



Inductive and combinatorial reasoning in international educational context: assessment, measurement invariance, and latent mean differences

Hao Wu¹ · Andi Rahmat Saleh² · Gyöngyvér Molnár³

Received: 13 May 2021 / Revised: 9 February 2022 / Accepted: 21 February 2022 / Published online: 22 March 2022
© The Author(s) 2022

Abstract

Indonesian 15-year-old students' educational achievement has proved to be significantly lower than the international average based on one of the most prominent international large-scale assessments, OECD PISA. This prompted our interest in Indonesian students' cognitive development during secondary schooling in comparison with that of students who are more successful on the PISA scale. This study aims to describe the development of Indonesian students' thinking skills, especially inductive and combinatorial reasoning, to map the influencing factors in the development of reasoning skills and make a cross-national comparison study with their Hungarian peers. The sample was drawn from 250 Indonesian and 864 Hungarian 8th and 11th graders. The reliabilities of the tests were acceptable, with Cronbach's alpha varying between .70 and .86. Results mainly indicated that: (a) the measures of these two constructs were partially invariant across nationalities; (b) the Indonesian students' reasoning skills did not sufficiently develop between the 8th and 11th grades; (c) the Indonesian students' inductive and combinatorial reasoning achievement was significantly worse than that of the Hungarian students in 11th grade, but not in 8th grade; and (d) the students' learning strategies proved to be predictive factors in students' reasoning skills achievement in both Indonesian and Hungarian contexts, but the level of prediction varied. Overall, the results of the current study provide support for an understanding of Indonesian students' cognitive development compared to an international benchmark. The findings have implications for revising educational methods in Indonesia and may form the basis for further studies.

Keywords Cognitive development · Cross-national comparison · Indonesian education · Reasoning skills

Introduction

One of the most prominent large-scale educational assessments is the Programme for International Student Assessment (PISA) conducted by the Organisation for Economic Co-operation and Development (OECD) since the turn of the millennium. Nowadays, it has become “the world's premier

yardstick for evaluating the quality, equity and efficiency of school systems” (OECD, 2016, p. 3) and has come to be considered as “an important mechanism for shifting, influencing, and shaping educational policy around the world” (de Roock & Espeña, 2018, p. 304). Indonesian 15-year-old students' performance was unsatisfactory in all three measured literacy domains (mathematics, reading, and science) in PISA's latest released results (i.e. PISA 2018). Their educational achievement was much lower than the average level among all the participating countries (OECD, 2019a), and only a negligible percentage of Indonesian students were identified as top performers in at least one subject (OECD, 2019b). Another worrying fact was that Indonesian 15-year-old students' educational achievement did not show substantial improvement in the past two decades (OECD, 2019b). That is, according to the international PISA results there is huge potential in the development of Indonesian students'

✉ Hao Wu
haowu@edu.u-szeged.hu

¹ Center for Learning and Instruction, University of Szeged, MTA-SZTE Digital Learning Technologies Research Group, Petőfi Sándor sgt. 32–34, Szeged 6722, Hungary

² Doctoral School of Education, University of Szeged, Szeged, Hungary

³ Institute of Education, University of Szeged, MTA-SZTE Digital Learning Technologies Research Group, Szeged, Hungary

educational achievement and cognitive development, a topic that must be examined closely.

This study aims to contribute to a further understanding of Indonesian students' cognitive development, especially in the areas of inductive reasoning (IR) and combinatorial reasoning (CR). Reasoning is normally understood as a generalized capability to acquire, apply and transfer knowledge (Molnár et al., 2017). It plays a significant role in school education (see e.g. de Castro, 2004; Kambeyo & Csapó, 2018) and in almost all higher-order cognitive processes, such as general intelligence, problem solving, knowledge acquisition, and knowledge application (Wu & Molnár, 2018; Molnár et al., 2017; Csapó, 1997; Söderqvist et al., 2012). Therefore, insufficiently developed reasoning skills would very likely influence students' educational achievement and reflect on their performance in educational assessment projects such as PISA.

Literature review

Inductive reasoning: concept and development

IR is the cognitive process of moving from the specific to the general (Sandberg & McCullough, 2010). It “entails using existing knowledge or observations to make predictions about novel cases” (Hayes et al., 2010, p. 278). To be more specific, Klauer (1990) defined IR as the discovery of regularities in detecting similarities and/or dissimilarities in the attributes of or relations to or between objects. Klauer's (1990) definition of IR is widely employed in IR assessment design (see Fig. 1 in the “Methods” section). Csapó (1997) pointed out that IR is a basic component of thinking and it forms a central aspect of intellectual functioning. Bisanz et al. (1994) also claimed IR as central to many types of learning and one of the most important factors in cognitive development. Some empirical studies have discovered the importance of IR in higher-order cognitive processes. For example, Molnár et al. (2013) and Wu & Molnár (2018) demonstrated and discussed the significant links between

IR and problem solving, Nikolov and Csapó (2018) found that students' performance in language learning is highly correlated with their IR skills, and Vartanian et al. (2003) confirmed that IR has a remarkable influence on divergent thinking.

The development of IR skills starts at a very early age (Perret, 2015; Schulz et al., 2008) and covers a broad age range (Csapó, 1997), the whole period of primary and secondary education (Molnár et al., 2013), offering opportunities for enhancement. Some researchers have suggested that explicit training is the best way to promote students' IR development (e.g. Klauer & Phye, 2008; Lipman, 1985), which can be realised effectively in both face-to-face and technology-based environments (see Molnár, 2011; Mousa & Molnár, 2020). However, an explicit IR training programme is not commonly applied in most schools; as a result, students' IR skills achieve relatively slow development (approximately one quarter of a standard deviation per year) (Molnár et al., 2013).

Combinatorial reasoning: concept and development

CR is “the process of creating complex constructs out of a set of given elements that satisfy the conditions explicitly given or inferred from the situation” (Adey & Csapó, 2012, p. 31; see Fig. 2 in the “Methods” section). Cognitive operations, such as combinations, arrangements, permutations, notations, and formulae, are employed in the process (English, 2005; Gál-Szabó & Bede-Fazekas, 2020). CR has been considered as one of the basic components of formal thinking (Batanero et al., 1997). The relationship between CR and higher-order cognitive processes has been frequently discussed. Csapó (1999) claimed CR plays a significant role in school learning (e.g. mathematics; see English, 2005) and everyday thinking. English's (2005) study highlighted the essential meaning of CR in several types of problem situations, such as those involving selections, distributions, and partitions. Csapó (1999, p. 51) emphasized that well-developed CR has the potential to “improve fluency of thinking when considering

Fig. 1 Sample items for the IR test (the original items were in Hungarian and Indonesian)

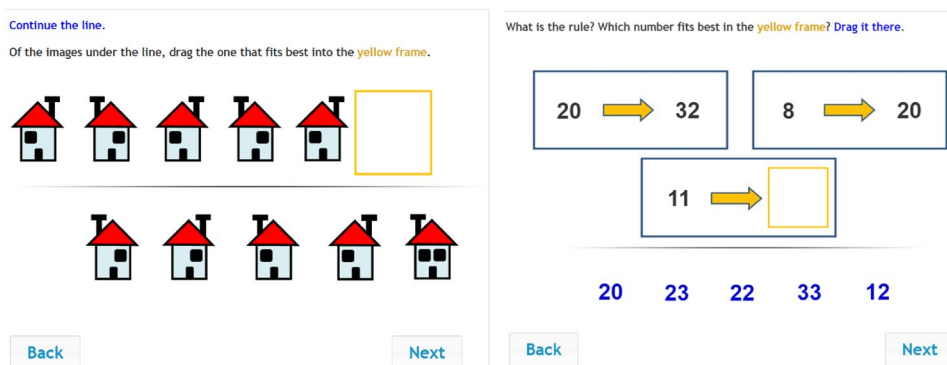


Fig. 2 Sample item for the CR test (the original items were in Hungarian and Indonesian)

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

Drag the different pieces of clothing to the pictures of the boy.

Create different sets of clothes for each picture. Put together all the possible sets of clothes.

Be careful. There are more pictures of the boy than there are options to dress him.

Back Next

different solutions to a problem; finding unusual relationships between certain elements, concepts, propositions; or generating a large variety of patterns from given units”.

In Piaget’s theory, the development of combinatoric operations is one of the important components of cognitive growth (see e.g. Inhelder & Piaget, 1958). Thus, many studies have linked the development of CR skills with the Piagetian stages of cognitive development (Batanero et al., 1997; English, 2005; Gál-Szabó & Bede-Fazekas, 2020). Children at Stage I “use random listing procedures, without trying to find a systematic strategy”, they “use trial and error, discovering some empirical procedures with a few elements” at Stage II, and, finally, “after the period of formal operations, adolescents discover systematic procedures of combinatorial construction, although for permutations, it is necessary to wait until children are 15 years old” (Batanero et al., 1997, p. 182). Some studies have found the development of CR does not always precisely follow the Piagetian stages. For instance, Fischbein’s (1975) study concluded, without specific teaching or training, students’ combinatorial problem-solving capacity may lag behind what the level should be. On the other hand, English’s (1993) study found some students attempt to use systematic combinatorial strategies even before the formal operational stage. English (1991, 1993) suggested that a well-designed context can prompt students to use combinatorial strategies or methods beyond their current stage of development.

The role of learning strategies in the development of reasoning skills

Thinking skills like IR and CR develop over a number of years, offering opportunities for enhancement. Both IR and CR can be improved through explicit training (e.g. Molnár, 2011; Fischbein, 1975; Klauer & Phye, 2008). Without direct enhancement, students’ reasoning skills can also develop as a by-product of ordinary school learning activities (de Konig, 2000).

The learning strategies (i.e. “the plans students select to achieve their goals”; Artelt et al., 2003, p. 13) have a strong influence on students’ academic achievement (Riding, & Rayner, 2013). Moreover, Aizpurua et al. (2018) pointed out that teaching and promoting the application of learning strategies have a positive influence not only on students’ educational achievement, but also on their ability to learn. In some studies (e.g. Artelt et al., 2003; Ghiasvand, 2010), learning strategies are placed into two categories: cognitive (“learn, remember, and understand the material”; Yukselturk & Bulut, 2007, p. 73) and metacognitive strategies (“planning, monitoring, and regulating their cognition”; Yukselturk & Bulut, 2007, p. 73).

Both types of learning strategies have the potential to influence the development of reasoning skills. Different cognitive strategies—such as memorization and elaboration strategies—operate different information processing skills (Artelt et al., 2003), resulting in different developmental

levels of information processing skills, and, in parallel, in the application of different higher-order cognitive processes. Csapó & Molnár's (2017) empirical study found significant correlations between cognitive strategies (memorization and elaboration) and certain higher-order cognitive processes, such as problem solving. In addition, Aizpurua et al.'s (2018) empirical study indicated that the use of cognitive strategies contributes to divergent thinking and creative intelligence in a positive way.

When students apply metacognitive strategies, "both knowledge and cognitive skills are planned, monitored, analyzed, evaluated, and reflected by students based on their own goals" (Lee et al., 2018, p. 43). In this process, students are able to develop advanced in-depth learning and active cognitive processing, thus fostering their reasoning skills (Lee et al., 2018; Zimmerman, 2002). Training in metacognitive strategies is, therefore, considered as an effective tool for improving students' reasoning skills (Lee et al., 2018). Lestari and Jailani (2018) conducted an empirical study at an Indonesian junior high school. Two groups of students were required to engage in collaborative learning, while metacognitive strategies were embedded in the learning activities in the first group only. A reasoning test was delivered to the students after the learning activities were done, on which the students in the first group showed a statistically higher performance than their peers. The results thus suggested positive effects from the use of metacognitive strategies to the development of reasoning skills.

IR and CR: research endeavours in Indonesia

Some studies have explored the feasibility of assessing reasoning skills in the Indonesian context. Lubis and Maulina (2017) designed a reliable figural inductive reasoning test for Indonesian high school students. Novia and Riandi (2017) investigated scientific reasoning skills in a group of Indonesian junior high school students, including CR. Generally, the available research has not focused on the development of reasoning skills themselves, but tested the reliability and validity of a newly developed instrument or used the developmental level of students' reasoning skills as a basis for understanding the development of another phenomenon.

Siswanto (2014) assessed Indonesian high school students' mathematical reasoning from the perspective of IR, but the main aim of the study was to confirm the effectiveness of the Student Teams Achievement Divisions (STAD) teaching method. Sudria et al. (2018) conducted an IR assessment in an Indonesian secondary school. The assessment results were used to explain and analyse students' learning activities in domain-specific learning. The role of CR in subject teaching (e.g. mathematics; Septiati, 2016) and in learning strategies and activities (Sumarmo et al., 2012) has also been in the focus of previous empirical

studies in Indonesia. That is, even though there are some researches on Indonesian students' reasoning skills, an in-depth analysis focusing on Indonesian students' IR and CR and their development is still missing.

Significance of research

PISA indicated that Indonesian students' educational achievement was at the bottom among all the participating countries/economies. It is urgent to conduct a study to explore the reasons behind Indonesian students' unsatisfactory educational achievement and ascertain potential methods to improve the situation. Considering the important role of reasoning skills in students' higher-order cognitive processes as well as educational achievement, we suppose there is a developmental difference in reasoning skills between Indonesian students and their international peers. However, there are barely any studies which place the developmental level of Indonesian students' reasoning skills in international context. As a result, we cannot even confirm if the developmental difference really exists, and—of course—we are not able to discuss it further (e.g. to investigate the age group in which the developmental difference starts). Cross-national comparative studies of students' reasoning skills are therefore urgently needed in Indonesia. Such studies will be able to help instructors, researchers and policy-makers to discover and understand the advantages, disadvantages and potential problems of the development of Indonesian students' reasoning skills.

Furthermore, there are several currently unanswered research questions that can be answered by well-designed cross-national studies. There is a basic and important question that has to be answered first: is the underlying measurement model for reasoning skills equivalent between the Indonesian and international contexts. Some studies have pointed out the possibility of measuring reasoning skills invariantly across nations or cultures (e.g. Lakin, 2012; van de Vijver, 2002). However, there is a lack of studies on this issue in Indonesian context. Moreover, previous studies have shown that students' level of reasoning skills is influenced by their use of learning strategies. It will be important to confirm if this influence exists in Indonesian context. The result will be significant in revising school education methods. For instance, Lestari and Jailani's (2018) study made a contribution in this area. Their study confirmed the positive influence of metacognitive strategies on students' development of reasoning skills in Indonesian context. However, their study only focused on the influence of metacognitive strategies, and the reasoning skills they measured are relevant to a specific knowledge domain (mathematics). Thus, how the two types of learning strategies influence students' domain-general reasoning skills in Indonesian context is still an unexplored

topic. In general, this study attempts to fill these gaps by assessing Indonesian students' reasoning skills using an international benchmark.

Research aims

The aim of the present analysis is to gain a better understanding of Indonesian students' interpretation and development of reasoning skills and influencing factors in international context. Specifically, this study seeks to elaborate the constructs of IR and CR and their operationalization in Indonesia and international benchmark. Hungary's performance on PISA 2018 fell in the middle of all participating countries/economies. Hungarian students' scores for reading, mathematics, and science were 476, 481, 481, respectively, which were significantly higher than 15-year-old students' average achievement in Indonesia (371, 379, and 396, respectively), but very close, although still below the OECD average (487, 489, and 489, respectively) (OECD, 2019a). Thus, Hungary can be used as an international benchmark for Indonesia. Furthermore, to gain more information about Indonesian students' unsatisfactory performance on PISA, this study aims to put the level of students' thinking skills in developmental context by exploring students' reasoning skills before and after PISA age (15 years old). This study, therefore, focuses on two groups of students, 8th and 11th graders, in both Indonesia and Hungary.

Specifically, we aim to explore: (1) measurement invariance across nationalities and grades; (2) developmental differences between Indonesian and Hungarian students; and between students before and after PISA age; and (3) the influence of learning strategies on IR and CR achievement. We will thus be able to answer the following research questions.

(RQ1) Can IR and CR assessment instruments be measurement invariant across nationalities and grades in the contexts of both Indonesia and Hungary?

(RQ2) Can developmental differences in IR and CR be detected between students before and after PISA age (i.e. 8th and 11th grades) in the contexts of both Indonesia and Hungary?

(RQ3) Can developmental differences in IR and CR be detected between Indonesian and Hungarian students before and after the PISA age?

(RQ4) How do learning strategies influence students' IR and CR achievement in both the Indonesian and Hungarian contexts?

Methods

Participants

The study sample ($N = 1114$) consisted of an Indonesian and a Hungarian sample. Participants were randomly drawn from 8 to 11th grades in Indonesian and Hungarian public primary and secondary schools. A total of 345 Indonesian students took part in the study. However, students who had more than 80% data on any of the measures missing were excluded from all the analyses. In the end, data from 250 Indonesian students were available for the analyses, which consisted of 56 8th graders (mean age = 14.20, $SD = 0.519$) and 194 11th graders (mean age = 17.08, $SD = 0.568$). Because the Hungarian students are used for the benchmark settings, we involved a much larger sample in Hungary. 864 Hungarian students were available for the analyses after the data cleaning, which consisted of 690 8th graders (mean age = 14.82, $SD = 0.567$) and 174 11th graders (mean age = 17.95, $SD = 0.611$). In addition, we used randomly selected subjects from the Hungarian sample based on demographic data. Except the age and grades noted above, the gender ratio between the Indonesian and Hungarian samples was also matched. The Indonesian sample consisted of 100 boys and 150 girls, while the Hungarian sample consisted of 404 boys and 460 girls. A χ^2 test confirmed no significant difference between the gender distribution in these two samples ($\chi^2 = 3.58, p > .05$).

Instruments

Inductive reasoning

The IR test was originally developed in Hungary (see Molnár, 2011; Csapó et al., 2009; Molnár & Csapó, 2011). It comprised analogy and series items in both numerical and figural formats. It has been applied widely in national and international contexts (see e.g. Saleh & Molnár, 2018; Wu & Molnár, 2018; Pásztor et al., 2018; Mousa & Molnár, 2020). The present online version was developed in Hungarian and then adapted for and translated to Indonesian. The translation was done by a group of language experts. The translation was double-checked by two Indonesian language teachers and a Hungarian teacher who is bilingual in Indonesian and Hungarian. The test comprised drag-and-drop-based multiple-choice items (see Fig. 1). The students were expected to solve (1) series reasoning problems (in which they observed a sequence of numbers/figures following a certain pattern and filled in the missing number(s)/figure(s) in a given series; see the left part of Fig. 1); and (2) analogy reasoning problems (in which they

discovered the similarity between two sets of numbers/figures and filled in the missing number/figure in the third set; see the right part of Fig. 1).

Forty-three (43) IR test items were delivered to the Indonesian students, while 36 items were provided for their Hungarian peers. Twenty-five (25) anchoring items appeared in both the Indonesian and Hungarian test versions. Only the 25 anchoring items were used for the analysis in this study. All of the IR items were scored dichotomously (1 for the correct answer; 0 in all other cases) and automatically by the eDia online assessment platform.

Combinatorial reasoning

The development and use of the CR test also enjoy a long history. It was originally developed and designed by Csapó (1999) as a paper-and-pencil test. In the current study, we used the improved, computerized version (Pásztor & Csapó, 2014) in both languages, Indonesian and Hungarian. The test contains both figural and verbal items. Each item provides certain elements (figures/images or letters/numbers) and a clear requirement for combing the elements. Students are expected to combine figures/images or letters/numbers and create different combinations which fit the given requirement using drag-and-drop (in the case of figural items; see Fig. 2) or typing the answers in a text box (in the case of verbal items). Students' performance was scored according to a specially developed J index (Csapó, 1988). The J index takes into account the correct and redundant combinations relative to all possible combinations (Csapó, 1988, p. 54). The index can take a value between 0 and 1 for each task, where a value of 1 means a list of all correct combinations without unnecessary combinations. The J index is computed by equation: $J = x(T - y)/T^2$. In this equation, T stands for the number of combinations belonging to the complete list, x stands for the number of correct combinations provided by the test taker, and y stands for the number of superfluous/redundant combinations provided by the test taker (Csapó, 1988). Moreover, Csapó (1988) determined that if y is larger than T , the J index will be zero. The CR test consisted of ten items in the Indonesian version and eight items in the Hungarian one. There were seven anchoring items in both versions, which were used for the analyses in the study.

Learning strategies questionnaire

The questionnaire in this study focused on students' learning strategies in their daily learning activities. The learning strategies questions were adapted from the internationally widely used PISA 2000 learning strategies questionnaire (Artelt et al., 2003), which was available in both languages. The questionnaire measured both cognitive strategies (involving information processing) and metacognitive strategies

(involving conscious regulation of learning). It contains 13 statements about different learning habits. The items can be clustered around three different strategies: (1) elaboration strategies (cognitive strategies to link the new material with previous knowledge or real life; four questions; sample item: When I study, I try to relate new material to things I have learned in other subjects), (2) memorization strategies (cognitive strategies to memorize knowledge without further processing; four questions; sample item: when I study, I try to memorise everything that might be covered), and (3) control strategies (metacognitive strategies to ensure learning goals are reached; five questions; sample item: when I study, I start by figuring out what exactly I need to learn). A five-point Likert scale (Never, Rarely, Sometimes, Often and Always) was used to indicate the frequency of a given habit in daily study.

Procedures

The tests and the questionnaire were administered online via the eDia assessment platform (Csapó & Molnár, 2019) in the ICT room of the participating schools. A pilot test had confirmed the feasibility and reliability of using the eDia platform in the Indonesian environment (Saleh & Molnár, 2018). Test completion was divided into two sessions, each lasting approximately 45 min; there was a ten-minute break between each session. In session 1, students worked on the CR test and the questionnaire. In session 2, they completed the IR test. Based on our practical experience from previous studies in Hungary and Indonesia (e.g. Pásztor et al., 2018; Saleh & Molnár, 2018), the testing time provided is sufficient for students to finish the tests. Moreover, there was no time limit on answering single items. The tests therefore put more emphasis on assessing students' ability rather than speed at which they complete the reasoning items. Testing sessions were supervised either by research assistants or teachers, who had been trained in test administration. The tests and questionnaire were prepared in the students' native languages, that is, Indonesian in Indonesia and Hungarian in Hungary.

Data analysis plan

Measurement invariance across nationalities and grades was tested by means of multi-group confirmatory factor analysis (MGCFA). The model fit for the measurement models was represented by two incremental fit indices, the Tucker–Lewis Index (TLI) and the comparative fit index (CFI), as well as an absolute fit index [the root mean square error of approximation (RMSEA)]. TLI and CFI $\geq .90$ as well as RMSEA $\leq .08$ are typically considered adequate (van de Schoot et al., 2012). This criterion has been employed in a large number of studies (e.g. Blevins

et al., 2015; Chan et al., 2009; Fong & Ng, 2012; Roberson et al., 2018; etc.). Moreover, Cudeck and Browne (1992) suggested a looser cut-off criterion for RMSEA, they considered RMSEA scores between .08 and .10 as a marginal fit, and they stated that they “would not want to employ a model with a RMSEA greater than 0.1” (Cudeck & Browne, 1992, p. 239). This looser criterion for RMSEA is also widely employed (see e.g. Furlong et al., 2005; Krause et al., 2003; Ryberg et al., 2020; Swami & Chamorro-Premuzic, 2008; etc.). Furthermore, Hu and Bentler (1999) proposed a more stringent cut-off criterion (i.e. TLI and CFI \geq .95 as well as RMSEA \leq .06). In this study, we considered TLI and CFI \geq .90 paired with RMSEA \leq .10 as an acceptable model fit and considered the criterion proposed by Hu and Bentler (1999) as evidence of a very good model fit.

According to Byrne and Stewart (2006), three models of invariance are distinguished: (1) configural invariance to investigate if the instrument has the same factor structure across groups; (2) strong factorial invariance to indicate the cross-group equality in the loadings and intercepts; and (3) strict factorial invariance to determine if the compared groups have the same item residual variances. Measurement invariance exists if the model fit parameters do not result in a significant difference between the nested models. Otherwise, between-group differences may reflect different psychometric properties of the items (Byrne & Stewart, 2006).

The most classic way to identify the differences between the invariance models is to test the significance of the change in χ^2 between the nested models. Yoon and Lai (2018, p. 202) pointed out that when analysing measurement invariance, “large imbalances in group sizes can affect the results because the chi-square statistics include a weighting by sample size”. This study ran three measurement invariance analyses to answer RQ1. In all three measurement invariance analyses, we faced a large imbalance in group sample sizes (Indonesian vs Hungarian samples; 8th vs 11th graders in both countries). Therefore, in this study, the differences between the invariance models was not identified by a χ^2 difference test, but by using a more traditional approach and focusing on the changes in CFI and RMSEA values (see e.g. Cheung & Rensvold, 2002; Putnick & Bornstein, 2016; Rutkowski & Svetina, 2014). The invariance models have a statistically equal model fit when the absolute Δ CFI is under

or equal to .020 or, according to the strictest perspective, .010, while absolute Δ RMSEA is under or equal to .015 (see Chen, 2007; Vincent-Höper & Stein, 2019; Vandenberg & Lance, 2000; Wang et al., 2013; Zhang & Bian, 2020).

If at least strong factorial invariance is established, mean comparisons can be meaningfully interpreted. In comparing the 8th and 11th graders’ ability levels, we used the 8th graders as a reference group, and we constrained their IR and CR latent mean values to zero in the latent mean comparison analyses (to answer RQ2). With the country-level latent mean comparison analyses, we used the Hungarian students’ achievement as a reference point and set their latent mean values to zero in both grades (to answer RQ3). The latent means were computed with Mplus (Muthén & Muthén, 2010).

The structural equation modelling (SEM; Bollen, 1989) approach was used to analyse the relationships between IR and CR on a latent level and the influential role of the learning strategies under examination (to answer RQ4). Two models were built (computed with Mplus) based on both the Indonesian and Hungarian samples. The quality of the structural equation models was evaluated by the indices CFI, TLI, and RMSEA.

Results

Reliability and measurement invariances across nationalities and grades (RQ1)

The internal consistencies of the reasoning tests were acceptable in both of the countries. Cronbach’s alpha for the IR and CR tests was .86 and .70 in the Indonesian context and .86 and .80 in the Hungarian context.

Measurement invariance analyses were conducted across nationalities for both of the IR and CR tests (see Table 1). The invariance models showed acceptable model fits except for the strict factorial invariance model for the CR test (CFI and TLI were lower than the .90, while RMSEA was higher than .10). Moreover, the model of strong factorial invariance did not result in a remarkable difference in model fit compared to the model of configural invariance (indicated by absolute Δ CFI and Δ RMSEA) in both cases, but CR

Table 1 Goodness of fit indices for testing invariance across nationalities

Test	Invariance model	CFI	TLI	RMSEA	Δ CFI	Δ RMSEA
IR	Config. inv	.918	.954	.055	–	–
	Strong factorial inv	.926	.952	.055	.008	.000
	Strict factorial inv	.920	.955	.054	.006	.001
CR	Config. inv	.948	.915	.087	–	–
	Strong factorial inv	.942	.922	.084	.006	.003
	Strict factorial inv	.898	.881	.104	.044	.020

proved to be not invariant in a strict sense. (Strict factorial invariance is not a prerequisite for group comparisons of means and variances; see Byrne & Stewart, 2006; Csapó et al., 2014). That is, the measurement invariance of the CR test partly held across nationalities. Latent mean differences could be interpreted as true differences in IR and CR and were not due to the psychometric issue. The results obtained in Indonesia and Hungary for IR and CR can be represented on the same international scale.

Measurement invariance analyses were conducted across grades for both the Indonesian (Table 2) and Hungarian samples (Table 3) to confirm that the results of the 8th and 11th graders can be represented on the same scale in both nations. In the area of IR, the model fits for the invariance models were acceptable, while ΔCFI was lower than .020 and ΔRMSEA was noticeably lower than .015 in both countries, indicating no remarkable difference between the three invariance models for IR. Thus, the results suggested that IR was measurement invariant across grades, independent of nation.

In the area of CR, the model fits for the Indonesian invariance models were very good. However, the Indonesian strict factorial invariance model showed a remarkable decrease in model fit (the ΔCFI was in the .010–.020 range, but ΔRMSEA was above .015; see Table 2). Thus the CR test proved to be measurement invariant across grades at a strong level of factorial invariance in the Indonesian sample. By contrast, the fit indices of the Hungarian invariance models failed to meet the cut-off values (see Table 3), so the measurement invariance across grades did not hold in the Hungarian sample.

To sum up, our IR and CR tests proved to be measurement invariant across nationalities at least at a strong level of invariance. Therefore, students in both countries

conceptualize the construct in the same way (Milfont & Fischer, 2010) and employ the same conceptual framework to answer the test items (configural level). Strong invariance indicated that the means, variances and covariances for the latent variables can be compared between these two samples (Csapó et al., 2014). The strict factorial level measurement invariance across nationality did not hold for the CR test, which indicated that the Indonesian and Hungarian students have different item residual variances. As strict factorial invariance is not a prerequisite for country comparisons of latent factor means and variances, we can conclude that, in general, the measurement invariance analyses confirmed the feasibility for comparing the group means at a latent level between the Indonesian and Hungarian contexts.

Similarly, the measurement invariance across grades of the IR and CR tests held at least at a strong level of invariance in the Indonesian context. Therefore, the latent mean comparison between the 8th and 11th graders in the Indonesian context is possible. However, in the Hungarian context, the measurement invariance across grades can only be detected on the IR test. Thus, a latent mean comparison of the CR test cannot be made between the Hungarian 8th and 11th graders.

Mean comparisons on latent level (RQ2 & RQ3)

As measurement invariance across grades for the IR test is sufficiently met for both nationalities, we compared the IR achievement of the 8th and 11th graders on a latent level in both contexts. The results indicated that the Indonesian 8th and 11th graders demonstrated a statistically similar performance on the IR test ($M_{IDN_8} = 0; M_{IDN_{11}} = -.02; SE = .05; p > .05$). By contrast, the Hungarian 11th graders showed

Table 2 Goodness of fit indices for testing measurement invariance across grades—Indonesia

Test	Invariance model	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA
IR	Config. inv	.909	.916	.094	–	–
	Strong factorial inv	.922	.925	.089	.013	.005
	Strict factorial inv	.910	.915	.094	.012	.005
CR	Config. inv	1.000	1.000	.000	–	–
	Strong factorial Inv	1.000	1.000	.000	.000	.000
	Strict factorial Inv	.980	.976	.038	.020	.038

Table 3 Goodness of fit indices for testing invariance across grades—Hungary

Test	Invariance model	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA
IR	Config. inv	.967	.967	.030	–	–
	Strong factorial inv	.963	.964	.032	.004	.002
	Strict factorial inv	.969	.968	.030	.006	.002
CR	Config. inv	.933	.892	.104	–	–
	Strong factorial inv	.921	.892	.104	.012	.000
	Strict factorial inv	.899	.882	.109	.022	.005

a significantly higher performance on the IR test than the Hungarian 8th graders ($M_{HUN_8}=0$; $M_{HUN_11}=.41$; $SE=.11$; $p<.001$).

In the CR achievement comparison across grades, the Indonesian 11th graders even displayed a slightly but significantly lower performance than the Indonesian 8th graders ($M_{IDN_8}=0$; $M_{IDN_11}=-.07$; $SE=.03$; $p<.05$). However, the grade difference on CR achievement was not comparable in the Hungarian sample, since measurement invariance did not hold.

The stagnation in the development of Indonesian students' reasoning skills led to an unsatisfactory result in the international comparison with Hungarian students. Results showed Indonesian 8th graders had the same level of development as that of Hungarian 8th graders in CR skills ($M_{IDN_8}=.02$; $M_{HUN_8}=0$; $SE=.01$; $p>.05$). The Indonesian 8th graders even showed significantly better performance than the Hungarian 8th graders on the IR test ($M_{IDN_8}=.31$; $M_{HUN_8}=0$; $SE=.05$; $p<.001$). However, due

to Indonesian students did not achieve sufficient development of reasoning skills between ages 14 and 17, the Indonesian 11th graders showed significantly worse performance on both the IR ($M_{IDN_11}=-.25$; $M_{HUN_11}=0$; $SE=.08$; $p<.01$) and CR ($M_{IDN_11}=-.29$; $M_{HUN_11}=0$; $SE=.03$; $p<.001$) tests than their international peers.

Learning strategies' influence on IR and CR (RQ4)

To investigate the impact of potential influencing factors on students' reasoning skills, we expected that learning strategies predicted performance in both cultures. We used SEM within each nation to indicate the relationship between the level of reasoning skills under examination and identify the predictive power of learning strategies to the levels of IR and CR skills (see Figs. 3 and 4).

The structural models fit the data reasonably well (IDN: $\chi^2=230.47$, $df=128$, $CFI=.90$, $TLI=.92$, $RMSEA=.06$; HUN: $\chi^2=746.67$, $df=201$, $CFI=.93$, $TLI=.96$,

Fig. 3 A structural model presenting the relationships between reasoning skills and learning strategies—Indonesian sample; **Significant at 0.01 level ($p<.01$)

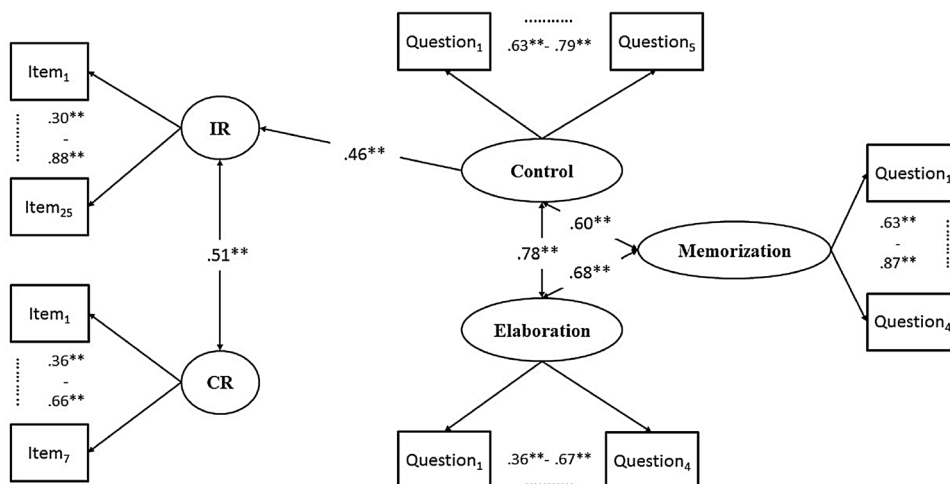
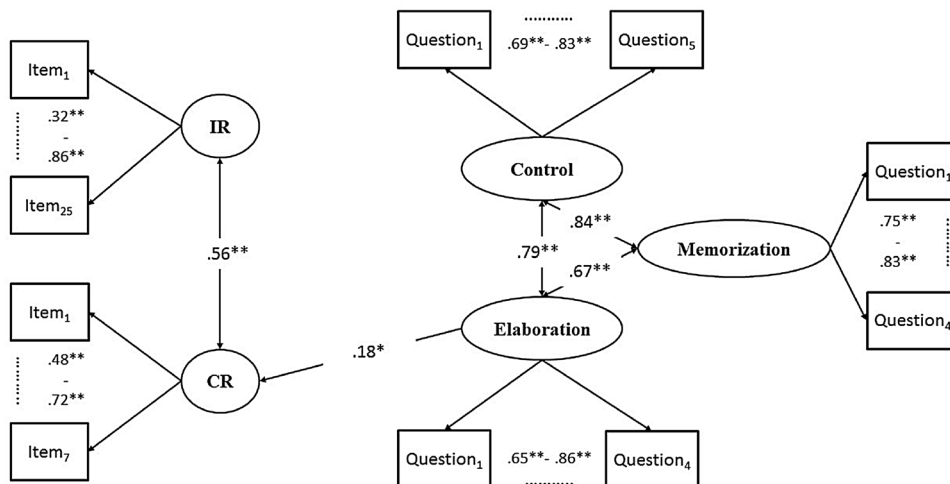


Fig. 4 A structural model presenting the relationships between reasoning skills and learning strategies—Hungarian sample; *Significant at 0.05 level ($p<.05$), **Significant at 0.01 level ($p<.01$)



RMSEA = .06). All the reasoning skills and learning strategies were built as latent variables, which were constructed by the corresponding test items or questionnaire questions. All the items or questions showed moderate to strong and significant ($p < .01$) factor loadings in both of the models. Results showed IR and CR are strongly correlated, independent of the cultural contexts (IDN: $r = .51, p < .01$; HUN: $r = .56, p < .01$). A similar phenomenon was noticed among the learning strategies; that is, the application of the three learning strategies under examination proved to be highly correlated (IDN: $r = .60-.78, p < .01$; HUN: $r = .67-.84, p < .01$) in both cultures. The results indicated that the use of learning strategies is not isolated. In other words, the frequencies for students using the three learning strategies under consideration are related. The use of control strategies was strongly predictive of IR ($\beta = .46, p < .01$) in the Indonesian sample. By contrast, the use of elaboration strategies showed moderate predictive power for CR ($\beta = .18, p < .05$) in the Hungarian sample. Only memorization strategies did not show any significant influence on reasoning skills in both models.

Discussion

The aim of the present study is to enhance our understanding of IR and CR as mental processes measurable through computer-based assessment with great relevance to education and to ascertain the development level of these skills in international context. More specifically, the results of the current study provide support for insights into Indonesian students' cognitive development using international benchmarks and has implications for revising educational methods in Indonesia.

The measures of these two reasoning skills were invariant (at least at a strong level of invariance) at the level of nationality (RQ1—measurement invariance across nationalities). That is, the Indonesian and Hungarian students' developmental level of reasoning skills can be represented on a common scale. They employ the same conceptual framework to answer the test items and the variances for latent variables, even though these two countries are very different in network, language, and cultural background. The findings indicate the possibility of conducting studies to further explore the developmental differences in reasoning skills between students from the Asia-Pacific and European regions.

The IR and CR measures were partially invariant across grades in the Indonesian sample. However, only IR was invariantly measured across grades in the Hungarian sample (RQ1—measurement invariance across grades). That is, the 8th and 11th graders conceptualize the construct of the IR test in the same way independent of their nationality, but this was not the case for the CR test. Compared to the IR test, the

students' answers on the CR test were more subjective. In addition, the scoring for the CR test was relatively complicated (especially compared to the dichotomous scoring on the IR test) and was influenced by a number of factors (see Csapó, 1988). Therefore, the measurement invariance analysis for the CR test was more strongly impacted by students' mental and/or behavioural differences. As a result, none of our measurement invariance analyses of the CR test held at a strict level of invariance (that is, at a strong level of invariance in the analyses across nationalities and across grades in the Indonesian sample, non-invariant in the analysis across grades in the Hungarian sample). Further study is recommended to improve our understanding of students' mental and/or behavioural differences on the CR test.

A latent mean comparison demonstrated that the Indonesian students' IR and CR skills did not sufficiently develop between 8 and 11th grades (i.e. between ages 14 and 17); by contrast, the Hungarian students showed significant development in IR during the same period of time (CR was not compared) (RQ2). The results pointed out a serious problem in the development of the Indonesian students' reasoning skills. As we have shown, the development of reasoning skills covers a broad age range: the whole period of compulsory schooling. However, the 8th to 11th grades may not be the most effective time to develop reasoning skills (Molnár et al., 2013). Students at this age still have the potential to achieve solid growth in reasoning skills, as we saw among the Hungarian students. Nonetheless, the Indonesian students' IR and CR skills did not develop as they should have.

Latent mean comparison analyses have also shown that the Indonesian students' IR and CR achievement was significantly worse than the Hungarian students' mean achievement in 11th grade, but not in 8th grade (RQ3). Results suggested the Indonesian students started to fall behind their international peers after 8th grade, that is, 1 year before the PISA age (15 years old). As regards the importance of IR and CR in the higher-order cognitive processes, there is reason to believe that unsatisfactory development of reasoning skills was one of the reasons for Indonesian students' poor performance on the PISA assessments. The poor performance on the PISA assessments reflects a low ability in mathematics, reading, and science. Therefore, we assume that insufficient growth in Indonesian students' reasoning skills hindered the development of these abilities at a certain level and caused the low PISA results at some level. The extent to which reasoning skills affect Indonesian students' performance on PISA as well as their mathematics, reading, and science abilities falls outside the scope of this study, but may be an interesting topic for the future.

The Indonesian students' disadvantage in reasoning skills growth could be caused by a lack of direct and indirect school development because proper subject education is able to impact—even implicitly—students' reasoning skills

(Primi et al., 2010; Xin & Zhang, 2009). This has been confirmed by Daniel (2013), who reported that education in Indonesia was less effective and innovative than that of its neighbours (e.g. Singapore, Australia, Malaysia, and Thailand). Establishing explicit training is an effective way to promote development of students' reasoning skills (Klauer & Phye, 2008; Lipman, 1985). However, so far, explicit training is not very commonly used in schools. Revising teaching methods in Indonesian schools may be a more feasible method to enhance Indonesian students' reasoning skills.

Results indicated that the students' use of learning strategies predicted their performance on the reasoning skill tests (RQ4). Based on the present results, it is suggested that teaching methods be revised so that students are encouraged to use specific learning strategies, which have shown positive predictive effects on reasoning skills achievement. For instance, the use of control strategies positively predicted the Indonesian students' IR achievement. Therefore, in Indonesian education, teachers can encourage and guide students to make more frequent use of control strategies and thus to indirectly enhance their IR achievement. Similarly, the Hungarian students' CR achievement was positively predicted by the use of elaboration strategies. Corresponding revision in Hungarian education could also be designed and implemented. However, it is not clear why learning strategies played different roles in the different cultures. We assumed these differences might be caused by the differing cultural backgrounds and traditions. Further empirical studies are needed to explore the reasons.

The study provides important insights into the international validity of reasoning skills measurements and points to the possibility and feasibility of explaining international differences in educational achievement by exploring the developmental differences of certain cognitive skills. Furthermore, the results shed light on the influences of students' learning strategies on their reasoning skills achievement, thus enhancing our understanding of the links between students' daily learning activities and their cognition. To sum up, the study provides a basis for further international studies and may contribute to revising teaching methods.

Limitations and future research

The original sample size of the Indonesian students was larger ($N = 345$). However, due to the data cleaning process, all students who had more than 20% missing data were deleted from the databases. Thus, the Indonesian sample size (especially the 8th graders) involved in the final analyses was relatively low. We assume this situation is due to the fact that (1) the Indonesian students were not sufficiently motivated on the test and (2) they were not very familiar with the computer-based assessment environment. Moreover,

considering that our samples were randomly collected, we cannot guarantee that our results have not been influenced by other background or environmental factors. Thus, our samples did not have sufficient power to be representative of the Indonesian students or of their Hungarian peers. The generalizability of the findings is therefore limited. Furthermore, we employed a relatively loose cut-off criterion in this study in evaluating the quality of the measurement invariance models and the structural equational models, which may weaken the validity of the results. Therefore, the study would need to be repeated for validation.

The CR test was found to be measurement non-invariant across grades in the Hungarian context. Thus, the comparison of CR achievement between Hungarian 8th and 11th graders could not be made. Further studies are therefore recommended to ascertain a possible reason for the measurement non-invariance that was detected and explore the true difference between Hungarian 14- and 17-year-old students' CR achievement.

The research only involved one country, Hungary, as the benchmark for the comparison. If we have more countries in a future comparison study, we might be able to acquire more knowledge about the behaviour and potential influencing factors of the reasoning skills being measured and the reasons for the poor Indonesian educational achievement in international context.

Funding This study has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the OTKA K135727 funding scheme and supported by the Research Programme for Public Education Development of the Hungarian Academy of Sciences (KOZOKT2021-16).

Data availability The data used to support the findings cannot be shared at this time as it also forms part of an ongoing study.

Code availability The codes used to analyse the data are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- 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). Nemzeti Tankönyvkiadó.
- Aizpurua, A., Lizaso, I., & Iturbe, I. (2018). Learning strategies and reasoning skills of university students. *Revista De Psicodidáctica*, 23(2), 110–116.
- Artelt, C., Baumert, J., Julius-McElvani, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning. Results from PISA 2000*. OECD.
- Batanero, C., Navarro-Pelayo, V., & Godino, J. D. (1997). Effect of the implicit combinatorial model on combinatorial reasoning in secondary school pupils. *Educational Studies in Mathematics*, 32(2), 181–199.
- Bisanz, J., Bisanz, G., & Korpan, C. A. (1994). Inductive reasoning. In R. Sternberg (Ed.), *Thinking and problem solving* (pp. 179–213). Academic Press.
- Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, 28(6), 489–498.
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.
- 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.
- Chan, D. K., Lonsdale, C., Ho, P. Y., Yung, P. S., & Chan, K. M. (2009). Patient motivation and adherence to postsurgery rehabilitation exercise recommendations: The influence of physiotherapists' autonomy-supportive behaviors. *Archives of Physical Medicine and Rehabilitation*, 90(12), 1977–1982.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.
- Csapó, B. (1988). *A kombinatív képesség struktúrája és fejlődése*. Akadémiai Kiadó.
- 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). Swets & Zeitlinger.
- 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. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia System. *Frontiers in Psychology*, 10, 1522.
- Csapó, B., Molnár, G., & Tóth, J. (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). Luxembourg: Office for Official Publications of the European Communities.
- Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, 106(3), 639.
- Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika*, 57(3), 357–369.
- Daniel, E. G. S. (2013). Asia Pacific science education in a knowledge society. *Asia Pacific Journal of Education*, 33(2), 170–182.
- de Castro, B. (2004). Pre-service teachers' mathematical reasoning as an imperative for codified conceptual pedagogy in algebra: A case study in teacher education. *Asia Pacific Education Review*, 5(2), 157–166.
- de Koning, E. (2000). *Inductive reasoning in primary education. Measurement, teaching, transfer*. Kerckebosch.
- de Roock, R. S., & Espeña, D. M. (2018). Constructing underachievement: The discursive life of Singapore in US federal education policy. *Asia Pacific Journal of Education*, 38(3), 303–318.
- English, L. D. (1991). Young children's combinatoric strategies. *Educational Studies in Mathematics*, 22(5), 451–474.
- English, L. D. (1993). Children's strategies for solving two- and three-dimensional combinatorial problems. *Journal for Research in Mathematics Education*, 24(3), 255–273.
- 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). Springer.
- Fischbein, H. (1975). *The intuitive sources of probabilistic thinking in children*. Reidel.
- Fong, T. C. T., & Ng, S. M. (2012). Measuring engagement at work: Validation of the Chinese version of the Utrecht work engagement scale. *International Journal of Behavioral Medicine*, 19(3), 391–397.
- Furlong, M. J., Greif, J. L., Bates, M. P., Whipple, A. D., Jimenez, T. C., & Morrison, R. (2005). Development of the California school climate and safety survey-short form. *Psychology in the Schools*, 42(2), 137–149.
- Gál-Szabó, Z., & Bede-Fazekas, Á. (2020). Formalization of odometer thinking and indices for the classification of combinatorial strategies. *International Electronic Journal of Mathematics Education*, 15(1), 1–16.
- Ghiasvand, M. Y. (2010). Relationship between learning strategies and academic achievement; based on information processing approach. *Procedia-Social and Behavioral Sciences*, 5, 1033–1036.
- Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(2), 278–292.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. Basic Books.
- Kambeyo, L., & Csapó, B. (2018). Scientific reasoning skills: A theoretical backgrounds to science education. *Reform Forum: Journal for Educational Research in Namibia*, 26(1), 27–36.
- Klauer, K. J. (1990). Paradigmatic teaching of inductive thinking. In H. Mandl, E. De Corte, S. N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction European research in an international context Analysis of complex skills and complex knowledge domains* (pp. 23–45). Pergamon Press.
- Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research*, 78(1), 85–123.
- Krause, E. D., Mendelson, T., & Lynch, T. R. (2003). Childhood emotional invalidation and adult psychological distress: The mediating role of emotional inhibition. *Child Abuse & Neglect*, 27(2), 199–213.
- Lakin, J. M. (2012). Multidimensional ability tests and culturally and linguistically diverse students: Evidence of measurement invariance. *Learning and Individual Differences*, 22(3), 397–403.
- Lee, Y., Capraro, M. M., Capraro, R. M., & Bicer, A. (2018). A meta-analysis: Improvement of students' algebraic reasoning through

- metacognitive training. *International Education Studies*, 11(10), 42–49.
- Lestari, W., & Jailani. (2018). Enhancing an ability mathematical reasoning through metacognitive strategies. *Journal of Physics: Conference Series*, 1097(1), 1–7.
- Lipman, M. (1985). Thinking skills fostered by philosophy for children. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills* (pp. 83–108). Lawrence Erlbaum Associates.
- Lubis, A. D., & Maulina, D. (2017). Development of an Indonesian figural-inductive reasoning test for high school students based on the Cattell-Horn-Carroll theory. *1st International Conference on Intervention and Applied Psychology (ICIAP 2017)*. Cham: Atlantis Press.
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–130.
- Molnár, G. (2011). Playful fostering of 6- to 8-year-old students' inductive reasoning. *Thinking Skills and Creativity*, 6(2), 91–99.
- Molnár, G., & Csapó, B. (2011). Az 1–11 évfolyamot átfogó inductív gondolkodási kompetenciakálá készítése a valószínűségi tesztelmélet alkalmazásával. *Magyar Pedagógia*, 111(2), 127–140.
- 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.
- Mousa, M., & Molnár, G. (2020). Computer-based training in math improves inductive reasoning of 9- to 11-year-old children. *Thinking Skills and Creativity*, 37, 100687.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Muthén & Muthén.
- Nikolov, M., & Csapó, B. (2018). The relationships between 8th graders' L1 and L2 reading skills, inductive reasoning and socio-economic status in early English and German as a foreign language programs. *System*, 73, 48–57.
- Novia, N., & Riandi, R. (2017). The analysis of students scientific reasoning ability in solving the modified Lawson Classroom Test of scientific reasoning (MLCTSR) problems by applying the levels of inquiry. *Jurnal Pendidikan IPA Indonesia*. <https://doi.org/10.15294/jpii.v6i1.9600>
- OECD. (2016). *PISA 2015 results (volume 1): Excellence and equity in education*. OECD Publishing.
- OECD. (2019a). *PISA 2018 results (volume 1): What students know and can do*. OECD Publishing.
- OECD. (2019b). *Indonesia - country note - PISA 2018 results*. OECD Publishing.
- 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., Kupiainen, S., Hotulainen, R., Molnár, G., & Csapó, B. (2018). *Comparing Finnish and Hungarian fourth grade students' inductive reasoning skills*. Paper presented at the EARLI SIG 1 Conference, August 29–31, University of Helsinki, Helsinki, Finland.
- Perret, P. (2015). Children's inductive reasoning: Developmental and educational perspectives. *Journal of Cognitive Education and Psychology*, 14(3), 389–408.
- 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.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90.
- Riding, R., & Rayner, S. (2013). *Cognitive styles and learning strategies: Understanding style differences in learning and behavior*. David Fulton Publishers.
- Roberson, P. N., Shorter, R. L., Woods, S., & Priest, J. (2018). How health behaviors link romantic relationship dysfunction and physical health across 20 years for middle-aged and older adults. *Social Science & Medicine*, 201, 18–26.
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57.
- Ryberg, R., Her, S., Temkin, D., Madill, R., Kelley, C., Thompson, J., & Gabriel, A. (2020). Measuring school climate: Validating the education department school climate survey in a sample of urban middle and high school students. *AERA Open*. <https://doi.org/10.1177/2332858420948024>
- Saleh, A. R., & Molnár, G. (2018). Inductive reasoning through the grades: Case of Indonesia. In *Edulearn 18. 10th international conference on education and new learning technology (Palma, 2nd-4th of July, 2018): Conference proceedings* (pp. 8790–8793). IATED Academy.
- 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). Routledge.
- Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., & Jenkins, A. C. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition*, 109(2), 211–223.
- Septiati, E. (2016). Kemampuan berpikir logis matematis mahasiswa pendidikan matematika pada mata kuliah matematika diskrit. In R. A. Siroj, H. Seri, Haryadi, & B. Hirza (Eds.), *Prosiding Seminar Nasional Pendidikan* (pp. 394–401). FKIP Universitas Muhammadiyah Palembang.
- Siswanto, R. (2014). Improvement of reasoning ability and mathematical connections through the application of stad cooperative learning model with geogebra software. *Jurnal Pendidikan Dan Keguruan*, 1(1), 209662.
- 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.
- Sudria, I. B. N., Redhana, I. W., Kirna, I., & Aini, D. (2018). Effect of Kolb's learning styles under inductive guided-inquiry learning on learning outcomes. *International Journal of Instruction*, 11(1), 89–102.
- Sumarmo, U., Hidayat, W., Zukarnaen, R., Hamidah, H., & Sarining-sih, R. (2012). Kemampuan dan disposisi berpikir logis, kritis, dan kreatif matematik (eksperimen terhadap siswa sma menggunakan pembelajaran berbasis masalah dan strategi think-talk-write). *Jurnal Pengajaran MIPA*, 17(1), 17–33.
- Swami, V., & Chamorro-Premuzic, T. (2008). Factor structure of the body appreciation scale among Malaysian women. *Body Image*, 5(4), 409–413.
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492.
- van de Vijver, F. J. (2002). Inductive reasoning in Zambia, Turkey, and the Netherlands: Establishing cross-cultural equivalence. *Intelligence*, 30(4), 313–351.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.
- Vartanian, O., Martindale, C., & Kwiatkowski, J. (2003). Creativity and inductive reasoning: The relationship between divergent thinking

- and performance on Wason's 2–4–6 Task. *The Quarterly Journal of Experimental Psychology Section A*, 56(4), 1–15.
- Vincent-Höper, S., & Stein, M. (2019). The role of leaders in designing employees' work characteristics: Validation of the health-and development-promoting leadership behavior questionnaire. *Frontiers in Psychology*, 10, 1049.
- Wang, M., Armour, C., Wu, Y., Ren, F., Zhu, X., & Yao, S. (2013). Factor structure of the CES-D and measurement invariance across gender in mainland Chinese adolescents. *Journal of Clinical Psychology*, 69(9), 966–979.
- 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.
- 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.
- Yoon, M., & Lai, M. H. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 201–213.
- Yukselturk, E., & Bulut, S. (2007). Predictors for student success in an online course. *Journal of Educational Technology & Society*, 10(2), 71–83.
- Zhang, Y., & Bian, Y. (2020). Emotion regulation questionnaire for cross-gender measurement invariance in Chinese university student. *Frontiers in Psychology*, 11, 2940.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.