2010). *PISA 2012 field trial problem solving framework.* Paris: OECD Publishing.

OECD (2013). *PISA 2015 Draft collaborative problem solving framework.* Paris: OECD Publishing.

OECD (2014a). *Results: Creative problem solving – Students' skills in tackling real-life problems. Volume V.* Paris: OECD Publishing.

OECD (2014b). *PISA 2012 Results: What students know and can do. Student performance in Mathematics, Reading and Science (Volume I).* Paris: OECD Publishing.

OECD. (2017a). *PISA 2015 Collaborative problem-solving framework*. Paris: OECD Publishing.

OECD. (2017b). *PISA 2015 results: Collaborative problem solving (Volume V)*. Paris: OECD Publishing.

Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. *Journal of Psychoeducational Assessment, 9*(1), 45–53.

Pachler, N., Daly, C., Mor, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education, 54*(3), 715–721.

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology, 80*(1), 44–57.

Pásztor, A., & Csapó, B. (2014). *Improving combinatorial reasoning through inquiry-based science learning.* Paper presented at Science and Mathematics Education Conference, June 24-25, Dublin City University, Dublin, Ireland.

Pásztor, A., Hotulainen, R., Kupiainen, S., Molnár, G. & Csapó, B. (2018): *Comparing Finnish and Hungarian fourth grade students' inductive reasoning skills*. Paper presented at EARLI SIG 1 Conference, Helsinki.

Pásztor, A., Molnár, G., Korom, E. B., Németh, M., & Csapó, B. (2017). Online assessment of inductive reasoning and its predictive power on inquiry skills in science. *17th biennial conference of the European Association for Research on Learning and Instruction (EARLI)* (p. 509). Tampere, Finland, August 29-September 2, 2017.

Pásztor, A., Molnár, G., & Csapó, B. (2015). Technology-based assessment of creativity in educational context: the case of divergent thinking and its relation to mathematical achievement. *Thinking Skills and Creativity, 18*, 32–42.

Pásztor-Kovács, A. (2018). *The assessment of collaborative problem solving skills*. Doctoral dissertation, University of Szeged.

Pellegrino, J. W., & Glaser, R. (1982). Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Eds.), *Advances in instructional psychology, Vol. 2* (pp. 269–345). Hillsdale, NJ: Erlbaum.

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. The University of Michigan, Ann Arbor, MI.

Piffer, D. (2012). Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. *Thinking Skills and Creativity, 7*(3), 258–264.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*(2), 83–96.

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences, 20*(5), 446–451.

Quach, A. S., Epstein, N. B., Riley, P. J., Falconier, M. K., & Fang, X. (2015). Effects of parental warmth and academic pressure on anxiety and depression symptoms in Chinese adolescents. *Journal of Child and Family Studies, 24*(1), 106–116.

Ramalingam, D., Philpot, R., & McCrae, B. (2017). The PISA 2012 assessment of problem solving. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 75–91). Paris: OECD Publishing.

Raven, J. (2000). Psychometrics, cognitive ability, and occupational performance, *Review of Psychology, 7* (1-2), 51–74.

Riding, R., & Rayner, S. (2013). *Cognitive styles and learning strategies: Understanding style differences in learning and behavior.* London: David Fulton Publishers.

Robinson, K. (2009). *The element: How finding your passion changes everything.* New York: Penguin Books.

Rosen, Y., & Tager, M. (2013). *Computer-based assessment of collaborative problem-solving skills: Human-to-agent versus human-to-human approach.* Research & Innovation Network, Pearson Education.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96.

Sandberg, E. H., & McCullough, M. B. (2010). The development of reasoning skills. In E. H. Sandberg & B. L. Spritz (Eds.), *A clinician's guide to normal cognitive development in childhood* (pp. 179-189). New York: Routledge

Schleicher, A. (2017). Foreword. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 3–4). Paris: OECD Publishing.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312.

Sewell, W. H., & Shah, V. P. (1968). Parents' education and children's educational aspirations and achievements. *American Sociological Review*, 191–209.

Shayer, M., & Adey, P. (Eds.). (2002). *Learning intelligence: Cognitive acceleration across the curriculum from 5 to 15 years.* Milton Keynes: Open University Press.

Simonton, D. K. (2012). Taking the US patent office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity Research Journal, 24*(2–3), 97–106.

Sokol, S. M., & McCloskey, M. (1991). Cognitive mechanisms in calculation. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 85–116). Hillsdale, NJ: Lawrence Erlbaum Associates.

Söderqvist, S., Bergman Nutley, S., Ottersen, J., Grill, K. M., & Klingberg, T. (2012). Computerized training of non-verbal reasoning and working memory in children with intellectual disability. *Frontiers in Human Neuroscience, 6*, 271.

Stanovich, K. E., & Cunningham, A. E. (1991). Reading as constrained reasoning. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 3–60). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sternberg, R. J. (2013). *Thinking and problem solving (Vol. 2)*. Academic Press.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96*(3), 471.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Tang, X, & Ren, Y. (2006). Ji yu wen ti shi xue xi jiao xue mo shi de tan jiu yu shi jian. *Jiao Xue Yan Jiu (Hebei)*, (1), 24–25. (In Chinese)

Thimbleby, H. (2007). *Press on: Principles of interaction programming.* MIT Press, Cambridge, MA.

Tóth, K., Rölke, H., Goldhammer, F., & Barkow, I. (2017). Educational process mining: New possibilities for understanding students' problem-solving skills. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 193–209). Paris: OECD Publishing.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. (2005). Uncovering expertise‐related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*(2), 205–221.

VanLehn, K. (1989). Discovering problem solving strategies: What humans do and machines don ’ t (yet). In A. M. Segre (Ed.), *Proceedings of the Sixth International Workshop on Machine* (pp. 215-217). Ithaca, NY: Cornell University Press.

Tomic, W. (1995). Training in inductive reasoning and problem solving. *Contemporary Educational Psychology, 20*(4), 483–490.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*, 75–100.

Vroom, V. H. (1964). *Work and motivation*. NY: John Wiley & sons.

Vygotsky, L. S. (1978). *Mind and Society: The development of higher psychological processes*. Cambridge, MA: Harvard University press.

Wang, F., & Wu, H. (2018). Returns to education in rural and urban China: An empirical study. *Asian Journal of Social Science Studies, 3*(4), 18–28.

Wang, Z. & Zhang, T. (2003). *Hua xue wen ti she ji yu yan jiu (Vol. 8)*. Higher Education Press (In Chinese).

Willoughby,T. & Wood, E. (1994). Elaborative interrogation examined at encoding and retrieval. *Learning and Instruction, 4*, 139-149.

Wittmann, W., & Hattrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science, 21* (4), 393–409.

Wittmann, W., & Süß, H. M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, traits, and content determinants* (pp. 77–108). Washington, DC: APA.

Woods, D. T., Crow, C. M., Hoffman, T. W., & Wright, J. D. (1985). Challenges to teaching problem solving skills. *Chem 13 News, 155*, 1–12. Waterloo, Ontario, Canada: University of Waterloo.

Worthington, T., & Wu, H. (2015). Time-shifted learning: Merging synchronous and asynchronous techniques for e-learning. In *2015 10th International Conference on Computer Science & Education (ICCSE)* (pp. 434–437). Piscataway: IEEE.

Wu, H., & Molnár G. (2018): Computer-based assessment of Chinese students' component skills of problem solving: A pilot study. *International Journal of Information and Education Technology, 8*(5), 381–386.

Wu, M., Raymond, J., Wilson, M., & Haldane, S. (2003). *ConQuest*, Australia: Australian Council or Educational Research.

Wüstenberg, S. (2013). *Nature and validity of complex problem solving.* Doctoral dissertation, Universität Heidelberg.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving – More than reasoning?. *Intelligence, 40*(1), 1–14.

Wüstenberg, S., Greiff, S., Molnár, G., & Funke, J. (2014). Cross-national gender differences in complex problem solving and their determinants. *Learning & Individual Differences, 29*, 18–29.

Xie, W., & Liu, S. (2015). Shanghai PISA Yin qi de guo ji fan xiang- Lai zi ou mei deng guo de zui xin dong tai ji qi si kao. *Jiao Yu Ce Liang Yu Ping Jia: Li Lun Ban*, (7), 20–24. (In Chinese)

Xin, Z., & Zhang, L. (2009). Cognitive holding power, fluid intelligence, and mathematical achievement as predictors of children's realistic problem solving. *Learning and Individual Differences, 19*(1), 124–129.

Yang, D. (2014). PISA Shanghai di yi de si kao. *Hebei Educaiton*, (2). (In Chinese)

Zhang, J. (2000). Ji yu wen ti jie jue de zhi shi jian gou. *Jiao Yu Yan Jiu, 10*, 58–62. (In Chinese)

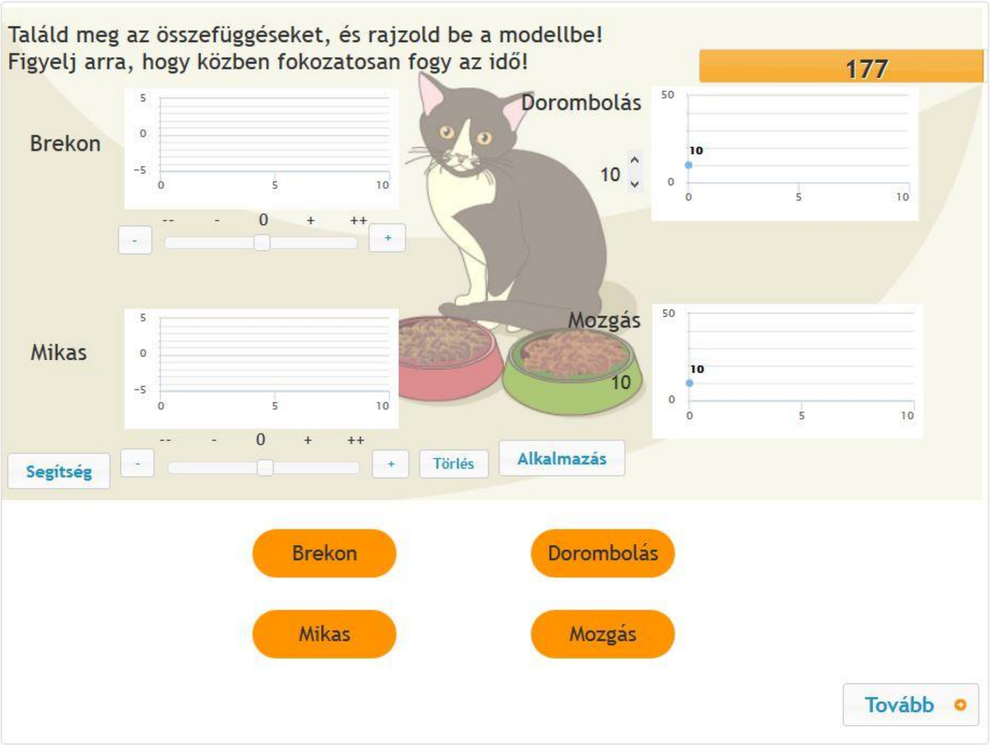
Zhu, L. (1997). Yi Xue jiao yu zhong de wen ti jie jue shi xue xi. *Fu Dan Jiao Yu Lun Tan*, (1), 25–26. (In Chinese)

Zoanetti, N., & Griffin, P. (2017). Log-file data as indicators for problem-solving processes. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 177–191). Paris: OECD Publishing.

Appendixes

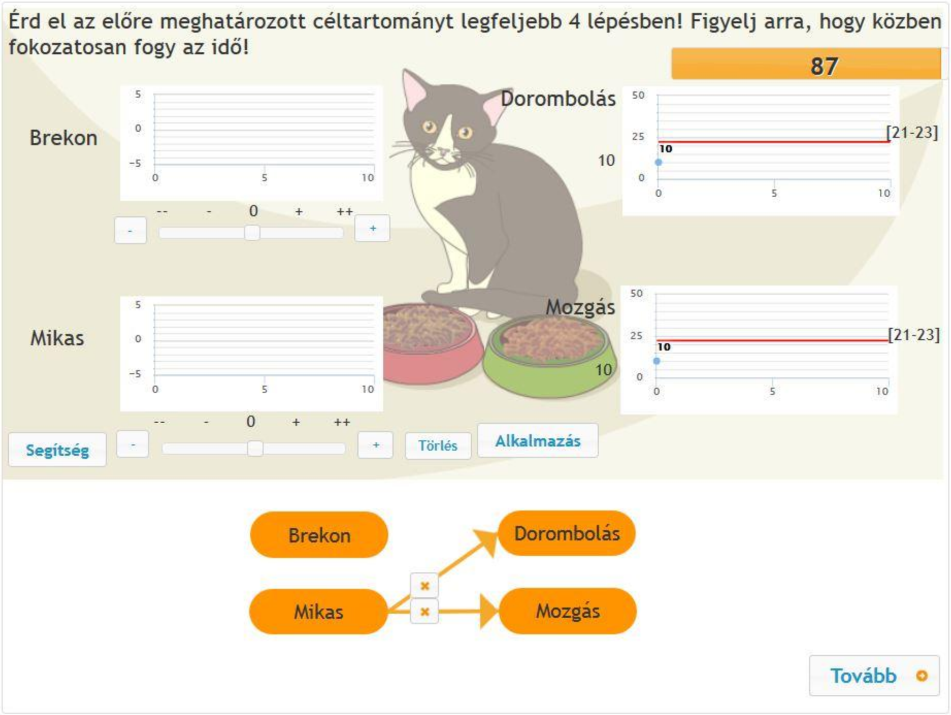
Appendix A: The Hungarian and the Chinese version of the IPS problem Cat

Hungarian Version



Találd meg az összefüggéseket, és rajzold be a modellbe!
Figyelj arra, hogy közben fokozatosan fogy az idő!

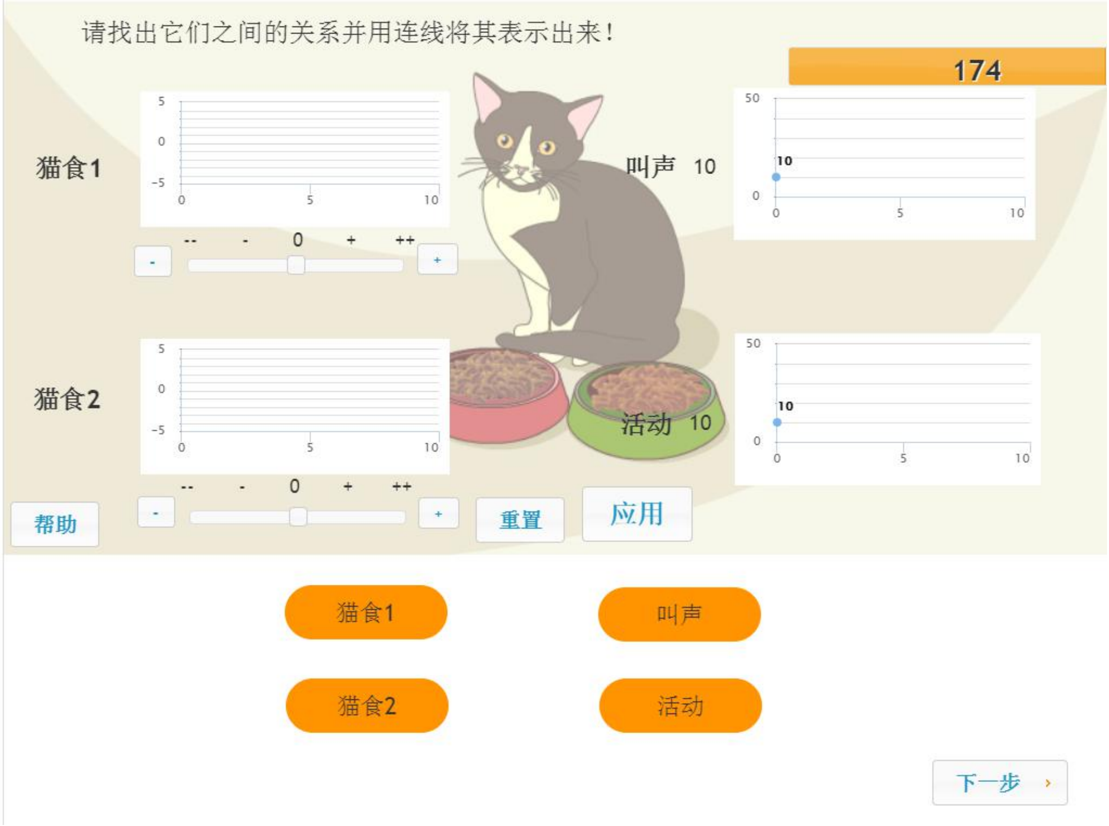
177
Brekon
5
0
-5
0
5
10
-- - 0 + ++
Dorombolás
10
50
0
10
0
5
10
Mikas
5
0
-5
0
5
10
-- - 0 + ++
Mozgás
10
50
0
10
0
5
10
Segítség
Törlés
Alkalmazás
Brekon
Dorombolás
Mikas
Mozgás
Tovább



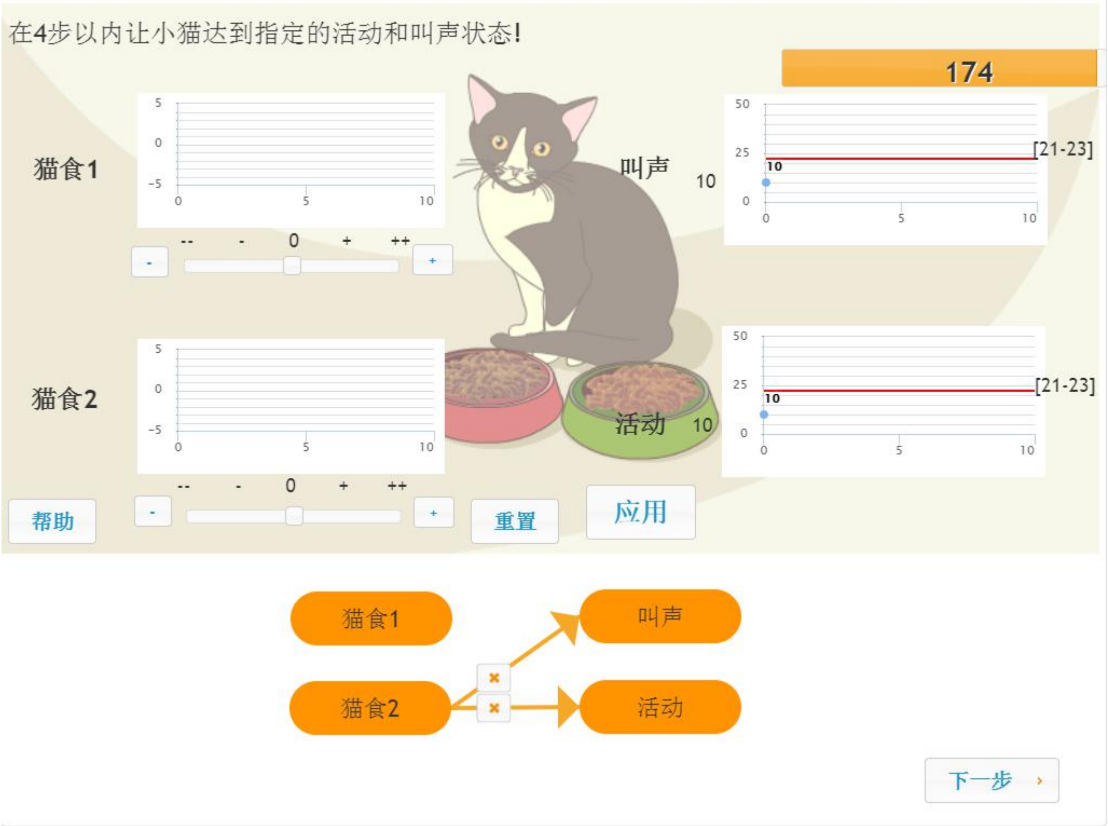
Érd el az előre meghatározott céltartományt legfeljebb 4 lépésben! Figyelj arra, hogy közben fokozatosan fogy az idő!

87
Brekon
5
0
-5
0
5
10
-- - 0 + ++
Dorombolás
10
50
25
0
[21-23]
10
0
5
10
Mikas
5
0
-5
0
5
10
-- - 0 + ++
Mozgás
10
50
25
0
[21-23]
10
0
5
10
Segítség
Törlés
Alkalmazás
Brekon
Dorombolás
Mikas
Mozgás
Tovább

Chinese Version


请找出它们之间的关系并用连线将其表示出来!

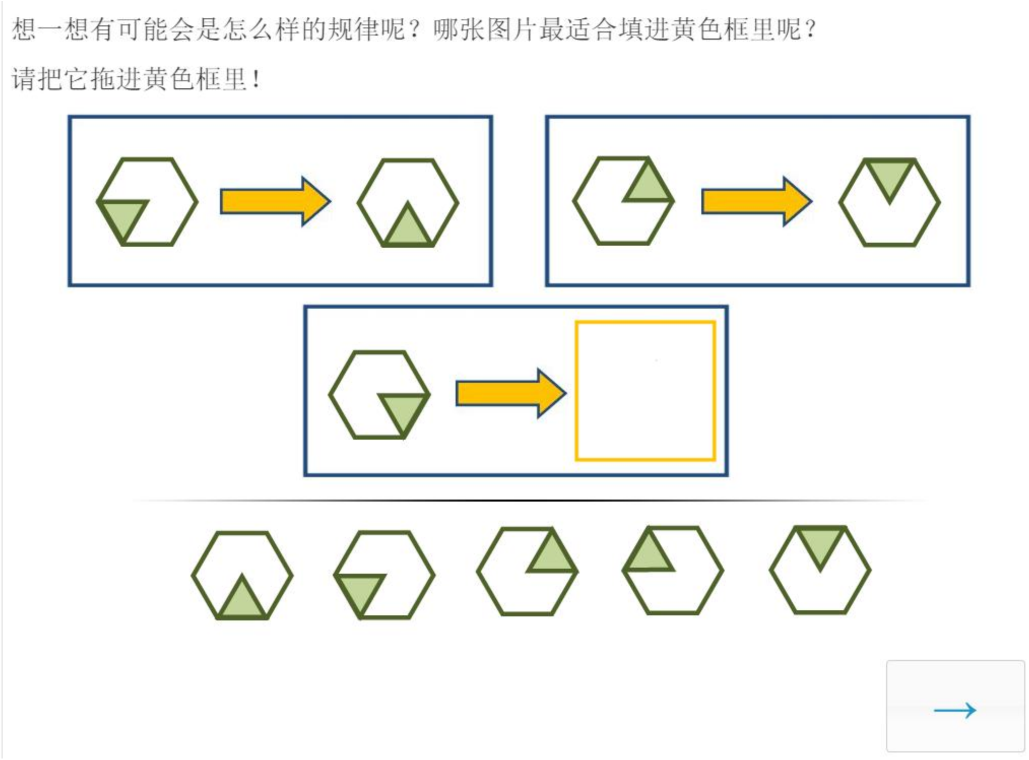
174
猫食1
猫食2
叫声 10
活动 10
5
0
-5
50
10
-- - 0 + ++
-
+
帮助
重置
应用
猫食1
叫声
猫食2
活动
下一步 ›


在4步以内让小猫达到指定的活动和叫声状态!

174
猫食1
猫食2
叫声 10
活动 10
50
25
0
10
[21-23]
-- - 0 + ++
帮助
重置
应用
猫食1
叫声
猫食2
活动
×
下一步 ›

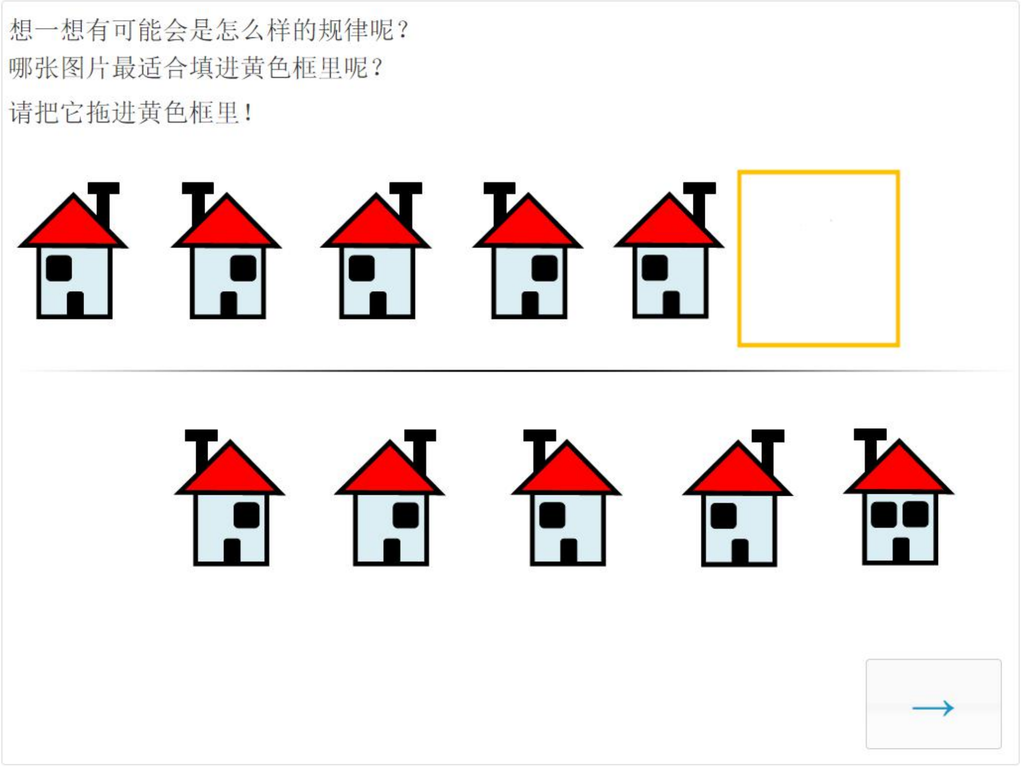
Appendix B: Sample items from the Chinese version of the inductive reasoning test

Figural Analogy Items



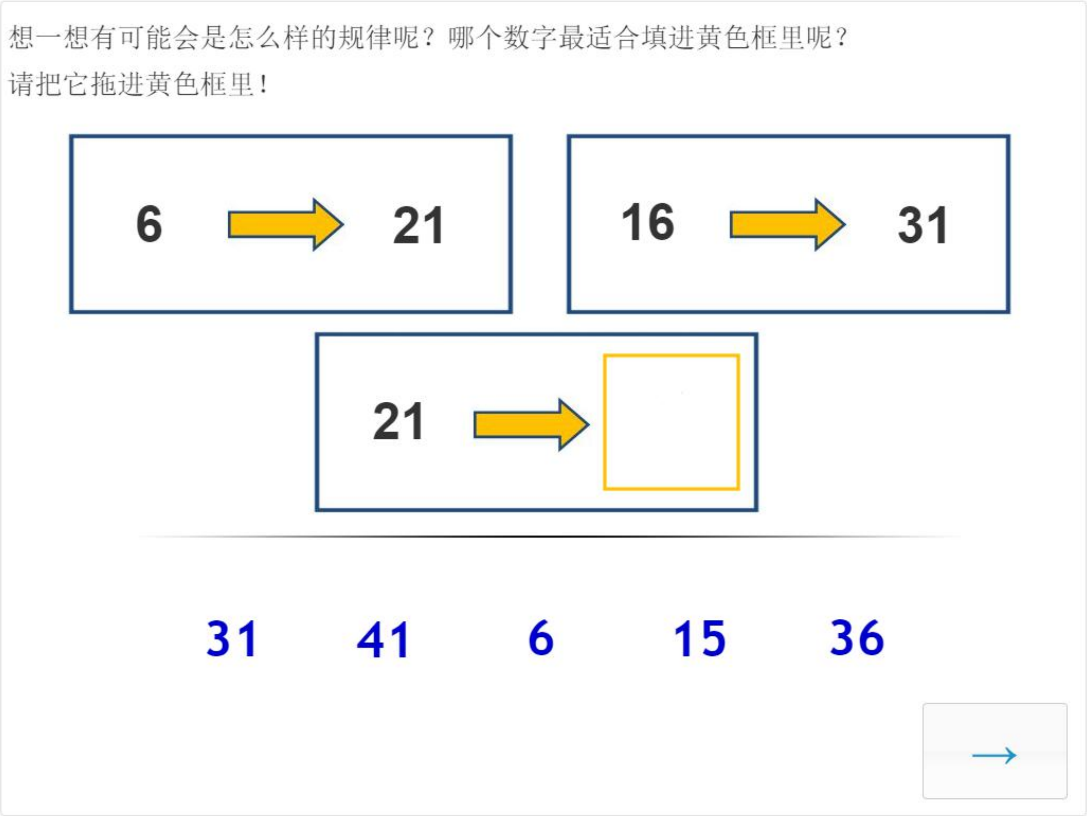
想一想有可能会是怎么样的规律呢？哪张图片最适合填进黄色框里呢？
请把它拖进黄色框里！

Figural Series Items



想一想有可能会是怎么样的规律呢？
哪张图片最适合填进黄色框里呢？
请把它拖进黄色框里！

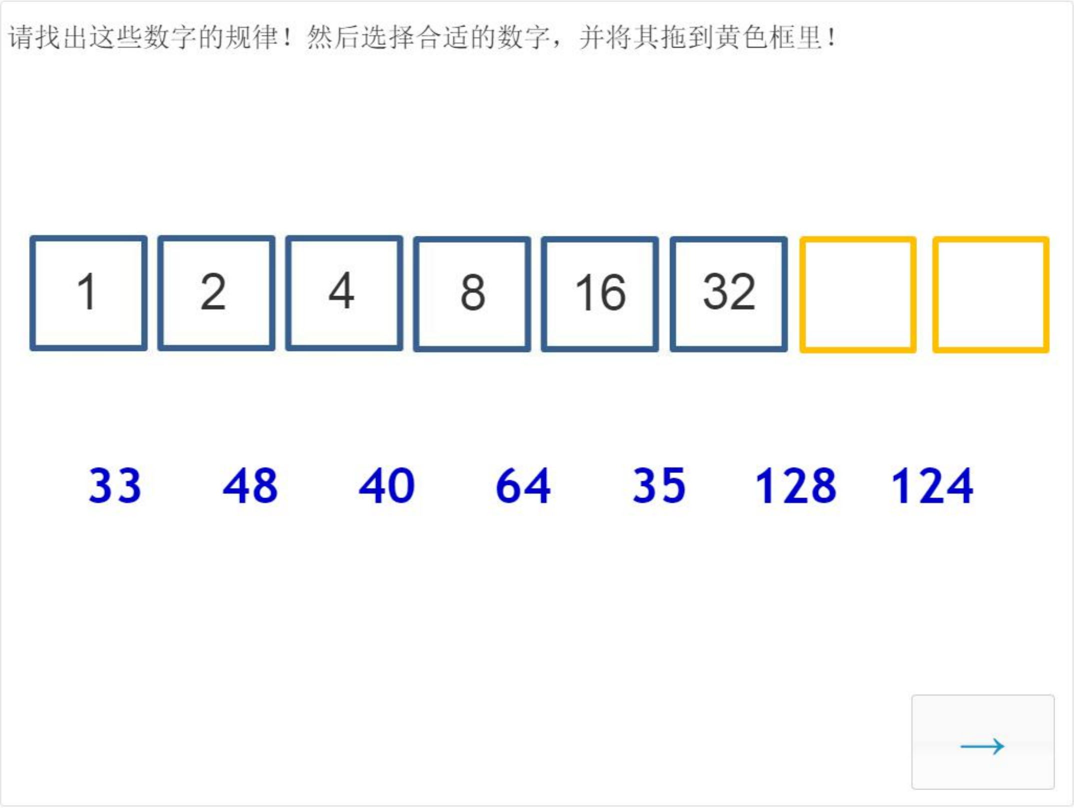
Number Analogy Items



想一想有可能会是怎么样的规律呢？哪个数字最适合填进黄色框里呢？
请把它拖进黄色框里！

6
21
16
31
21
31
41
6
15
36
→

Number Series Items

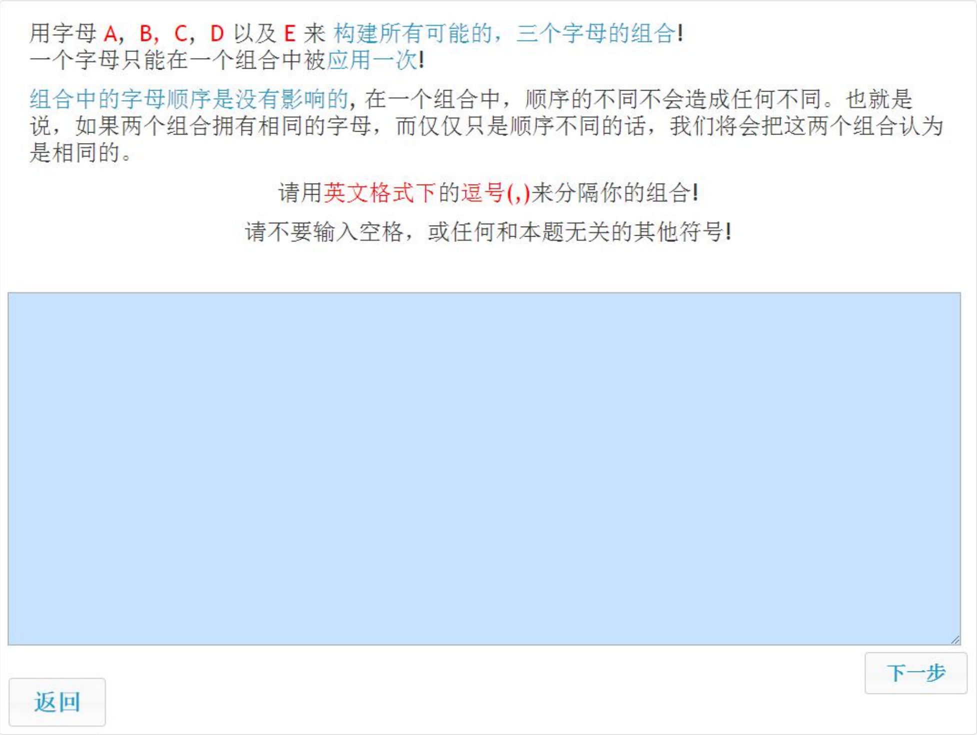


请找出这些数字的规律！然后选择合适的数字，并将其拖到黄色框里！

1
2
4
8
16
32
33
48
40
64
35
128
124
→

Appendix C: Sample items from the Chinese version of the combinatorial reasoning test

Verbal Items



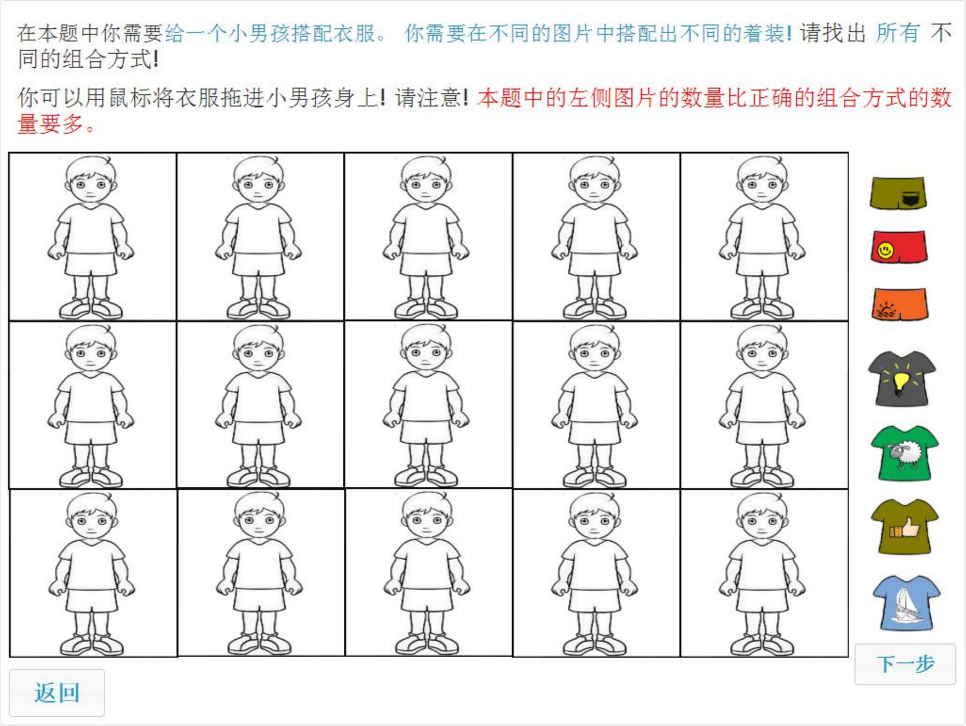
用字母 A，B，C，D 以及 E 来 构建所有可能的，三个字母的组合!
一个字母只能在一个组合中被应用一次!

组合中的字母顺序是没有影响的，在一个组合中，顺序的不同不会造成任何不同。也就是说，如果两个组合拥有相同的字母，而仅仅只是顺序不同的话，我们将会把这两个组合认为是相同的。

请用英文格式下的逗号(,)来分隔你的组合!

请不要输入空格，或任何和本题无关的其他符号!

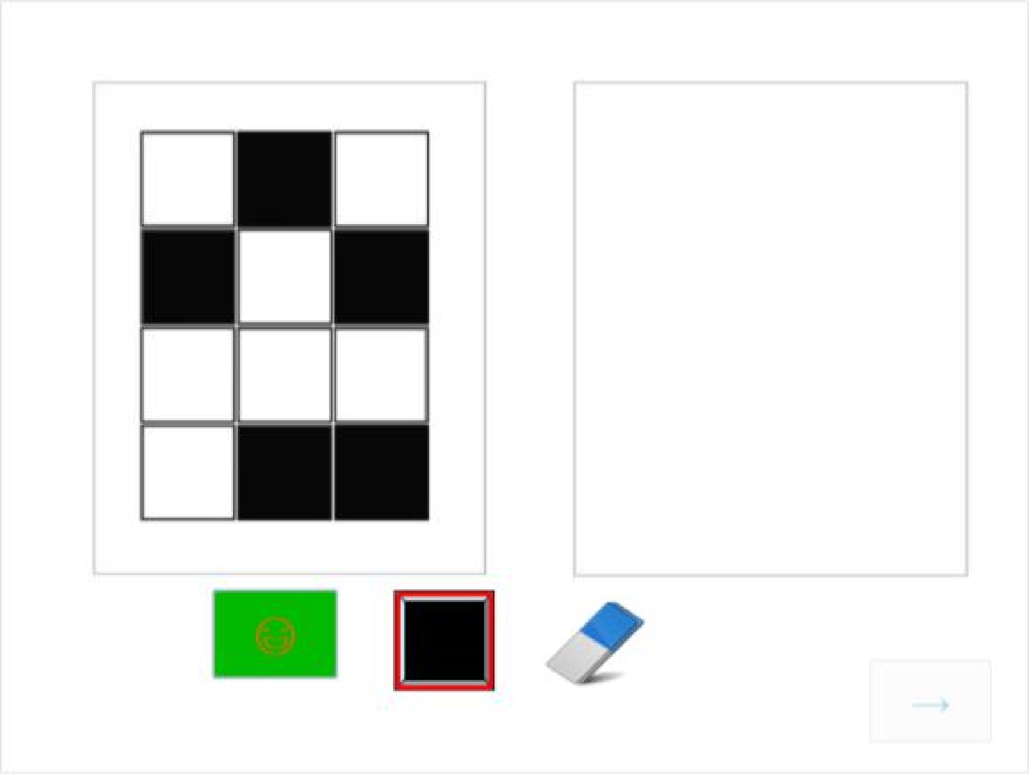
Figural Items

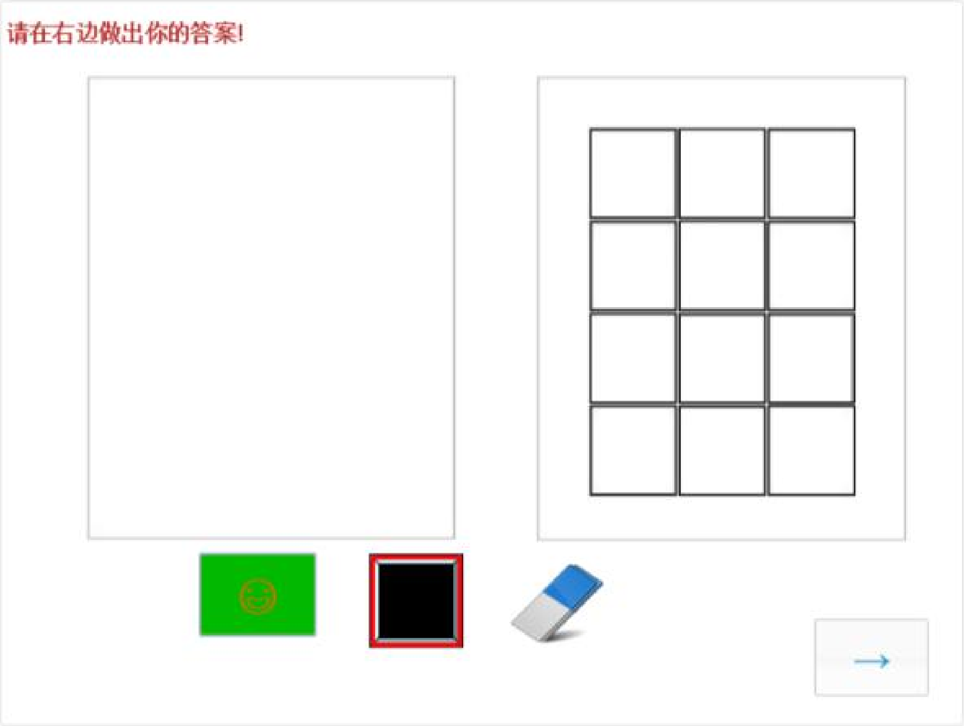


在本题中你需要给一个小男孩搭配衣服。 你需要在不同的图片中搭配出不同的着装! 请找出 所有 不同的组合方式!

你可以用鼠标将衣服拖进小男孩身上! 请注意! 本题中的左侧图片的数量比正确的组合方式的数量要多。

Appendix D: Sample item from the Chinese version of the working memory test

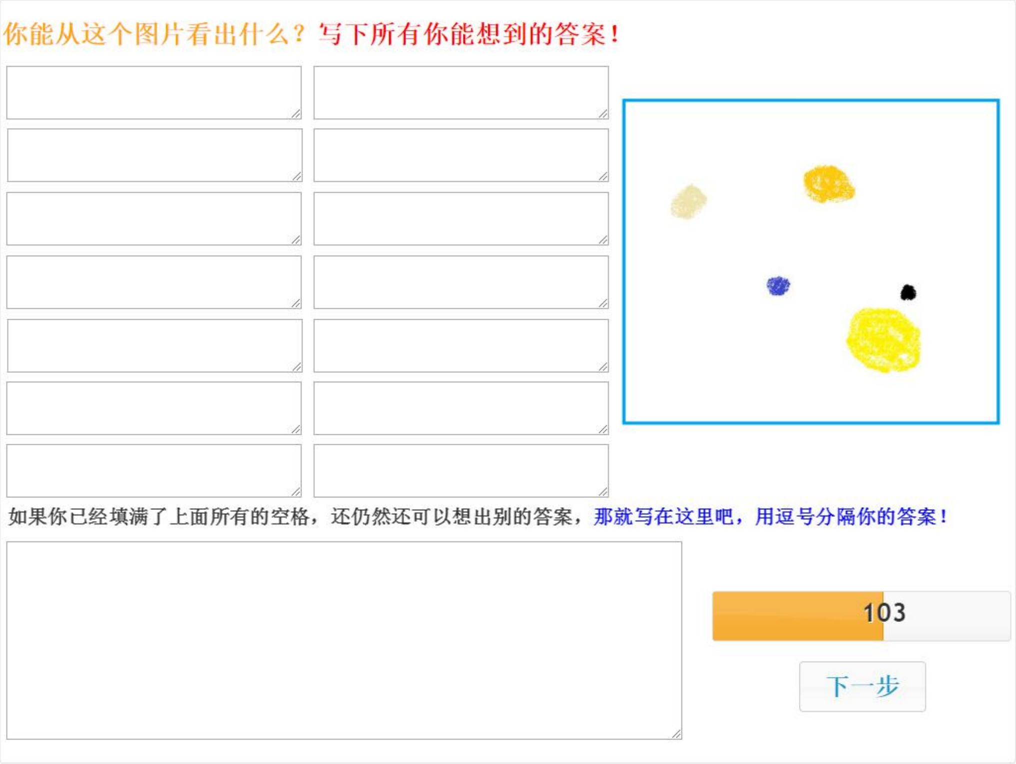




请在右边做出你的答案!

Appendix E: Sample item from the Chinese version of the creativity test

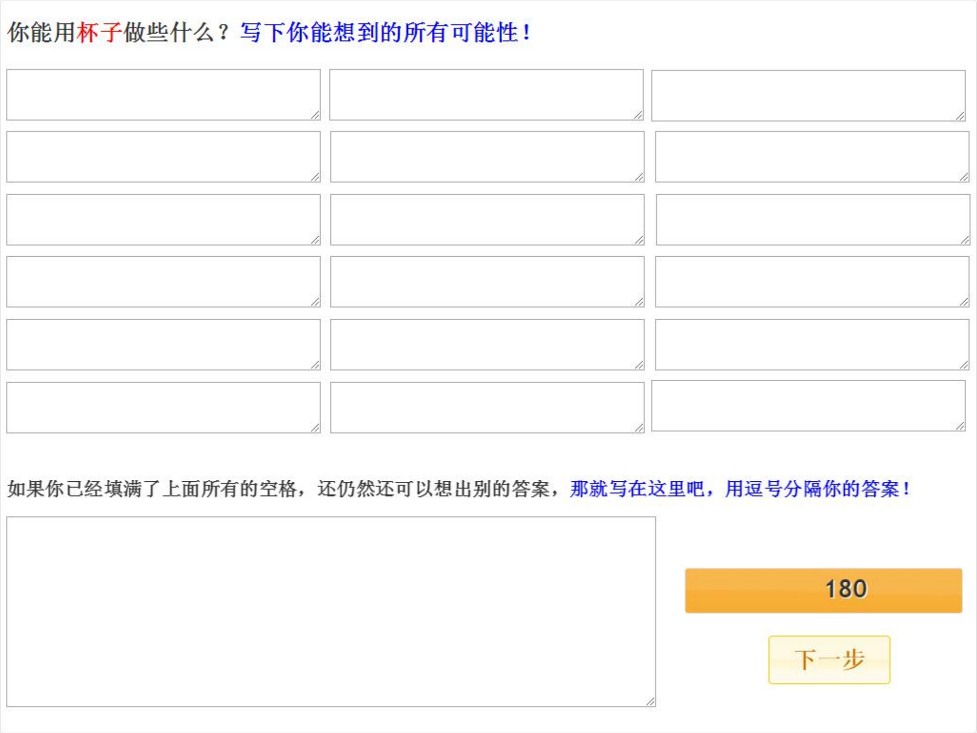
Figural-verbal Creativity Items



你能从这个图片看出什么？写下所有你能想到的答案！

如果你已经填满了上面所有的空格，还仍然还可以想出别的答案，那就写在这里吧，用逗号分隔你的答案！

Verbal-verbal Creativity Items



你能用杯子做些什么？写下你能想到的所有可能性！

如果你已经填满了上面所有的空格，还仍然还可以想出别的答案，那就写在这里吧，用逗号分隔你的答案！

Appendix F: Background questionnaire (in English)

Are you a ____?

Boy

Girl

How old are you?

____

What is your mother's educational level?

Primary School

Junior or Senior High School

Vocational School

Undergraduate Degree

Postgraduate Degree

I Don't Know

What is your father's educational level?

Primary School

Junior or Senior High School

Vocational School

Undergraduate Degree

Postgraduate Degree

I Don't Know

Elaboration Strategies

When I study, I try to relate new material to things I have learned in other subjects

Never Rarely Sometimes Often Always

When I study, I figure out how the information might be useful in the real world

Never Rarely Sometimes Often Always

When I study, I try to understand the material better by relating it to things I already know

Never Rarely Sometimes Often Always

When I study, I figure out how the material fits in with what I have learned

Never Rarely Sometimes Often Always

Memorisation strategies

When I study, I try to memorise everything that might be covered

Never Rarely Sometimes Often Always

When I study, I memorise as much as possible

Never Rarely Sometimes Often Always

When I study, I memorise all new material so that I can recite it

Never Rarely Sometimes Often Always

When I study, I practice by saying the material to myself over and over

Never Rarely Sometimes Often Always

Control strategies

When I study, I start by figuring out what exactly I need to learn

Never Rarely Sometimes Often Always

When I study, I force myself to check to see if I remember what I have learned

Never Rarely Sometimes Often Always

When I study, I try to figure out, as I read, which concepts I still haven't really understood

Never Rarely Sometimes Often Always

When I study, I make sure that I remember the most important things

Never Rarely Sometimes Often Always

When I study, and I don't understand something, I look for additional information to clarify the point

Never Rarely Sometimes Often Always