Thinking Skills and Creativity 46 (2022) 101132

Contents lists available at ScienceDirect

Thinking Skills and Creativity

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

Assessing Indonesian student inductive reasoning: Rasch analysis

Soeharto Soeharto [a,*], Benő Csapó [b,c]

[a] Doctoral School of Education, University of Szeged, 32-34, Petőfi S. sgt., Szeged, H-6722, Hungary
[b] Institute of Education, University of Szeged, Hungary
[c] MTA-SZTE Research Group on the Development of Competencies, Hungary

A R T I C L E I N F O

Keywords:
Inductive reasoning
Rasch analysis
Evaluation
Gender
Grade schools

A B S T R A C T

Inductive reasoning is an ability related to student academic achievement and is embedded in 21st-century competency frameworks. The purpose of this study was to evaluate Indonesian students' inductive reasoning, validate an adapted inductive reasoning test for Indonesia, and classify the difficulty of items and students' reasoning ability. The participants were 856 students in Grades 10, 11, and 12 in senior high schools as well as undergraduate students in higher educational institutions in West Kalimantan province, Indonesia. Four different tasks were employed to assess student inductive reasoning. The data collection comprised an onlinetest through the eDia assessment platform and a traditional paper-based test. The results from the Rasch analysis demonstrated that the adapted inductive reasoning test met the validity and reliability criteria based on Rasch parameters. Differential item functioning (DIF) analysis revealed that only one of the 40 items is in the moderate to large DIF category based on the test method. The results further revealed that students performed better in solving figural items than numeric items. Furthermore, older students in higher grades had higher logit measures than younger students, however, no significant differences were found between students in Grades 10 and 11. No significant differences were found between females and males. The difficulty of items and students' abilities were also classified to understand the evaluation of students' inductive reasoning.

1. Introduction

Inductive reasoning is an ability that children need to promote their cognitive development and develop their intelligence. Inductive reasoning facilitates individuals' tunderstanding the abstraction of basic rules through their application to various fields. Inductive reasoning theories have been integrated from various disciplines, including mathematics, science, philosophy, psychology, and artificial intelligence (Perret, 2015; Sosa-Moguel & Aparicio-Landa, 2021). Worldwide 21st-century competency frameworks have also regarded inductive reasoning components as imperative. Inductive reasoning components are embedded into creativity and problem-solving in various significant competencies such as those found in computer science and technology, communication, social skills, collaboration, and teamwork. Employers, researchers, and policymakers have exhibited an interest in enhancing these related competencies (Chu et al., 2017; Van Vo & Csapó, 2020; Zhu & Neupert, 2021). Reasoning skills play an essential role in workplace environments and various educational contexts. The ability to reason inductively is needed and has been learned and is relevant in

Open Access Funding: This study has received funding from the University of Szeged Open Access Fund. Grant number: 5540.
* Corresponding author.
E-mail addresses: soeharto.soeharto@edu.u-szeged.hu, soehatofisika@gmail.com (S. Soeharto).

https://doi.org/10.1016/j.tsc.2022.101132
Received 9 November 2021; Received in revised form 23 August 2022; Accepted 6 September 2022
Available online 8 September 2022
1871-1871/

recent social changes. Furthermore, it is a factor that predicts student outcomes and achievements.

Many studies have revealed inductive reasoning is of importance in various contexts. A relationship between students' inductive reasoning and problem-solving ability and academic success has been found (Csapó, 1997; Csapó & Molnár, 2019; Korom et al., 2017; Perret, 2015; Sosa-Moguel & Aparicio-Landa, 2021). Wu and Molnár (2018) found that inductive reasoning can predict students' interactive problem-solving ability, which is a multi-dimensional cognitive process in specific thinking skills. Furthermore, it can facilitate student decision-making and be employed to establish various causal relationships (Lafraire et al., 2020; Leighton & Sternberg, 2003). It appears that it continue to increase in 5th, 6th, 7th, 9th, and 11th grade students in the Vietnamese context, inductive reasoning continues to increase (Van Vo & Csapó, 2020). However, this study did not show how inductive reasoning develops in students at higher levels, for example, by comparing senior high school students with those at undergraduate levels. Measuring inductive reasoning among students at higher levels is crucial to addressing the lack of information related to the development of students' inductive reasoning. Evaluation it may be beneficial for evaluating curriculum success, supporting student cognitive development, and providing an overview of student success rates for educators, especially for tstudents who are about to complete their education and start working. An objective measurement such as Rasch analysis can be applied to explore further information related to student and item investigation to assess inductive reasoning. Based on the literature review, we find no examples of the application of Rasch analysis being used to assess inductive reasoning in the Indonesian context whereas Rasch analysis can help researchers to extend the result and literature because it has several advantages such as meeting the requirements of fundamental measurement to transform raw data into a linear interval scale (logits), allowing researchers to investigate student performance and item difficulty using item-person maps, and being a psychometric technique developed to improve measurement accuracy whereby researchers can construct instruments and monitor instrument quality (Boone, 2016; Kleppang et al., 2020; Tavakol & Dennick, 2013).

In Indonesia, thinking skills were included in the 2013 Indonesian core curriculum (Hasan, 2013; Prastowo & Fitriyaningsih, 2020). This curriculum supported three main domains: attitude, skills, and knowledge,with the learning material designed to relate to core competencies in different disciplines (Hasan, 2013). This curriculum has a crucial problem in term of assessment methods especially in evaluating attitude assessment. The attitude assessment was completely new and difficult to adapt to the classroom context. Badaruddin and Hawi (2022) reported that the majority of teachers complained about the difficulty of assessing student attitudes, and their own of knowledge in choosing the method and developing the assessment instrument . However, assessing knowledge and skills was easy (Natsir et al., 2018). To enhance students' thinking skills, the teacher may employ various learning models on different materials and subjects (Prastowo & Fitriyaningsih, 2020). Although inductive reasoning is not taught and trained directly in schools, the inductive reasoning test has been utilized in the general basic skills knowledge test when applying for jobs at the government and company levels. Thus, limited data have been employed to describe students' inductive reasoning at schools and even at higher education institutions. Additionally, some limitations in data collection may occur related to internet access and the computer laboratory. Consequently, we used online and paper-based tests, and this method is also used to confirm bias issues regarding the data collection method. Therefore, it is imperative that studies related to the evaluation of inductive reasoning in the Indonesian context are conducted. It is possible that because inductive reasoning is closely related to mathematics, reading, and science performance (De Koning et al., 2002; Nikolov & Csapó, 2018; Soeharto et al., 2019), it was students' low inductive reasoning ability that resulted in Indonesia's low ranking in the 2018 PISA report in which the country was placed 71st out of 77 participating countries (OECD, 2020). Because inductive reasoning was not embedded in the Indonesian core curriculum directly, only a paucity of related studies have been conducted in school and educational contexts during the past ten years (Istikomah et al., 2017; Siswono et al., 2020). No studies have employed objective measurements such as Rasch analysis to evaluate students' inductive reasoning in Indonesia. Rasch analysis was thus utilized in this study to evaluate students' inductive reasoning skills so as to validate an adapted inductive reasoning test in the Indonesian context and classify the difficulty of inductive reasoning items and students' inductive reasoning abilities. It is expected that the current study will form the foundation of research to explore the level of Indonesian students' inductive reasoning and provide information to support teacher and student development.

1.1. Inductive reasoning

Inductive reasoning may be defined as the cognitive activity of generating inferences that meet two criteria: direction and confidence level. In relation to direction, students move from specific observation cases to formulate general principles. With regard to confidence level, students start reasoning from a position of uncertainty to form related hypotheses (Feeney & Heit, 2007; Perret, 2015). Inductive reasoning is a form of reasoning, which can be broadly defined as the process of drawing conclusions that aim to solve problems and arrive at decisions (Lee et al., 2021; Sternberg et al., 2012). Inductive reasoning is concerned with deriving logically sound conclusions from a collection of premises (Feeney & Heit, 2007). In reasoning, one starts from the known to reach and/or evaluate a new conclusion (Sternberg et al., 2012). Inductive reasoning is the process of applying prior information to generate predictions about new cases (Hayes & Heit, 2017). Numerous interpretations of inductive reasoning can be found in various disciplines, including mathematics, philosophy, and psychology. Inductive reasoning is generally considered to be a cognitive process to enable one to generalize the rules from initial observations to arrive at a general conclusion (Csapó, 1997; Stephens et al., 2020).

Inductive reasoning plays a vital role in various cognitive activities such as feature attribution, analogical reasoning, causal reasoning, and probabilistic judgment. Furthermore, it is considered a pivotal element for understanding knowledge on a regular basis and for determining concepts and categories in daily activities (Klauer, 1996). In the inductive process, hypothetical rules are generated to solve unfamiliar problems that can be tested on further action and observation (Perret, 2015). In essence, inductive reasoning plays a role in understanding various kind of knowledge and their application to solve unfamiliar cases. Furthermore, it is included as one of seven factors in mental abilities that describe individual intelligence (Csapó, 1997; Kinshuk et al., 2006; Nikolov &

Csapó, 2018; Perret, 2015)

1.2. The role of inductive reasoning in student development

Inductive reasoning can predict fluid intelligence and crystallized intelligence, which refers to students' ability to solve new problems in working memory (Feeney & Heit, 2007; Perret, 2015). Strobel et al. (2019) demonstrated that fluid intelligence can be measured with an inductive reasoning test. Additionally, student inductive reasoning was also defined as students' ability to elaborate on various insights in one's long-term memory (Perret, 2015). In turn, students' inductive reasoning can affect intelligence in similar ways. Inductive reasoning can be employed to solve new problems and support strategies in solving the same problems in different contexts (Feeney & Heit, 2007). Several studies have also demonstrated that inductive reasoning is an essential predictor of academic achievement and science performance (Van Vo & Csapó, 2020, 2021). In this study, it was assumed that inductive reasoning is a construct or latent factor, and it can be tested empirically using an inductive reasoning test. In the learning context, inductive reasoning plays a vital role in facilitating the learning process because inductive reasoning ability can help a student solve a complex problem. Therefore, assessing students' inductive reasoning is more profitable in education than assessing intelligencen. Many studies have also provided evidence that the higher the students' inductive reasoning ability, the higher their ability in various areas such as natural science, mathematics, attitudes, and languages (Childers & Exemplars, 2020; Kambeyo & Wu, 2018; Lee et al., 2021; Nikolov & Csapó, 2018; Sosa-Moguel & Aparicio-Landa, 2021; Van Vo & Csapó, 2021).

The primary objective of several inductive reasoning studies has focused on gender and grade at school. The majority of studies on student inductive reasoning have examined gender differences. Therefore, findings related to the effect of gender differences on student inductive reasoning are inconsistent in relation to the particular context and culture. Some studies have revealed that the inductive reasoning ability of male students is superior to that of female students (Strobel et al., 2019; Venville & Oliver, 2015). On the contrary, Díaz-Morales and Escribano (2013) found that female students' inductive reasoning abilities were superior to that of male students in predicting school achievement. Several studies also concluded that there were no significant gender differences between females and males in assessing inductive reasoning (Molnár & Csapó, 2011; Kambeyo & Wu, 2018; Kinshuk et al., 2006; Sosa-Moguel & Aparicio-Landa, 2021). Grade levels or age groups also affected students' inductive reasoning. As noted previously, Van Vo and Csapó (2020) demonstrated that students' inductive reasoning ability tended to increase regularly among 5th, 7th, 9th, and 11th grade students in Vietnam. While students' inductive reasoning tended to improve gradually from the 3rd grade (8–9 year-olds) to the 11th grade (16–17 year-olds), those in the 7th grade exhibited rapid development (Csapó, 1997; Díaz-Morales & Escribano, 2013; Molnár & Csapó, 2011; Pásztor et al., 2018; Sosa-Moguel & Aparicio-Landa, 2021). No studies have been conducted on the effect grade has on inductive reasoning in Indonesia.

1.3. Assessing inductive reasoning in the educational context and Rasch analysis

In a meta-analysis, Waschl and Burns (2020) demonstrated that 40 different test types had been employed to measure inductive reasoning. The most common inductive reasoning test was designed to measure reasoning problems related to series completion, analogies, geometric matrices, and classification (Csapó, 1997; Hayes & Heit, 2017; Nikolov & Csapó, 2018; Stephens et al., 2020; Van Vo & Csapó, 2020; Waschl & Burns, 2020; Wu & Molnár, 2018). Series completion tasks require students to determine the relationships in a given completion series such as numbers, letters, objects, and words. Students can solve a series of completion tasks if they can find

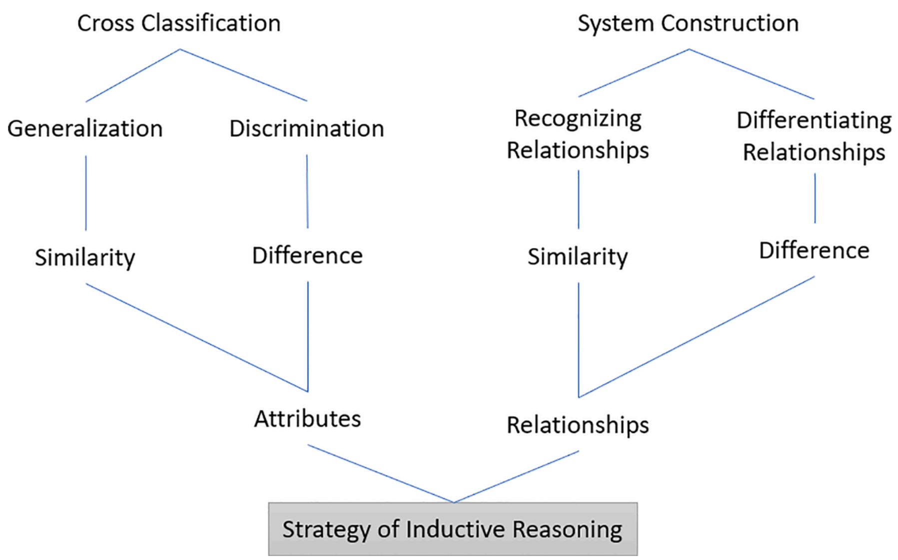

Fig. 1. A genealogy of tasks in inductive reasoning (Klauer & Phye, 2008).

the relation between the given components so as to determine the next component as a solution (Klauer & Phye, 2008; Leighton & Sternberg, 2003; Waschl & Burns, 2020). The analogy task, which involves the structure of a display on an object such as figures, numbers, and letters, can be solved by assessing the sample information in the task. This task is frequently used to measure students' intelligence (Hotulainen et al., 2018; Klauer & Phye, 2008; Stephens et al., 2020; Strobel et al., 2019; Venville & Oliver, 2015). The classification task involves combining various forms of problems that comprise words, figures, and numbers that require students to identify answers that are unrelated to the others. The geometric matric task, in which a set of images is provided in a matrix where the rows and columns have particular rules, requires students to determine relationships and information to find missing images (Csapó, 1997; Klauer, 1996; Klauer & Phye, 2008; Waschl & Burns, 2020). Klauer and Phye (2008) formulated the genealogy of tasks in inductive reasoning to create inductive reasoning items so as to help researchers assess students' inductive reasoning. In Fig. 1, Klauer's diagram that depicts the genealogy of tasks in inductive reasoning is portrayed. This study adapted and translated the inductive reasoning test developed by Csapó (1997) for Indonesian purposes.

Rasch analysis is an objective measurement, developed by the Danish mathematician George Rasch. In Rasch analysis, each item and person are combined into a single scale the called Log odd units (logit) scale (Andrich, 2018). By applying Rasch analysis in the context of inductive reasoning, researchers can investigate a student's inductive reasoning ability with precision measurement, map items on an inductive reasoning test based on difficulty level, and investigate differential item functioning based on various background variables such as test method, gender and grade. Students' inductive reasoning test generates dichotomous data that can be assessed in accordance with the difficulty of each item and a person's personal abilities with the following mathematical equation:

$$log\frac{P_{ni1}}{P_{ni0}} = B_n - D_i$$

where

P_{ni1} or P_{ni0} is the probability that person n encountering item i is observed in category 1 or 0,

B_n is the measure of the ability of person n, and

D_i is the measure of the difficulty of item I, the point where the highest and lowest categories of the item are equally probable (Linacre, 2021b).

The analysis results are used to assess the reliability and validity of the instrument, unidimensionality, respondents' abilities, item difficulties, differential item functioning (DIF), item-respondent map, a comparison of respondents' logits based on gender and grade, and confirmation of outliers or answers that have been guessed (Barbic & Cano, 2016; Khine, 2020).

1.5. Research questions

The purpose of this study was to assess an adapted Indonesian version of the inductive reasoning test by determining its validity and reliability so as to evaluate Indonesian students' inductive reasoning and to classify their inductive reasoning levels by grade and gender. The data were analyzed by employing the Rasch measurement approach. WINSTEPS software was utilized to conduct the analysis. R software was employed with the yarrr (Phillips, 2017). Furthermore, the ggplot2 package (Wickham, 2016) was employed to depict students' inductive reasoning development related to gender and grade. The following four research questions were formulated:

1 In accordance with the Rasch parameters, what is the reliability and validity of the adapted inductive reasoning test?
2 Is DIF detected between paper-based and online-based tests?
3 Is there any DIF based on gender and grade levels?
4 How is the evaluation of Indonesian students' inductive reasoning across grade and gender?
5 What is the classification of the difficulty of inductive reasoning items and students' inductive reasoning abilities when employing Rasch analysis?

Methods

2.1. Participants

A cross-sectional research design with a quantitative method was employed in this study. Stratified random sampling was utilized to select 856 students in the 10^{th} to 12^{th} grade in senior high schools and undergraduate students at universities in West Kalimantan

Table 1
Demographic profiled of participants.

Grade	N	Female/Male ratio (%)	Mean age (years)
10^{th}	231	29.4/70.6	16.02
11^{th}	291	67.7/32.3	17.11
12^{th}	153	41.2/58.8	17.99
Undergraduate students	181	66.3/33.7	19.17
Total	856		

province, Indonesia. Students provided written consent before completing the inductive reasoning test. The anonymity of the students was assured to protect their personal identification. The participants were given 50 minutes to complete the inductive reasoning test under teacher supervision and guidance during regular class time. The demographic profile of the participants is presented in Table 1.

2.2. Instruments

2.2.1. Inductive reasoning test

Csapó's (1997) inductive reasoning test was adapted and employed in this study. The original test comprised four subscales in Hungarian and English. The test has been employed in various empirical studies with different cultural contexts and school-aged samples to establish its reliability and predictive validity. These various cultural contexts include Hungary (Csapó, 1997; Pásztor et al., 2018), Finland (Hotulainen et al., 2018), Namibia (Kambeyo & Wu, 2018), Vietnam (Van Vo & Csapó, 2020), and China (Wu & Molnár, 2018). The adapted inductive reasoning test was translated into Indonesian by two language specialists and comprises two main sections using back-and-forth translation approaches. The first section encompasses a background questionnaire on gender, grade, parents' employment, parents' education, and science and mathematics achievement scores from the previous semester. Only information related to gender and grade was used in this study. The second section included inductive reasoning items in four tasks: number analogies (NA), number series (NS), figural series (FS), and figural analogies (FA). Each subscale comprises ten items. While a correct answer is allocated one point, incorrect answers are not awarded any points. Thus, respondents who answered all the answers correctly score a maximum of 40 points. Participant responses were included in the dataset automatically and in a traditional way into the Statistical Package for Social Sciences (SPSS) dataset before Rasch analysis was performed. Examples of the items in the four tasks are depicted in Fig. 2.

2.3. Procedures

The data collection at both the schools and universities was conducted before the COVID19 pandemic in Indonesia using online and paper-based tests. Csapó, Molnár, & Tóth, 2009 confirmed no bias effect or significant differences between the two modes. However, to ensure this effect, we also employed DIF based on online-based and paper-based tests. The researchers obtained ethical approval from the Institutional Review Board at the University of Szeged. Permission was also sought from the schools and universities to administer the test. The Electronic Diagnostic Assessment System (eDia) platform, which was developed by the Center for Research on Learning and Instruction at the University of Szeged, Hungary (Csapó & Molnár, 2019), was employed to disseminate the test. The eDia platform can be used to implement various types of tests items, such as multiple-choice tests, open-ended questions and drag-and-drop items. The eDia is an easy to use as diagnostic instrument that contains item banks to support teaching and learning in a digital pedagogy system. The participants were able to access the eDia platform through Google Chrome, Mozilla Firefox, and other standard internet browsers. Where the schools and universities lacked the necessary infrastructure and technological support, the paper test was employed. We collaborated with the teachers in observing and giving guidance when finalizing the inductive reasoning test.

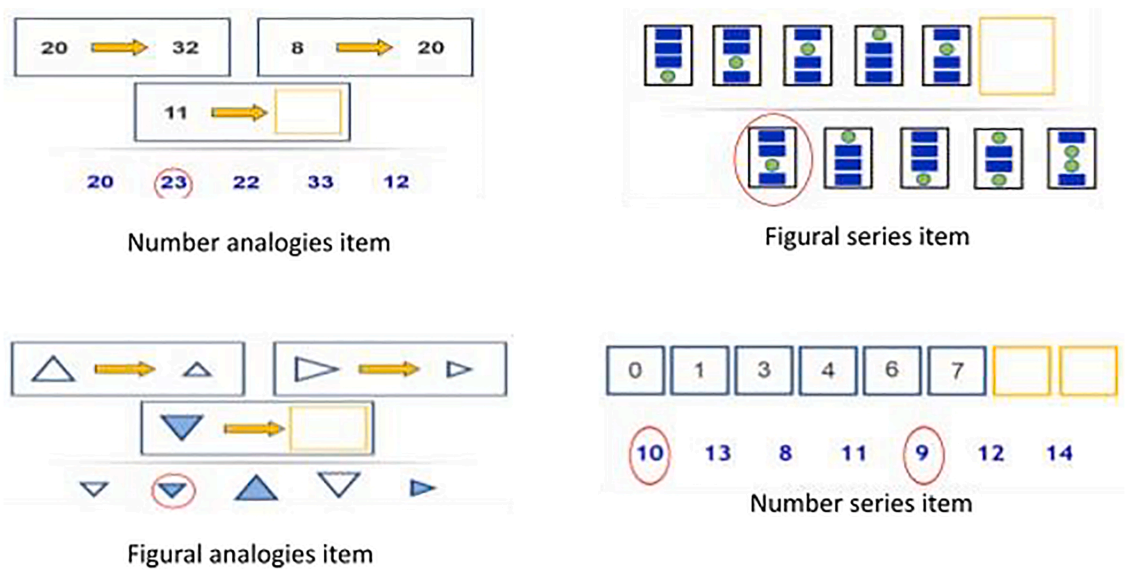

Fig. 2. Sample items on inductive reasoning test based on four tasks.

2.4. Data analysis

SPSS version 25 (IBM Corp, 2017) was employed to perform descriptive statistics, Cronbach's alpha (α), and McDonald's omega (ω). Furthermore, R software version 1.4.1717 (R Core Team, 2020) with graphical packages such as the yarrr (Phillips, 2017) and the ggplot2 (Wickham, 2016) were used to depict the interactive pirate plot of the participants' inductive reasoning development. WINSTEPS version 5.1.4 software (Linacre, 2021a) for Rasch measurement was utilized to perform data analysis. Rasch analysis included conducting Rasch modelling using joint maximum likelihood estimation (JMLE) in which student scores were converted into the logit scale (interval data), ranging from negative infinity to positive infinity. Rasch parameter evaluation was employed to assess the validity and reliability based on unidimensionality and local independence and by checking person and item reliability criteria. The Wright map was presented to confirm targeting criteria between item and person. DIF analysis was used to evaluate item bias in accordance with the test method. The logit value of person (LVP) and logit value of item (LVI) were classified using the COUNTIF function in Microsoft Excel in accordance with mean logits and the standard deviation as recommended by Chan et al. (2020) and Adams et al. (2020). The COUNTIF function was used to perform automatic calculations of the person ability measure and item difficulty measure based on the mean logits and the standard deviation categorization. This process made it possible to achieve accuracy in the grouping of persons and items with a large number of respondents and to reduce human errors when grouping manually.

3. Results

3.1. Validity and reliability of inductive reasoning test

3.1.1. Validity

Rasch analysis was performed by employing JMLE estimation for dichotomous data to validate the inductive reasoning test that had been adapted for Indonesia. The item and person parameters were used to validate the inductive reasoning test. Person and item fit validity were identified in accordance with the mean of infit and outfit mean square (MNSQ), where an acceptable range is from 0.5 to 1.5 even though 1.6 is still regarded as acceptable. Furthermore, the ideal values for fit criteria are close to 1.00 logits (Andrich, 2018; Boone et al., 2014). The infit and outfit z-standardized (ZSTD) of persons and items were ignored because the sample was larger than 500 students (Azizan et al., 2020) could differentiate person abilities as latent traits. In addition, item separation revealed that the inductive reasoning test includes a range of easy and difficult items (Boone et al., 2014). It is imperative that separation values should be more than 2 logits in which the larger the separation index, the more higher the quality of the test (Boone et al., 2014; Fisher, 2007; Planinic et al., 2019). The results of Rasch analysis are presented in Table 2. The results confirmed that inductive reasoning for the reasoning test adapted for Indonesia achieved validity in accordance with the Rasch parameter for each task and entire test. It was considered that the FA task met the person separation threshold, with person separation close to 2 logits.

WINSTEPS software estimates unidimensional Rasch model, but it also can give benefits to multidimensional model by assessing the subtest (Linacre, 2021b). In this study, we evaluated the task as a subtest as a unidimensional model based on a recommendation by Bond and Fox (2015), where the inductive reasoning test was developed to assess an underlying construct that is composed of distinct but related sub-dimensions. Aryadoust and Raquel (2019) also recommended assessing the unidimensionality of the subtest using WINSTEPS when using a test with a multidimensionality model as a basic assumption. Unidimensionality and local independence were assessed to confirm the construct validity of the inductive reasoning test. The values of raw variance by measure for all tasks are presented in Table 2. The results revealed that the reasoning test achieved an acceptable threshold of more than 30%. In addition, the unexplained variance for the first contrasting values was less than 2 for all the tasks that confirmed unidimensionality which indicates the test comprised close to four dimensions based on the tasks. Local independence proves that each item in the inductive reasoning test was independent. The raw residual correlation between pairs was also assessed to determine local independence. The acceptable threshold of the raw residual correlation between pairs of items should be less than 0.3 (Boone et al., 2014). The results revealed that all the items on the inductive reasoning test had a residual correlation ranging from 0.11 to 0.28, which supported the assumption of acceptable criteria for local independence.

Table 2
Summary of Rasch parameters for the inductive reasoning test and for each task.

Psychometrics attribute	Task				IR test
	FA	FS	NA	NS	
Number of items	10	10	10	10	40
Mean					
item outfit MNSQ	0.95	0.98	1.16	1.54	1.01
item Infit MNSQ	1.00	0.98	0.98	.99	1.00
person outfit MNSQ	0.95	.98	1.16	1.13	1.01
person Infit MNSQ	1.00	.99	0.98	0.96	1.00
Item separation	10.27	12.07	13.62	14.79	16.46
Person separation	1.98	2.18	2.25	2.82	2.92
Unidimensionality					
Raw variance by measure	30.2%	36.6%	36.1%	53.7%	
Unexplained variance 1st contrast	1.72	1.97	1.70	2.03	

The interaction between items and students is represented on the Wright map (Fig. 3). The Wright map reveals that the items and students matched the targeting criteria. In other words, all the items covered all students' abilities. The results further demonstrated that all the items in the inductive reasoning test met fit criteria based on the infit MNSQ values, ranging from 0.80 to 1.23 logits. While item NS8 (3.34 logits) was indicated as the most difficult item, FA3 (−2.52 logits) and FS4 (−2.55 logits) were the easiest items.

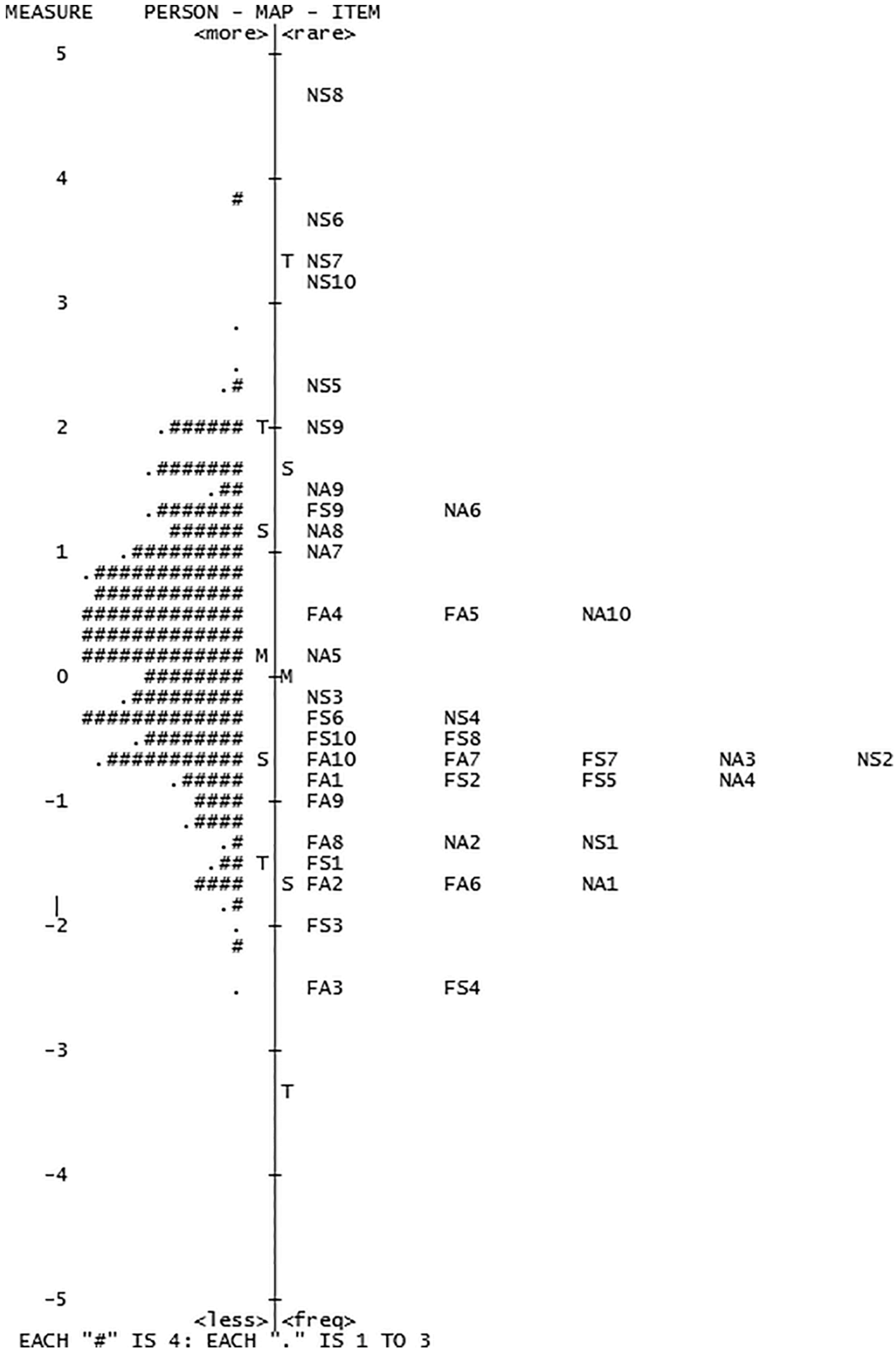

Fig. 3. Wright map.

3.1.2. Reliability

The reliability criteria were evaluated following several indicators, including Rasch parameters using person and item reliability (Fisher, 2007; Linacre, 2021b), Cronbach's alpha (α) (Taber, 2018) and McDonald's omega (ω) (Dunn et al., 2014). WINSTEPS software will generate person reliability, item reliability and Cronbach's Alpha (α), and SPSS was utilized to compute McDonald's omega (ω). Cronbach's Alpha (α) values ranged from 0.61 to 0.77 for all the tasks as well as the entire test, thus indicating sufficient reliability (Taber, 2018), and McDonald's omega (ω) ranges from 0.54 to 0.75, thus confirming acceptable reliability was achieved for only in the test level with 0.75 (Dunn et al., 2014). However, for person reliability and item reliability, the values range from 0.68 to 1.00. Fisher (2007) noted that values more than 0.67 demonstrated acceptable reliability. Overall, the adapted inductive reasoning test and all its tasks exhibited acceptable criteria for the Rasch reliability parameter. All the reliability results for both the tasks and the test are summarized in Table 3.

3.2. DIF between paper-based and online tests

The DIF analysis used in this study was the uniform DIF analysis that compares all ability levels of two or more groups. DIF analysis based on the test method was performed to evaluate whether any item bias between paper-based and online tests was detected in student abilities. DIF analysis indicates participant responses based on subgroups for each item in the test (Bond & Fox, 2015; Boone et al., 2014; Khine, 2020). DIF can be assessed in accordance with two categories: a significant probability ($p < 0.05$) and DIF size. There are 3 DIF size classifications: negligible, slight to moderate ($|\text{DIF}| \geq 0.43$ logits), and moderate to large ($|\text{DIF}| \geq 0.64$ logits) (Zwick et al., 1999). The results of the DIF analysis revealed that three of the 40 items had a significant probability ($p < 0.05$), namely, FS2, FA7, and NS6. However, NS6 had moderate to large DIF. Furthermore, the online test was more difficult for students than the paper-based test with regard to item NS6, with 0.94 logits of DIF size, $p < 0.05$. FS2 and FA7 were classified as having negligible DIF. The DIF analysis based on the test method is illustrated in Fig. 4.

3.3. DIF across grade and gender

The DIF analysis is also able to detect failures of invariance in this study context. As noted in section 3.2, a significant probability ($p < 0.05$) and DIF size were used to identify DIF across grade and gender. Based on gender, four items have $p < 0.05$ with various DIF sizes, FS1(0.23 logits), NS2(0.34 logits), NS7(0.15 logits), NS8(0.16 logits) as presented in the Figure 5. Therefore, we can categorize these four items as negligible DIF. The DIF analysis is also performed based on grade level in Fig. 6. Four items, FA1, NA4, NS3, NS5, have $p < 0.05$. However, these four items have DIF sizes below 0.43 logits. The highest DIF size between grade 10 and grade 12 has 0.42 logits which is still categorized as negligible DIF. We can assume the IR test can hold invariance confirming no DIF issue across grade and gender.

3.4. The evaluation of Indonesian students' inductive reasoning

The students' inductive reasoning based on the tasks of the test was evaluated. Rasch scales the students' ability from negative infinity to positive infinity whereby 0 logits is the average measure of students' ability (Bond & Fox, 2015; Sumintono & Widhiarso, 2014). In general, student abilities in solving the items of the inductive reasoning test were above average level (above 0 logits), with $M = 0.24$ logits; $SD = 0.79$. However, for the NA task ($M = -0.04$; $SD = 0.78$) and NS task ($M = -1.41$; $SD = 0.98$), the students' abilities were below average. These findings revealed that the students encountered some difficulties in solving numeric tasks, especially NS. This finding concurs with that of previous research on lower grades (Csapó, 1997; Díaz-Morales & Escribano, 2013; Molnár & Csapó, 2011; Pásztor et al., 2018; Sosa-Moguel & Aparicio-Landa, 2021; Van Vo & Csapó, 2020). The correlation matrix for all the tasks and the whole test were also evaluated. All correlation values were significant and ranged from 0.16 to 0.76. While the highest correlation was found between the FA task and inductive reasoning test ($r = 0.76$), the lowest correlation was observed between the FS and NS tasks, even though the latter relationship was positively significant. This finding implied that students with a higher score on a task would achieve a higher score on the inductive reasoning test. The students' abilities and correlations between the inductive reasoning test and tasks are summarized in Table 4.

The students' inductive reasoning abilities were also evaluated in accordance by gender and grade. An examination of Table 5 reveals that undergraduate students outperformed the students in other grades; $M = 0.59$; $SD = 0.63$. The 12^{th} grade students had higher logit values ($M = 0.31$; $SD = 0.66$) than the 10^{th} and 11^{th} grade students. Surprisingly, the 10^{th} and 11^{th} graders had the same

Table 3
Reliability indicators.

Reliability	Instrument				
	FA	FS	NA	NS	Test
Item reliability	0.99	0.99	0.99	1.00	1.00
Person reliability	0.69	0.68	0.68	0.71	0.79
Cronbach's alpha (α)	0.70	0.61	0.65	0.62	0.77
McDonald's omega (ω)	0.66	0.58	0.57	0.54	0.75

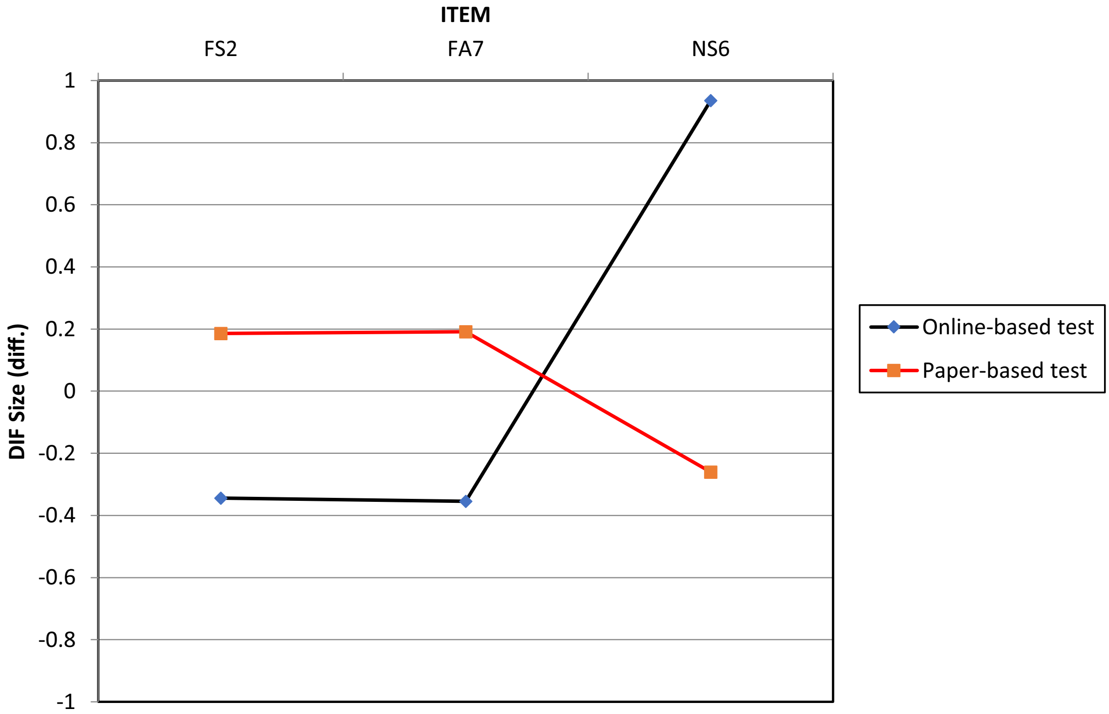

Fig. 4. DIF analysis based on the test method.

logit values. Furthermore, the female students showed superior performances (M = 0.28; SD = 0.88) in solving inductive reasoning problems in comparison to the male students. Person reliability for all the subgroups realized the minimum threshold criteria.

To depict the primary trend between gender and grade related to the development of student inductive reasoning, graphical packages such as the yarrr package (Phillips, 2017) and the ggplot2 package (Wickham, 2016) were employed by using R software to create a pirate plot that combined the boxplot and student logit value distribution. An examination of Fig. 7 reveals that males and females in each group were similar. The logit measure of females and males in all the groups remained stable, between 0 logits to 0.7 logits. No significant gender differences were identified in student inductive development. However, the distributions among the grades depicted different trends. While 10th and 11th grade students had similar inductive reasoning abilities, those in the 12th grade outperformed the former groups. The undergraduate students appeared to have the highest mean logit. However, further evaluation is needed to check any differences in students' inductive reasoning abilities across gender and grades.

An independent t-test was conducted to compare student inductive reasoning abilities between females and males for each grade. We classified the person logit values in accordance with grade and analyzed such by performing an independent t-test to determine gender differences. The results presented in Table 6 confirm that no significant differences were found between males and females for each grade. However, with the exception of the gGrade 11 students, the mean logit values of the females were higher than those of the males.

Moreover, a one-way ANOVA was employed using person logit values to check whether there were any differences among the grades. The results revealed significant differences among the grades for five groups that comprised four tasks and the inductive reasoning test: FA task [$F(3, 855) = 16.35$, $p < 0.001$], FS task [$F(3, 855) = 12.00$, $p < 0.001$], NA task [$F(3, 855) = 4.36$, $p < 0.05$],

Table 4
Result of student abilities and correlation based on inductive reasoning test and tasks.

Test-subscale	M (logits)	SD	Logit range(Min, Max)	Pearson correlation			
				FA	FS	NA	NS
FA	1.16	0.8	(-2.58, 3.97)				
FS	0.98	1.01	(-2.72, 4.31)	.45**			
NA	-0.04	0.78	(-2.76, 4.05)	.36**	.24**		
NS	-1.41	0.98	(-4.25, 4.30)	.17**	.16**	.45**	
IR test	0.24	0.79	(-5.41, 1.69)	.76**	.68**	.74**	.56*

Note. N = 856
*p < .05,
**p < .001, M = mean, SD = standard deviation, IR = inductive reasoning, FA = Ffigural analogies, FS = figural series, NA = number analogies, NS = number series

Table 5
Student inductive reasoning abilities based on gender and grade levels.

	N	M (score)	M (logit)	SD	Person Reliability
Grade					
10	231	21.5	0.09	0.96	0.82
11	291	21.4	0.09	0.99	0.83
12	153	23	0.31	0.66	0.68
Undergraduate students	181	24.7	0.59	0.63	0.67
Gender					
Female	448	22.7	0.28	0.88	0.78
Male	408	22.1	0.19	0.9	0.79

Table 6
The independent t-test for comparing student inductive reasoning abilities between females and males.

Grade	Female		Male		MD (logit)	t	p
	N	M/SD (logit)	N	M/SD (logit)			
10	68	0.17(1.08)	163	0.06(0.92)	0.11	0.74	.46
11	197	0.09(0.96)	94	0.11(1.08)	-0.02	-0.14	.89
12	63	0.35(0.69)	90	0.29(0.64)	0.06	0.11	.62
Undergraduate students	120	0.62(0.56)	61	0.52(0.76)	0.10	0.91	.37

NS task [$F(3, 855) = 6.33$, $p < 0.001$], and entire inductive reasoning test [$F(3, 855) = 15.01$, $p < 0.001$]. To evaluate if there were any significant differences among the grades, WINSTEPS was employ to perform an independent t-test. The results are presented in Table 7. At the test level, significant differences were found among all the grades on the inductive reasoning test, with the exception of the 10^{th} and 11^{th} grade students. In relation to the tasks, the results demonstrated that the older students outperformed the younger ones on all the tasks and the entire test. No significant difference was found between the 10^{th} and 11^{th} graders, thus demonstrating that these two groups of students had similar inductive reasoning abilities. Furthermore, the older students' performance was superior to that of their younger counterparts. Even though some significant differences were identified between tasks and the whole test, almost all mean differences revealed negative values, thus indicating that students in higher levels performed better than those at the lower levels, with the exception of the 10^{th} and 11^{th} grade students on the FA, NA, and NS tasks. The group comparisons for all the tasks and the entire test are presented in Table 7.

3.5. Item difficulty categorization for the inductive reasoning test

In evaluating the difficulties of the inductive reasoning items, LVI analysis was used in accordance with the mean logit of items (0.00 logits), 1SD (1.69 logits), −1SD (−1.69 logits), the mean logit + 2SD (3.38), and the mean logit + 2SD (−3.38). The SD value demonstrated a wide dispersion of logit measures in item difficulty level. Using these thresholds, the inductive reasoning items were classified into 5 categories: difficulty level I (LVI ≥ mean logit + 2SD), difficulty level II (mean logit + 2SD > LVI ≥ 1SD), difficulty level III (1SD > LVI ≥ mean logit), difficulty level IV (mean logit > LVI ≥ −1SD), difficulty level V (LVI < −1SD) (Table 8). The results of the LVI analysis revealed 2 items (5%) in difficulty level I, 5 items (12.5%) in difficulty level II, 8 items (20%) in difficulty level III, 22 items in difficulty level IV, and 3 items (7.5%) in difficulty level V. These difficulty levels may be described as very difficult (difficulty level I), difficult (difficulty level II), moderate (difficulty level III), easy (difficulty level IV), and very easy (difficulty level V).

An examination of Table 8 reveals that two NS items (5%) were classified as very difficult even though four NS items were classified as easy items. One NA item and four NS items were classified as difficult items and two FA items, one FS item, and five NA items were

Table 7
The independent t-test for grade comparison based on the inductive reasoning test and tasks.

Grade	IR test		FA		FS		NA		NS	
	MD (logit)	t	MD (logit)	t	MD (logit)	t	MD (logit)	t	MD (logit)	t
10th & 11th	-0.01	-0.04	0.01	0.05	-0.07	-0.45	0.03	0.23	0.1	0.81
10th & 12th	-0.22	-2.7**	-0.41	-3.23**	-0.5	-3.27**	-0.02	-0.02	-0.07	-0.51
10th & UNS	-0.5	-6.27**	-0.73	-6.09**	-0.83	-5.26**	-0.3	-2.32*	-0.43	-3.7**
11th & 12th	-0.22	-2.78**	-0.42	-3.49**	-0.43	-3**	-0.03	-0.23	-0.17	-1.28
11th & UNS	-0.49	-6.56**	-0.74	-6.6**	-0.76	-5.11**	-0.33	-2.62**	-0.53	-4.61**
12th & UNS	-0.27	-3.81**	-0.32	-2.68**	-0.33	-2.24*	-0.3	-2.1*	-0.36	-2.93**

Note. N = 856
*p < .05,
**p < .001, UNS = undergraduate student, MD = mean differences, IR = inductive reasoning, FA = figural analogies, FS = figural series, NA = number analogies, NS = number series

Table 8
Categorization of inductive reasoning item difficulty.

Task	Difficulty level I, LVI ≥ Mean logit + 2SD	Difficulty level II, Mean logit + 2SD > LVI ≥ 1SD	Difficulty level III, 1SD > LVI ≥ Mean logit	Difficulty level IV, Mean logit > LVI ≥ -1SD	Difficulty level V, LVI < -1SD
FA			FA4, FA5,	FA1, FA2, FA6, FA7, FA8, FA9, FA10	FA3
FS			FS9	FS1, FS2,FS5, FS6, FS7, FS8, FS10	FS3, FS4
NA		NA6	NA5,NA7, NA8, NA9, NA10	NA1, NA2, NA3, NA4	
NS	NS6, NS8	NS5, NS7, NS9, NS10		NS1, NS2, NS3, NS4	

classified as moderate. Finally, one FA item and two FS items were classified as very easy. This result corroborates with Fig. 3 in relation to the Wright map, which conveys item and person scaling. In essence, items in the inductive reasoning test revealed a wide range of difficulty levels in relation to students' abilities. One can assume that the FA and FS items were easier than NS and NA items, which revealed that students' ability to solve figural tasks was more enhanced than their ability to solve numeric tasks. The classification in accordance with the LVI analysis is displayed in Table 8.

3.6. ntudent inductive reasoning abilities categorization

LVP analysis was employed to classify students' ability to solve inductive reasoning problems in accordance with the mean logit of the person (0.24 logits), 1SD (0.89 logits), −1SD (−0.89 logits), the mean logit + 2SD (2.02 logits) and the mean logit + 2SD (−1.54 logits). LVP analysis was conducted in by gender and grade by utilizing the COUNTIF function in Microsoft Excel to perform an automatic calculation of person logit measures. The results of LVP analysis, which resulted in four categories in relation to gender and grade, are presented in Table 9.

The results revealed that ten (1.17%) females and three (0.35%) males were classified as having very high abilities. Furthermore, 237 (27.02%) females were classified as possessing high abilities, 202 (23.60%) males had high abilities. In addition, 191 (22.31%) females and 184 (21.50%) males were classified as having moderate abilities, and ten (1.17%) females and 19 (2.22%) males had low abilities. In relation to grade, 13 (1.52%) students were classified as having very high abilities, six of whom were in Grade 12. In addition, 439 (51.29%) students possessed high abilities. Of these, 37% were undergraduate students. While 375 (43.81%) students were classified as having moderate abilities, 29 students (3.39%) had low ability. Two (0.23%) 10th grade students were classified as having very high abilities, 102 (11.92%) high abilities, 114 (13.32%) moderate abilities, and 13 (1.52%) low abilities. In the 11th grade 6 (0.70%) students possessed very high abilities, 122 (14.25%) high abilities, 147 (17.17%) moderate abilities, and 16 (1.87%) low abilities. It is noteworthy that very few 12th graders and and undergraduate students possessed moderate abilities and none in these groups were classified as having low abilities. Rather, most were classified as having high abilities.

4. Discussion

The results showed that the adapted inductive reasoning test for the Indonesian context is valid and reliable in accordance with Rasch parameters for measuring students in Grade 10, 11, and 12 students at senior high school as well as undergraduate students. One may deduce that the inductive reasoning test can be employed among a wide range of grades in different cultural and country contexts. As noted previously, (Van Vo and Csapó (2020) found an adaptation of the test was valid and reliable among 5th, 7th, 9th, and 11th graders in Vietnam. Furthermore, Csapó (1997) used a paper-based version of the inductive reasoning test to assess the inductive reasoning of 3rd to 11th grade students in Hungary.

Moreover, DIF analysis based on the test method identified only one of the 40 items had a moderate to large DIF, thus implying that the items in the adapted version measured students' inductive reasoning, without the test method or media bias. This finding concurs with Csapó, Molnár, & Tóth, 2009 who revealed no media bias was found when paper-based and online tests were compared. The

Table 9
The categorisation of student inductive reasoning abilities.

Demographics	Very high, LVP > Mean Logit + 2SD	High, Mean Logit + 2SD ≥ LVP > Mean Logit	Moderate, Mean Logit ≥ LVP > Mean Logit - 2SD	Low, LVP < Mean Logit - 2SD
Gender				
Female	10	237	191	10
Male	3	202	184	19
Total	13	439	375	29
10th grade	2	102	114	13
11th grade	6	122	147	16
12th grade	2	74	77	0
Undergraduate student	3	141	37	0
Total	13	439	375	29

application of technology through online testing was supported in this study because technology can offer several benefits for teachers, including developing item banks, using the adaptive test, composing anchoring tests, and collecting data continuously from a large sample. Csapó & Molnár, 2019 described the eDia platform in which more than 25,000 innovative (multimedia-supported) items were created and the system has been operated in an experimental model in over 1,000 schools.

The evaluation of Indonesian students' inductive reasoning revealed that students tended to achieve higher performances on figural tasks with positive logits than numeric tasks. This finding concurs with previous studies (Feeney & Heit, 2007; Roberts et al., 2000; Van Vo & Csapó, 2020). The results further revealed that the females outperformed the males by 0.09 logits. However, no significant differences were found between females and males in all the grades (Table 6). These results concur with previous studies in Vietnam (Van Vo & Csapó, 2020), Namibia (Kambeyo & Wu, 2018), and Spain (Díaz-Morales & Escribano, 2013). In relation to grades, an independent t-test to compare grades also revealed significant differences among grades, with the exception of those in Grade 10 and 11. Undergraduate students had higher inductive reasoning abilities than the other groups. These findings seem to support previous studies related to a comparison of student inductive reasoning among grades (Csapó 1997; Csapó, Molnár, & Tóth, 2009, 2019; Díaz-Morales & Escribano, 2013; Stephens et al., 2020; Van Vo & Csapó, 2020; Wu & Molnár, 2018). The evaluation of student inductive reasoning further revealed that the development of students' inductive development slowed after 14 years of age.

The classification ofitem difficulty showed that while most of the numeric items were classified as very difficult and difficult in accordance with LVI analysis, most of the figural tasks were classified as moderate, easy, and very easy. This finding is in line with Van Vo and Csapó's (2020) evaluation of inductive reasoning that revealed students experienced difficulty solving numeric items: less than 40% of the answers were correct. The finding from

Van Vo and Csapó (2020) and Kambeyo & Wu (2018) also confirm that the figural tasks were relatively easy to solve compared to the numerics task. For instance, NS8 is the most difficult item in the numeric tasks based on Rasch scaling because this item requires complex pattern calculations with large numbers. Meanwhile, FS4 which is the easiest item in figural tasks only requires students to rotate the circle with a simple figure based on the previous pattern. Based on LVP analysis, students' inductive reasoning abilities were classified into four categories. The results revealed that 439 (51,28%) were classified as having high abilities and 375 (43.8%) moderate abilities. This finding confirmed that students in higher grades could solve student inductive reasoning problems well. This is in line with previous studies (Csapó, Hotulainen, Pásztor, & Molnár, 2019; Díaz-Morales & Escribano, 2013; Venville & Oliver, 2015).

This study contributes to the assessment of inductive reasoning using the Rasch measurement approach. A comprehensive analysis and application of the inductive reasoning assessment would extend the practical use of objective measurement in the educational field and encourage other researchers to explore inductive reasoning assessment in different contexts. Investigating person and item interactions based on individual-level statistics has allowed researchers to improve instrument quality and compared the result at the item level. This study also provides the item difficulty classification on the IR test. The person's ability measures was represented across grade and gender. Specific group comparison was also represented for four different tasks and the whole test. The DIF test performed in this study can examine the bias issue or failures of invariances in the assessment context.

Educators and teachers should be aware to identify student inductive reasoning that is related to their future academic achievement. The development of thinking skills in the learning process is embedded in the Indonesian core curriculum whereby inductive reasoning tests is also used as the primary test to examine student thinking skills in various job application. Additionally, inductive reasoning can promote their cognitive development, develop their intelligence, and facilitate the understanding of the application of basic knowledge. Therefore, teachers and educators must understand the importance of inductive reasoning assessment and improve student inductive reasoning in the learning process.

5. Conclusions

In conclusion, the results of this study have contributed to an understanding of the item and person interaction on inductive reasoning. The adapted inductive reasoning test was shown to be valid and reliable in Indonesia and other countries, thus indicating this instrument can be employed in a wide range of cultural contexts. The items on the test are free of bias and only NS6 had a moderate to large DIF. Even though females outperformed males in relation to inductive reasoning abilities, no significant gender differences were found among the grades. Significant differences were found among all the groups, with the exception of the 10^{th} and 11^{th} grades. The classification of the difficulty of items revealed a wide range of difficulty levels, where numeric items were more difficult than figural items. Most of the students were classified as having high or moderate abilities.

The findings in this study provided initial information related to Indonesian students' inductive reasoning ability. This information may be useful for teachers and researchers to predict student success rates in other related subjects such as mathematics and science. In the Indonesian 2013 national curriculum, inductive reasoning is not included clearly. We thus believe that inductive reasoning skills can be embedded and trained in a wide range of grades because inductive reasoning is an essential thinking skill for predicting student academic achievement. We believe this study may be the first to perform different tests and utilize the Rasch measurement to assess students' inductive reasoning in Indonesia.

5.2. Limitations and recommendations for future study

The study has several limitations. First, the sample size was small in comparison to the Indonesian student population. While endeavors were made to include all grades, only those in the West Kalimantan province were included. Therefore, the results cannot be generalized to different contexts. Second, a cross-sectional research design employing quantitative methods was employed to evaluate

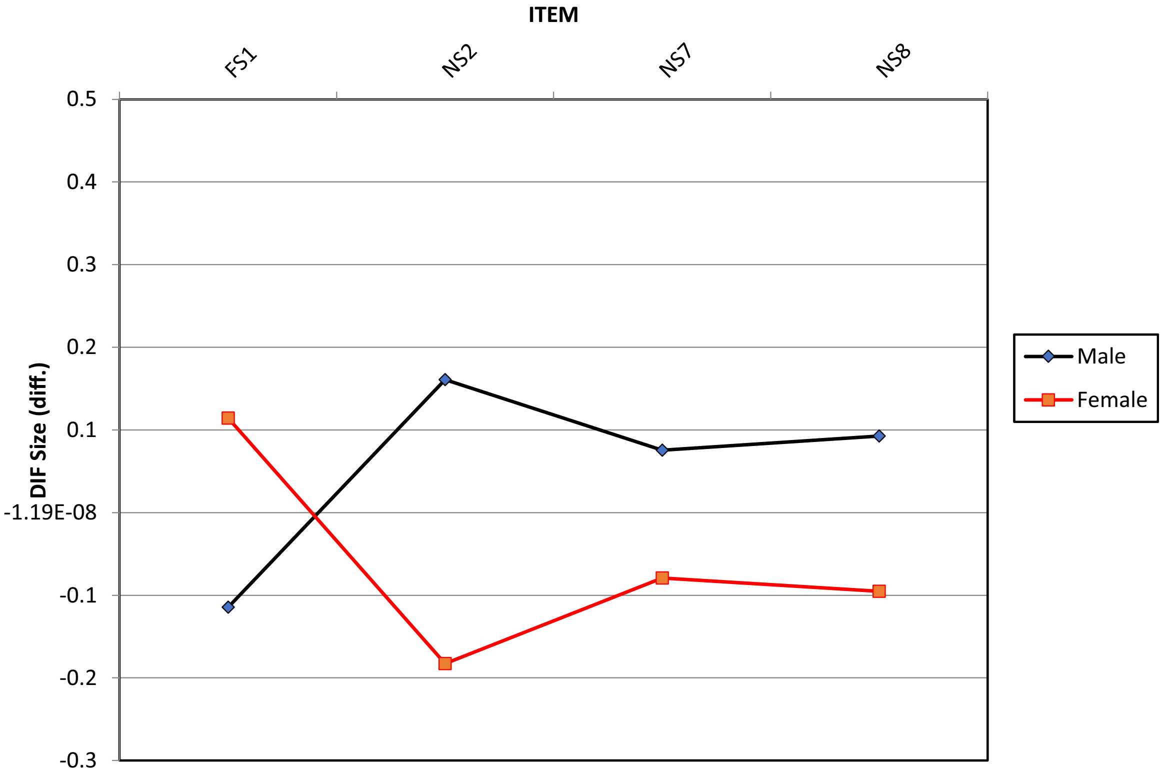

Fig. 5. DIF analysis across gender.

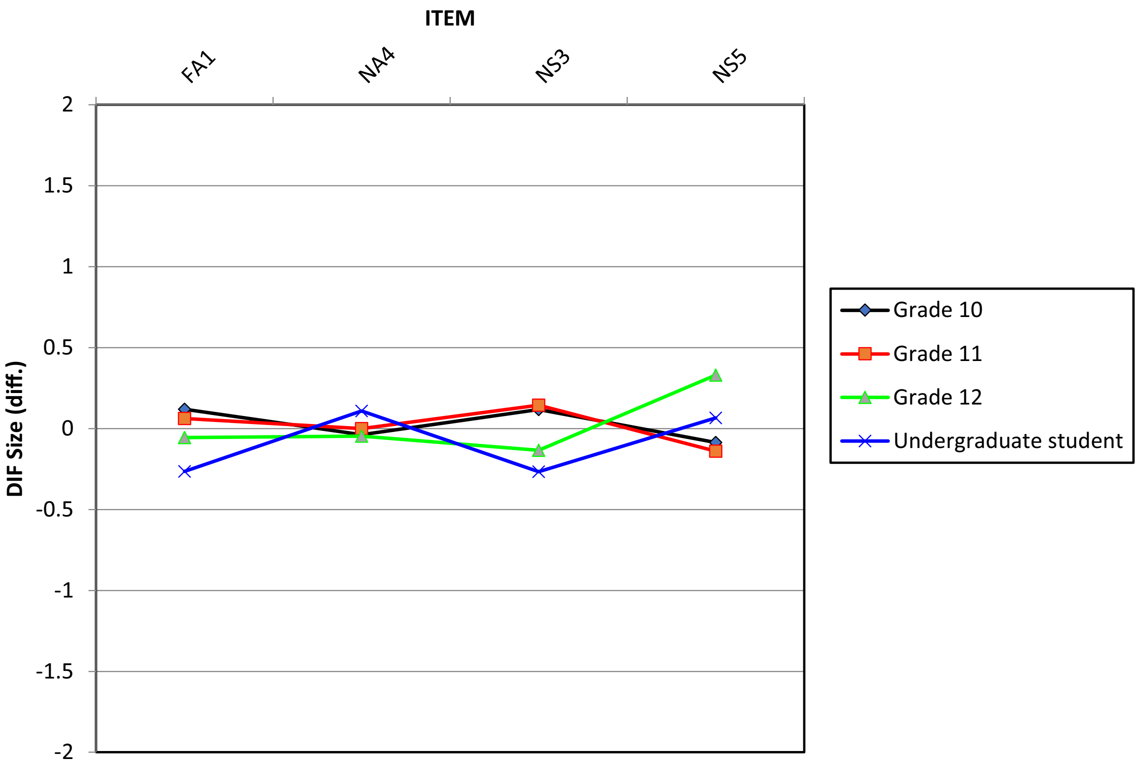

Fig. 6. DIF analysis across grade.

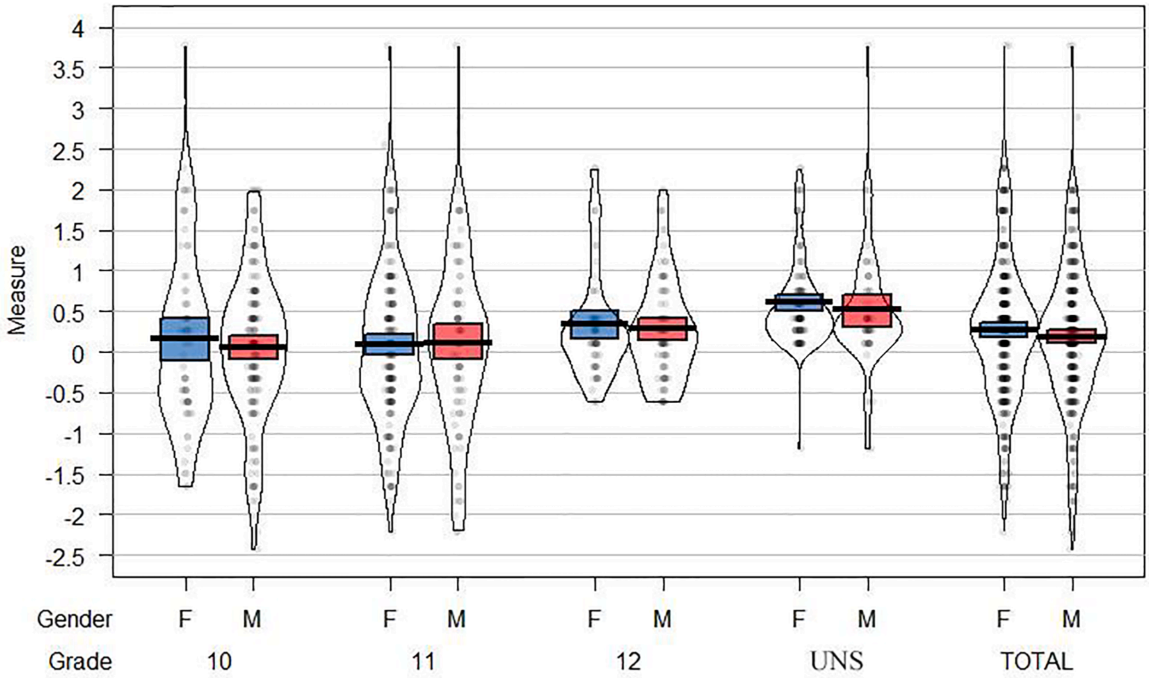

Fig. 7. Pirate plot for comparing student measure (logit) based on gender and grade.

student inductive reasoning abilities. It is recommended that future studies employ a longitudinal research design and include mixed methods to investigate comprehensive student thinking skills. Third, the adapted inductive reasoning test only covered four types of tasks: FS, FA, NS, and NA. It is recommended that future studies include a wide range of inductive reasoning subscales to acquire in-depth information.

CRediT authorship contribution statement

Soeharto Soeharto: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Benő Csapó:** Writing – original draft, Writing – review & editing.

References

Adams, D., Joo, M. T. H., Sumintono, B., & Pei, O. S. (2020). Blended learning engagement in higher education institutions: A differential item functioning analysis of students' backgrounds. *Malaysian Journal of Learning and Instruction, 17*(1), 133–158.

Andrich, D. (2018). Advances in social measurement: A Rasch measurement theory. *Perceived Health and Adaptation in Chronic Disease*, 66–91. Routledge.

Aryadoust, V., & Raquel, M. (2019). *Quantitative data analysis for language assessment volume I: Fundamental techniques*. Routledge.

Azizan, N. H., Mahmud, Z., & Rambli, A. (2020). Rasch Rating Scale Item Estimates using Maximum Likelihood Approach: Effects of Sample Size on the Accuracy and Bias of the Estimates. *International Journal of Advanced Science and Technology, 29*(4), 2526–2531.

Badaruddin, K., & Hawi, A. (2022). Assessment of Student Attitudes in the 2013 Curriculum: Its Implementation and Problems. *Webology, 19*(1), 6408–6419.

Barbic, S. P., & Cano, S. J. (2016). The application of Rasch measurement theory to psychiatric clinical outcomes research: Commentary on… Screening for depression in primary care. *BJPsych Bulletin, 40*(5), 243–244.

Bond, T., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd Edition). Routledge. https://doi.org/10.4324/9781315814698

Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, (4), 15. https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer.

Chan, S. W., Looi, C. K., & Sumintono, B. (2020). Assessing computational thinking abilities among Singapore secondary students: a Rasch model measurement analysis. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-020-00177-2. *0123456789*.

Childers, J. B., & Exemplars, M. (2020). Language and Concept Acquisition from Infancy Through Childhood. *Language and Concept Acquisition from Infancy Through Childhood*. https://doi.org/10.1007/978-3-030-35594-4

Chu, S. K. W., Reynolds, R. B., Tavares, N. J., Notari, M., & Lee, C. W. Y. (2017). Twenty-First Century Skills and Global Education Roadmaps. *21st Century Skills Development Through Inquiry-Based Learning* (pp. 17–32). Singapore: Springer. https://doi.org/10.1007/978-981-10-2481-8_2

Csapó, B. (1997). The Development of Inductive Reasoning: Cross-sectional Assessments in an Educational Context. *International Journal of Behavioral Development, 20* (4), 609–626. https://doi.org/10.1080/016502597385081

Csapó, B., Hotulainen, R., Pásztor, A., & Molnár, G. (2019). Az induktív gondolkodás fejlődésének összehasonlító vizsgálata: online felmérések Magyarországon és Finnországban. *Neveléstudomány, 2019*(3–4), 5–24. https://doi.org/10.21549/NTNY.27.2019.3.1

Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, (JULY), 10. https://doi.org/10.3389/fpsyg.2019.01522

Csapó, B., Molnár, G., & Tóth, K. R (2009). Comparing paper-and-pencil and online assessment of reasoning skills: A pilot study for introducing eletronic testing in large-scale assessment in Hungary. *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Luxemburg: Office for Official Publications of the European Communities (Issue 2).

De Koning, E., Hamers, J. H. M., Sijtsma, K., & Vermeer, A. (2002). Teaching inductive reasoning in primary education. *Developmental Review, 22*(2), 211–241. https://doi.org/10.1006/drev.2002.0548
Díaz-Morales, J. F., & Escribano, C. (2013). Predicting school achievement: The role of inductive reasoning, sleep length and morningness–eveningness. *Personality and Individual Differences, 55*(2), 106–111. https://doi.org/10.1016/j.paid.2013.02.011
Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*, 399–412. https://doi.org/10.1111/bjop.12046
Feeney, A., & Heit, E. (2007). *Inductive Reasoning Experimental, Developmental, and Computational Approaches*. Cambridge: University Press.
Fisher, W. P. J. (2007). Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions, 21*(1), 1095. http://www.rasch.org/rmt/rmt211m.htm.
Hasan, S. H. (2013). History education in curriculum 2013: A new approach to teaching history. *Historia: Jurnal Pendidik Dan Peneliti Sejarah, 14*(1), 163. https://doi.org/10.17509/historia.v14i1.2023
Hayes, B. K., & Heit, E. (2017). Inductive reasoning 2.0. *Wiley Interdisciplinary Reviews: Cognitive Science, 9*(3), 1–13. https://doi.org/10.1002/wcs.1459
Hotulainen, R., Pásztor, A., Kupiainen, S., Molnár, G., & Csapó, B. (2018). Entering school with equal skills? A two-country comparison of early inductive reasoning. In *August Paper Presented at the 9th Biennial Conference of EARLI SIG 1: Assessment and Evaluation: Assessment & Learning Analytics*. Paper: C_2_3.
IBM Corp. (2017). *IBM SPSS Statistics for Windows (Version 25) [Computer software]*. Armonk, NY: IBM SPSS Corp.
Istikomah, F., Rochmad, R., & Winarti, E. R. (2017). Analysis of 7th Grade Students' Inductive Reasoning Skill in PBL-Bertema Model Towards Responsibility Character. *Unnes Journal of Mathematics Education, 6*(3), 345–351. https://doi.org/10.15294/ujme.v6i3.17600
Kambeyo, L., & Wu, H. (2018). Online assessment of students' inductive reasoning skills abilities in Oshana Region, Namibia. *International Journal of Educational Sciences, 21*(1), 1–12, 11. 258359/KRE-86.
Khine, M. S. (2020). *Rasch Measurement*. Springer Singapore.
Kinshuk, Lin, T., & Mcnab, P (2006). Cognitive trait modelling: The case of inductive reasoning ability. *Innovations in Education and Teaching International, 43*(2), 151–161. https://doi.org/10.1080/14703290600650442
Klauer, K. J. (1996). Teaching inductive reasoning: some theory and three experimental studies. *Learning and Instruction, 6*(1), 37–57. https://doi.org/10.1016/S0959-4752(96)80003-X
Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*(1), 85–123. https://doi.org/10.3102/0034654307313402
Kleppang, A. L., Steigen, A. M., & Finbråten, H. S. (2020). Using Rasch measurement theory to assess the psychometric properties of a depressive symptoms scale in Norwegian adolescents. *Health and Quality of Life Outcomes, 18*(1), 1–8. https://doi.org/10.1186/s12955-020-01373-5
Korom, E., Németh, M., Pásztor, A., & Csapó, B. (2017). Relationship between scientific and inductive reasoning in grades 5 and 7. In *17thBiennial Conference of the European Association for Research on Learning and Instruction (EARLI)* (pp. 128–129).
Lafraire, J., Rioux, C., Hamaoui, J., Girgis, H., Nguyen, S., & Thibaut, J. P. (2020). Food as a borderline domain of knowledge: The development of domain-specific inductive reasoning strategies in young children. *Cognitive Development, 56*(June), Article 100946. https://doi.org/10.1016/j.cogdev.2020.100946
Lee, Y., Kim, M. L., & Hong, S. (2021). Big-data analytics: Exploring the well-being trend in South Korea through inductive reasoning. *KSII Transactions on Internet and Information Systems, 15*(6), 1–16. https://doi.org/10.3837/tiis.2021.06.003
Leighton, J. P., & Sternberg, R. J. (2003). *The Nature of Reasoning*. Cambridge University Press. https://doi.org/10.1017/CBO9780511818714 (J. P. Leighton & R. J. Sternberg (eds.)).
Linacre, John M. (2021a). *Winsteps® (Version 5.1.4) [Computer Software]*. Winsteps.com. www.winsteps.com/.
Linacre, John M. (2021b). *Winsteps® Rasch measurement computer program User's Guide*. Winsteps.com..
Molnár, G., & Csapó, B. (2011). *Az 1–11 évfolyamot átfogó induktív gondolkodás kompetenciaskála készítése a valószínűségi tesztelmélet alkalmazásával [Constructing inductive reasoning competency scales for years 1–11 using IRT models], 111*(2), 127–140.
Natsir, Y., Qismullah Yusuf, Y., & Fiolina Nasution, U. (2018). The Rise and Fall of Curriculum 2013: Insights on the Attitude Assessment from Practicing Teachers. *SHS Web of Conferences, 42*, 00010. https://doi.org/10.1051/shsconf/20184200010
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. https://doi.org/10.1016/j.system.2017.11.001
OECD. (2020). *Science performance (PISA) (indicator)*. 10.1787/91952204-en.
Pásztor, A., Kupiainen, S., Hotulainen, R., Molnár, G., & Csapó, B. (2018). Comparing Finnish and Hungarian fourth grade students' inductive reasoning skills. In *9th Biennial Conference of EARLI SIG 1: Assessment and Evaluation: Assessment & Learning Analytics*. *1*, Paper: A_1_3.
Perret, P. (2015). Children's Inductive Reasoning: Developmental and Educational Perspectives. *Journal of Cognitive Education and Psychology, 14*(3), 389–408. https://doi.org/10.1097/NNR.0b013e31824798ba
Phillips, N. D. (2017). Yarrr! The pirate's guide to R. *APS Observer, 30*(3).
Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research, 15*(2), 20111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111
Prastowo, A., & Fitriyaningsih, F. (2020). Learning Material Changes as the Impact of the 2013 Curriculum Policy for the Primary School/Madrasah Ibtidaiyah. *Edukasia : Jurnal Penelitian PendidikanIslam, 15*(2), 251. https://doi.org/10.21043/edukasia.v15i2.7947
R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.
Roberts, M. J., Welfare, H., Livermore, D. P., & Theadom, A. M. (2000). Context, visual salience, and inductive reasoning. *Thinking & Reasoning, 6*(4), 349–374. https://doi.org/10.1080/135467800750038175
Siswono, T. Y. E., Hartono, S., & Kohar, A. W. (2020). Deductive or Inductive? Prospective Teachers' Preference of Proof Method on An Intermediate Proof Task. *Journal on Mathematics Education, 11*(3), 417–438. https://doi.org/10.22342/jme.11.3.11846.417-438
Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia, 8*(2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649
Sosa-Moguel, L., & Aparicio-Landa, E. (2021). Secondary school mathematics teachers' perceptions about inductive reasoning and their interpretation in teaching. *Journal on Mathematics Education, 12*(2), 239–256. https://doi.org/10.22342/JME.12.2.12863.239-256
Stephens, R. G., Dunn, J. C., Hayes, B. K., & Kalish, M. L. (2020). A test of two processes: The effect of training on deductive and inductive reasoning. *Cognition, 199* (February), Article 104223. https://doi.org/10.1016/j.cognition.2020.104223
Sternberg, R. J., Sternberg, K., & Mio, J. (2012). *Cognitive psychology*. Cengage Learning Press.
Strobel, A., Behnke, A., Gärtner, A., & Strobel, A. (2019). The interplay of intelligence and need for cognition in predicting school grades: A retrospective study. *Personality and Individual Differences, 144*(March), 147–152. https://doi.org/10.1016/j.paid.2019.02.041
Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Trim Komunikata Publishing House.
Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2
Tavakol, M., & Dennick, R. (2013). Psychometric evaluation of a knowledge based examination using Rasch analysis: An illustrative guide: AMEE Guide No. 72. *Medical Teacher, 35*(1), e838–e848. https://doi.org/10.3109/0142159X.2012.737488
Van Vo, D., & Csapó, B. (2020). Development of inductive reasoning in students across school grade levels. *Thinking Skills and Creativity, 37*(March), 1–16. https://doi.org/10.1016/j.tsc.2020.100699
Van Vo, D., & Csapó, B. (2021). Exploring students' science motivation across grade levels and the role of inductive reasoning in science motivation. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-021-00568-8
Venville, G., & Oliver, M. (2015). The impact of a cognitive acceleration programme in science on students in an academically selective high school. *Thinking Skills and Creativity, 15*, 48–60. https://doi.org/10.1016/j.tsc.2014.11.004

Waschl, N., & Burns, N. R. (2020). Sex differences in inductive reasoning: A research synthesis using meta-analytic techniques. *Personality and Individual Differences, 164*(February), Article 109959. https://doi.org/10.1016/j.paid.2020.109959
Wickham, H. (2016). *Data analysis. In ggplot2* (pp. 189–201). Springer.
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.
Zhu, X., & Neupert, S. D. (2021). Dynamic awareness of age-related losses predict concurrent and subsequent changes in daily inductive reasoning performance. *British Journal of Developmental Psychology, 39*(2), 282–298. https://doi.org/10.1111/bjdp.12344
Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*(1), 1–28.