

---

*Research articles***The construct validity of mathematical reasoning and proof test instrument integrated with GeoGebra: Second-order confirmatory factor analysis****Yurizka Melia Sari<sup>1</sup>, Heri Retnawati<sup>2</sup>, Shofan Fiangga<sup>1</sup>**

**Abstrak** Instrumen penalaran dan pembuktian matematika merupakan alat untuk menilai kemampuan penalaran siswa dalam memecahkan masalah pembuktian. Pentingnya penalaran dan pembuktian dalam matematika didokumentasikan dalam keterampilan yang harus dikuasai siswa dalam beberapa kurikulum. Akan tetapi, terdapat kesenjangan penilaian penalaran dan pembuktian matematis di Indonesia, yaitu belum adanya instrumen yang memenuhi validitas konstruk sehingga instrumen yang dikembangkan hanya menegaskan interpretasi guru dengan menitikberatkan pada masalah matematika yang mengukur pengetahuan siswa. Oleh karena itu, ada kebutuhan untuk menunjukkan validitas konstruk item dalam penilaian penalaran dan pembuktian matematika. Untuk memenuhi kebutuhan tersebut, penelitian ini bertujuan mengembangkan dan menentukan validitas konstruk instrumen yang dikembangkan. Terdapat empat aspek instrumen penalaran dan pembuktian matematika terintegrasi GeoGebra, yaitu CMR (*Creative Mathematical Reasoning*), IR (*Imitative Reasoning*), FP (*Formal Proof*), dan EP (*Empirical Proof*). Data diperoleh dengan melakukan pengujian tes terhadap 300 siswa SMA di Jawa Timur, Indonesia. Data dianalisis menggunakan analisis faktor konfirmatori orde kedua (CFA) berbantuan perangkat lunak Lisrel 8.80. Hasil penelitian menunjukkan bahwa, dari semua item, sebanyak 8 item valid atau unidimensional, dengan nilai  $t > 1,96$  dan nilai faktor loading  $> 0.5$ . Hal ini menunjukkan bahwa parameter butir soal tersebut unidimensi sehingga dapat mengukur komposisi penalaran dan pembuktian matematis.

**Kata kunci** *Penalaran dan pembuktian, Validitas konstruk, Analisis faktor*

**Abstract** Mathematical reasoning and proof instruments assess students' reasoning in solving proof problems. The importance of reasoning and proof in mathematics is documented in skills that students need to be developed in several curricula. However, there are some issues in the assessment of mathematical reasoning and proof in Indonesia. One of them is no instruments that meet construct validity. The existing instruments only confirm the teacher's interpretation by focusing on mathematical problems that measure students' knowledge. Therefore, there is a need to determine the construct validity of items in assessing mathematical reasoning and proof. For this need, this research aims to develop and evaluate the construct validity of the developed instrument. There are four aspects of a mathematical reasoning and proof instrument integrated with GeoGebra, namely CMR (*Creative Mathematical Reasoning*), IR (*Imitative Reasoning*), FP (*Formal Proof*), and EP (*Empirical Proof*). The data was obtained by conducting tests on 300 high school students in East Java, Indonesia. Second-order confirmatory factor analysis (CFA) was used to analyze the data using Lisrel 8.80 software. The results showed that the developed eight items are valid or unidimensional, with a  $t$ -value  $> 1.96$  and a loading factor value  $> 0.5$ . This reveals that the parameter of the item is unidimensional; hence, it can measure students' mathematical reasoning and proof.

**Keywords** *Reasoning and Proof, Construct validity, Confirmatory factor analysis*

---

<sup>1</sup>Department of Mathematics Education, Universitas Negeri Surabaya, Indonesia, [yurizkasari@unesa.ac.id](mailto:yurizkasari@unesa.ac.id)

<sup>2</sup>Department of Mathematics Education, Universitas Negeri Yogyakarta, Indonesia

## Introduction

Reasoning and proof are essential components in learning mathematics. With mathematical reasoning, students can make conjectures and then compile evidence and manipulate mathematical problems and draw conclusions correctly and appropriately (Baylis, 1983; Stylianides, 2009; Buchbinder & McCrone, 2022). The importance of reasoning and proof skills in mathematics is listed in the abilities students need to master in several school curricula at every level of education (National Council of Teachers of Mathematics, 2000, 2009; Stylianides & Stylianides, 2017). However, students still have difficulties solving mathematical problems related to reasoning and proof (Sari et al., 2020; Ginting et al., 2018). This could be caused by the learning and assessment processes in schools that only provide examples and at the end of the lesson, they are given practice questions that tend only to memorize and apply formulas (Chotimah et al., 2020; Fiangga, 2014). This results in students' inability to solve problems related to reasoning and proof, such as understanding examples, counterexamples, and special cases (Sevimli, 2018; Stylianides, 2019). It is acknowledged that reasoning and proof assessment should give information about students' capacity to engage with mathematical processes and evaluate topic knowledge, such as justification and proof tasks (Maoto et al, 2018; Thompson, 2012). This means that there is a gap between theory and practice in assessing mathematics learning, especially the use of routine problems in assessing students' knowledge.

In fact, Indonesian students still have low reasoning abilities indicated by the results of the Program for International Student Assessment (PISA). The Organization for Economic Cooperation and Development (OECD) (OECD, 2019) defines mathematical literacy as an individual's capacity to formulate, use and interpret mathematics in various contexts, including mathematical reasoning. PISA divides mathematics proficiency into six levels. The level is ordered according to the scores achieved by countries on tests administered by PISA. Students are said to be capable of reasoning if their scores are levels 3-6. According to PISA results, Indonesia ranked 74th in 2018, or 6th place from the bottom. Indonesian students were ranked 73rd in mathematics literacy with 379 points (OECD, 2019). This means that Indonesian students have inadequate reasoning skills. At this level, students are only able to answer questions belonging to familiar contexts where all relevant information is presented and the questions are clearly defined.

In addition, many high school students have difficulties constructing and understanding evidence. Fu et al. (2022) show that students do not understand the meaning or purpose of proof, cannot distinguish proven or unproven empirical examples, lack knowledge of concepts, definitions, and notations and are not familiar with proof strategies. It includes how to start proving and a metacognitive strategy to observe their progress while doing the proof. Reiss et al. (2008) reveal that many students face severe difficulties in consistent reasoning and argumentation, especially in a mathematical proof. However, it is well captured that students at all levels have difficulties with proofs (Hemmi et al., 2013; Noto et al., 2019), especially in understanding the role of examples, counterexamples, and specific cases (Harel & Sowder, 2007; Doruk, 2019; Sevimli, 2018). Students are expected to develop reasoning by constructing new ideas so that mathematical problems can be solved with new answers that most students do not commonly use.

The development of students' mathematical reasoning and proof in schools requires related instruments to examine the skills. Some studies on the topic have also been carried out by Seah and Horne (2020), which focused on the geometric reasoning test item. Mumu and Tanujaya

(2019) developed a test instrument to measure reasoning skills from mathematics routine and non-routine tasks. Saeedullah (2021) also developed an instrument to measure mathematical reasoning in senior high school using five constructs of mathematical reasoning, such as mathematical inductive reasoning, deductive reasoning, generalization, adaptive reasoning, and problem-solving. In recent years, research on the assessment of mathematical reasoning and proof has been more on developing instruments that focus on revealing reasoning abilities, especially for mathematics students or teachers (Stylianides & Stylianides, 2009; Akkurt & Durmus, 2022; Sari, 2017). However, there is yet an instrument that reveals both of reasoning and proof ability. This can be said that there is no standardized instrument that can be used as a reference for reasoning and proof test instrument. Due to the lack of standardized references, data obtained from test scores should be analyzed by Item Response Theory (IRT) in order to obtain more accurate data information. In fact, Item response theory (IRT) is concerned with accurate test scoring and the development of test items better than Classical Test Theory (CTT) (An & Yung, 2014). To address this, research on a test that can effectively explain students' reasoning and proof with a valid instrument based on analysis with IRT and be utilized in evaluating students' reasoning and proof based on norms is required. There is also a need for research on mathematical reasoning and proof assessment that can be used in Indonesian high schools.

## **The Framework of Reasoning and Proof**

### **Reasoning schemes**

One way to test students' mathematical reasoning skills is to understand the arguments students use to draw a conclusion (Fischer et al., 2020; Hidayat & Prabawanto, 2018). The arguments students generated stem from their choices to employ strategies in solving a given mathematics problem or task (Lithner, 2008; Sumpter, 2013). It allows teachers and educational practitioners to identify their students' reasoning structures. This information can be used to enhance teaching guidelines and mathematics learning assessments.

There are different views about the construct of reasoning. Haylock and Thangata (2007) define two constructs of reasoning, namely inductive reasoning and deductive reasoning. Lithner (2008) developed the construct of students' mathematical reasoning based on the results of students' conclusion-drawing arguments, namely creative mathematical reasoning and imitative mathematical reasoning. Creative mathematical reasoning (CMR) needs to meet the conditions of novelty, reasonable or acceptable (plausibility) and based on mathematical knowledge. In comparison, imitative reasoning (IR) has a relationship with memory reasoning, where students choose strategies by using answers without any consideration and algorithmic reasoning, where students use strategies with a set of mathematical rules in solving given problems.

In this study, we limited the constructs of mathematical reasoning to creative mathematical reasoning (CMR) and imitative reasoning (IR) in developing an instrument model for assessing the reasoning ability of high school students. The description for each reasoning construct used is summarized in Table 1.

### **Proof schemes**

Mathematics education literature suggests various theoretical frameworks relating to the proof studied in school mathematics (Balacheff, 1991; Stylianides & Stylianides, 2008; Harel &

Showder., 2007). Mejia-Ramos and Inglis (2009) found that in a sample of 131 articles related to the theoretical framework of proof and argumentation in mathematics, only three articles focused on students' comprehension of given proofs. These findings suggest that more sophisticated ways of assessing students' comprehension of proof are needed. In regard to teaching and learning, teachers' understanding of constructs and proof schemes, especially in school mathematics, is essential in analyzing students' knowledge to prove a mathematical statement (Dickerson & Doerr, 2014; Blanton & Stylianou, 2014).

**Table 1.** Reasoning construct and its indicator

Reasoning construct	Type	Indicator
CMR (Creative Mathematical Reasoning)		<ul style="list-style-type: none"> <li>• Novelty: Students recreate new reasoning sequences that are created or that are forgotten.</li> <li>• Plausible: There are arguments in favour of strategy selection and strategy implementation that motivate why the conclusions are valid or reasonable.</li> <li>• Mathematical foundation: The argument is part of the intrinsic mathematical nature of the components involved in reasoning.</li> </ul>
IR (Imitative Reasoning)	MR (Memorized Reasoning)	<ul style="list-style-type: none"> <li>• Strategy selection is based on obtaining complete answers.</li> <li>• Implementation of the strategy will be done only in the written form contained in the question.</li> </ul>
	AR (Algorithmic Reasoning)	<ul style="list-style-type: none"> <li>• The choice of a strategy used is to remember the solution algorithm. Prediction arguments can be of various types but are not required to create a new solution.</li> <li>• Solving strategies are partly used because they are simple for pupils to understand. After all, the only thing that can go wrong in the calculation is negligence.</li> </ul>

NCTM (2000) outlines four proof constructions that need be developed in high schools, including direct proof, indirect proof, proof by example, and proof by mathematical induction. Stylianides and Stylianides (2008) suggest that two necessary logical inference rules can be considered as constructs of deductive proof, namely modus ponens (MP) and modus tollens (MT). Ponon's mode is the basis of direct proof, while Tollen's mode is the basis of indirect proof (including proof by contradiction and proof of contraposition).

Experts have various stances about the construct of proof in school mathematics. For example, Gutierrez, Pegg, and Lawrie (2004) only define two proof constructs in the context of school mathematics: empirical and deductive. In contrast, Shpiro (2014) created a proof scheme based on students' responses to familiar and unfamiliar mathematical content in his research. The proof scheme consists of deductive proof and particular proof. The first is divided into general proof, where the statement is proven by justifying each step. The latter shows some statement or justification can explain the truth of a statement.

Based on the preceding descriptions about constructs in proof in school mathematics, the scheme of proof developed by students' mindsets is essential to be analyzed so that teachers can develop problems that follow the schemes of students. As a result, it can give rise to another

mathematical mindset. In addition, the evidentiary constructs used in this study are formal deductive and empirical proof. The descriptions of the proof construct used in this study are presented in [Table 2](#).

**Table 2.** The construct of mathematical proof and its technique

Proof construct	Proof's technique
Formal Deductive	Transformative proof: mental operations that result in the transformation of the initial problem into another form of a problem. Structural proof: a proof that is a logical deduction from data, axioms, definitions, and theorems.
Empirical Proof	Perceptual proof: using images and perception of visual objects. Intellectual proof: proof based on empirical observation of the example, but the proof uses more abstract properties from the example.

### The use of technology for students' assessment

Technology is becoming increasingly important in our personal and professional life. It is also very significant in schools, particularly in mathematics instruction and assessment. To reach their greatest potential in education, technologies need be built with the characteristics of the intended students in mind. Researchers have also decided that technological literacy is a vital teaching skill since it allows pupils to comprehend, construct, and explore new problem-solving strategies (Bray & Tangney, 2017; Mainali & Key, 2012). Thus, the advantage of seamless technology integration with learning is that it broadens students' mental processing power to a new domain of knowledge representation via modeling, simulation, and visualization (Ziatdinov & Valles, 2022). Another crucial factor to consider is students' assessment. Technology has also changed the way students are assessed. Using computers as evaluation tools dates back to the 1960s (Green, 1964). There was an observation about a shift from behaviourism to constructivism throughout time (Karadag & McDougall, 2011). Hence, there was a movement from teacher-centered learning toward learner-centered and personalized learning. As a result, the need for adaptive evaluation is becoming more evident, especially in the online world.

GeoGebra, a web-based instructional tool, has been shown to play an important role in mathematics teaching and learning (Hohenwarter et al, 2008; Mthethwa et al, 2020). Furthermore, GeoGebra's capacity to integrate algebraic and geometric principles makes it ideal for discovering mathematical facts and relations and developing students' thinking and proof skills (Albano & Dello, 2019; Lepmann & Albre, 2008). Millions of users are already accustomed to the solid interface of GeoGebra, which has a wide range of features. In addition, GeoGebra can be used as an assessment tool by state-the-art theorem provers and also include characteristics that are challenging to alter, for example, the precise way to introduce the conjectures to be proved or the sort of output results (Botana et al., 2015). In this study, we conduct an assessment to examine students' reasoning and proof skills using Geogebra. Since the activities were carried out in Geogebra, students might discover or write their justification using the application.

## Methods

This study was a part of a dissertation project referring to design and development (D&D) research comprising three main stages: requirements assessment, design, and development and implementation. This article reports the development of the instrument in the third stage, which mainly focuses on construct validity. This study used a quantitative approach and the Confirmatory Factor Analysis (CFA). Data gathered using a quantitative technique, in the form of numbers, is statistically processed, and the results are explained. CFA aims to demonstrate that the model created is consistent with the theory developed by experts and field data. The expected final product is a reasoning and proof assessment instrument integrated with GeoGebra with high instrument quality.

## Participants

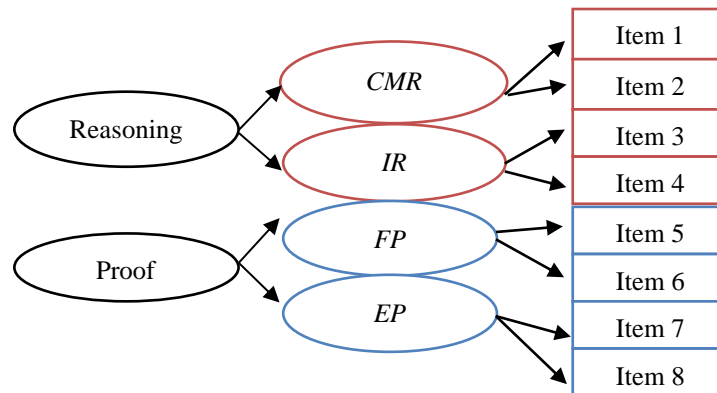
This research was conducted in East Java, involving 330 high school students. The schools involved in the development of the assessment instrument adjusted to the number of test participants used in the development of this assessment instrument. There is a rule of thumb for Structural Equation Modeling (SEM), which suggests the number of research samples is  $N > 200$  (Lacobucci, 2010). Another standpoint is that the number of research samples is 5-20 times the number of parameters used (Kline, 2005). Therefore, the number of students involved in this trial was 300 students. The number of observed variables or items determined the sample size in the CFA analysis. For the sample size, it is recommended to use the estimated Maximum Likelihood (ML) at 100-200 (Hair et al., 2006).

## The instrument

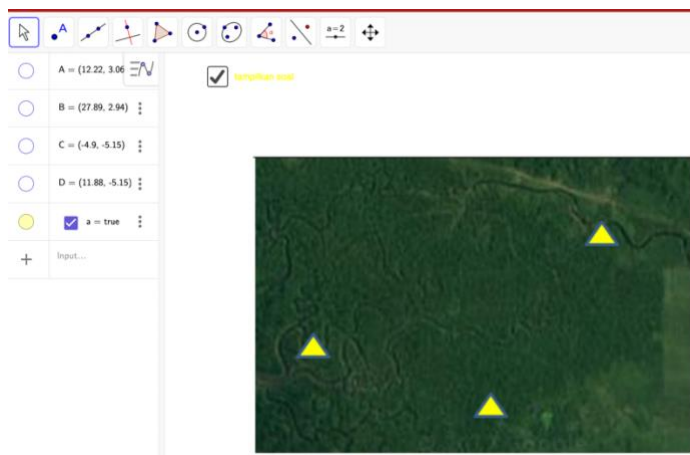
The data for this study was gathered using a questionnaire and a test instrument. The questionnaire was utilized to collect data on the experts' judgments on the developed instrument called Mathematical Reasoning and Proof Testing Tool for Project Improvement. Based on expert comments or ratings quantitatively, the statistical approach is used to measure the extent of instrument validity. In this research, we calculated the index following the Aiken model (1980, 1985) widely used in validating instrument items. According to Aiken Table (Aiken, 1985), the content validity index (V) required of the item is significant if above the cut-off value of 0.70 ( $V > 0.7$ ). The test instrument has fulfilled good criteria of content validity with Aiken indexes of 0.92. It was then used to collect data on students' mathematical reasoning and proof as empirical evidence to determine the instrument quality used in Indonesian schools. The instrument was made up of 8 items presented in Figure 1.

Figure 2 is one example of reasoning and proof assessment, specifically on creative mathematical reasoning. In this case, students are asked to determine the finish location of the race tournament from three different city locations. They could freely use GeoGebra to find the best location by using the concept of triangle or trigonometry. This integrated assessment test with GeoGebra could explore student creativity to find their best strategy based on their creative reasoning using GeoGebra tools, such as adding a point, drawing a line, dragging their sketch, and so on.





**Figure 1.** The model of second-order confirmatory factor analysis



*The figure illustrates the downtown locations of 3 neighbouring cities. The cities plan to hold races across the meadows around the city. Participants from each city must start from their respective city centres (The yellow sign). The finish point is somewhere around the meadow. Can you help determine the exact finish point for the race. Explain your reasons?*

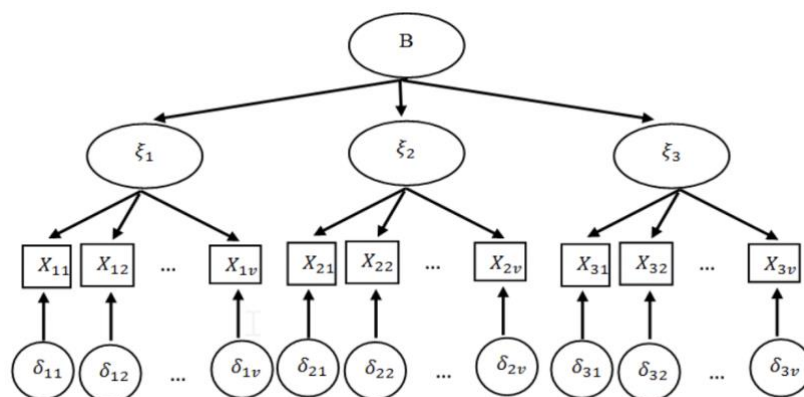
**Figure 2.** The excerpt of reasoning and proof test instrument integrated with GeoGebra

## Data analysis

Data analysis to examine the construct validity of the Mathematical Reasoning and Proof Test has two dimensions: Reasoning (R) and Proof (P), each with its own indicator, representing R and P. CFA was used to analyze the data. Construct validity analysis was conducted to determine whether the measurement results revealed the ability to be measured using factor loading data obtained from CFA (Brown & Moore, 2012). CFA was used to examine the instrument's validity based on the empirical data obtained (Brown & Moore, 2012). The criteria used to decide whether the model fits the data (valid) is based on the significance value (p-value) of Chi-Square (2) > 0.05 and Root Mean Square Error of Approximation (RMSEA) < 0.5 (Schumacker & Lomax, 2004).

The use of CFA is due to its ability to determine to construct discriminant and convergent validity, which were adjusted to the theory of measurement error. It is more comprehensive than the correlation analysis framework or multiple regression, which has the assumption that the variables involved are free from measurement errors (Harrington, 2009). In addition, CFA assumes that the variables that participate in the analysis have measurement errors (Brown, 2015).

Two forms of CFA analysis can be used to determine validity, namely first-order and second-order CFA (Marsh et al., 2014) assisted by LISREL 8 software. In this research, second-order confirmatory factor analysis was applied. The second-order CFA is a two-level measurement model. In the second CFA, the indicator variables cannot directly measure the latent variables. Therefore, some indicators cannot be measured directly, and more indicators are needed.



**Figure 3.** The model of second-order CFA

The first level of analysis is administered from the construct of latent aspects to the indicators, and the second level is carried out from the latent construct to the aspect construct (Jöreskog & Sörbom, 2006). The construct validity shows that the tested instruments suit the theoretical concept (Tavakol & Wetzel, 2020). It provides an overview of how perfect results can be achieved using theory-based measurement (Clark & Watson, 2019). In this research, the construct validity test of the second-order CFA was conducted by observing the factor loading value of ( $>0.5$ ) and t-value of ( $>1.96$ ). Hair et al. (2006) set 0.5 as the minimum factor loading value. Meanwhile, the construct reliability is considered good if the Construct Reliability (CR) equals or is greater than 0.70 and the variance extracted value equals or is greater than 0.50. The reliability is fulfilled if the construct reliability value shows  $> 0.70$  (Hair et al., 2010).

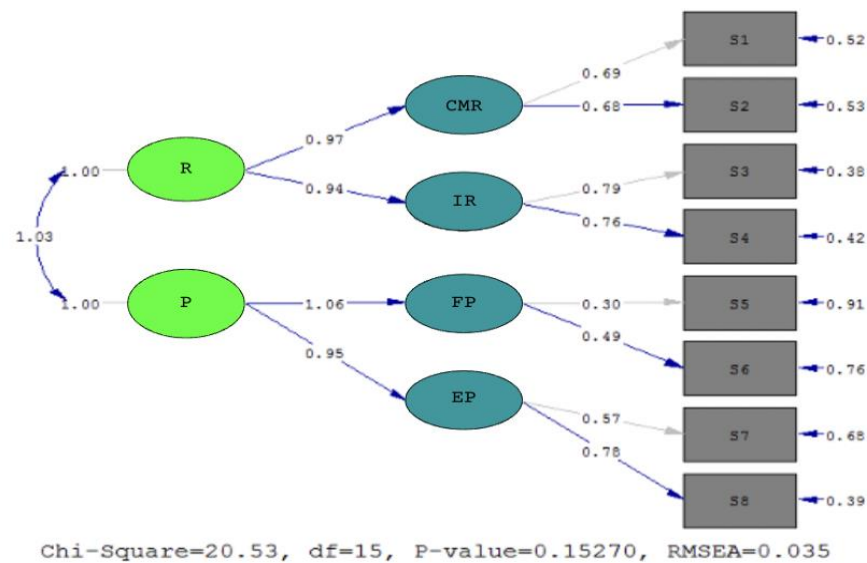
## Findings and Discussion

Results of the CFA analysis in this study show that all items of each latent variable, such as CMR (Creative Mathematical Reasoning), IR (Imitative Reasoning), FP (Formal Proof), and EP (Empirical Proof), measure the components of mathematical reasoning and proof. From each item constructed, it was intended to prove that each item measures a component of mathematical reasoning and proof. Dimensions follow a factor model and each dimension's item also has a large impact.

The data analysis resulted in the RMSEA (Root Mean Square Error of Approximation) value is 0.035, the chi-square is 20.53 with a p-value of 0.153, GFI is 0.98, and AGFI is 0.96. These indicate that the model fit has been well fulfilled by using the second-order CFA. Previous studies (Savalei, 2017; Xia & Yang, 2019) examined the impact of estimation methods on the SEM fit index. In particular, recent work by Xia and Yang (2019) has systematically tested root mean square error approximation (RMSEA; Steiger, 1990) and comparative fit index (CFI; Bentler, 1990) estimation methods. Using simulations and empirical examples, these authors



yield smaller RMSEAs and larger CFIs than those obtained with ML, suggesting that the model fits better. The complete result of the model fit analysis is shown in Figure 4.



**Figure 4.** The fit model of second-order confirmatory factor analysis

The results of the second-order CFA for the t-value and SLF (Standardized Loading Factor) and the calculation of the Alpha reliability coefficient in the first trial are presented in Table 3.

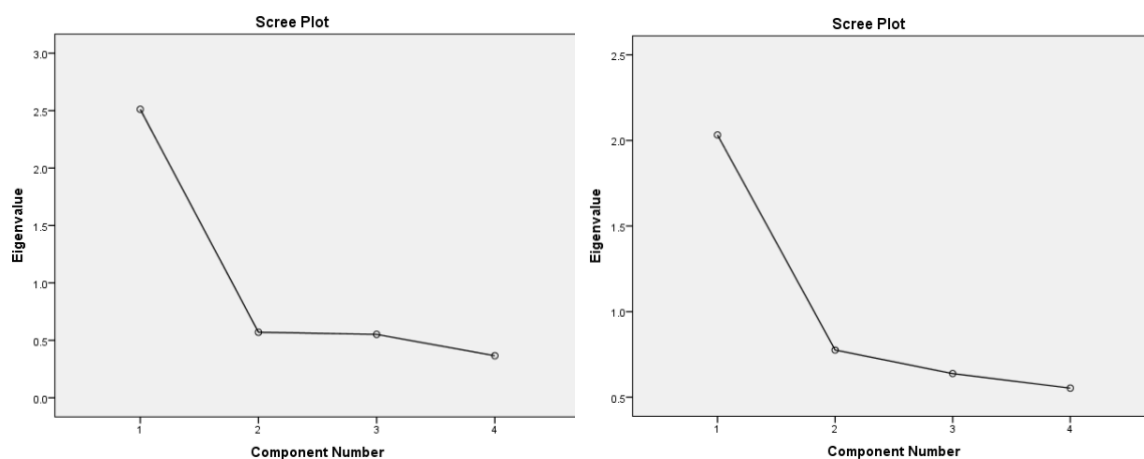
**Table 3.** The result of second-order CFA for t-value and SLF

Item	Aspect	Indicator	Second-order CFA				Validity	Construct reliability
			t-value	description	SLF	Error		
1	R	CMR	9,167	Significance	0,69	0,079	Valid	0,892
2			10,182	Significance	0,68	0,06	Valid	
3		IR	8,038	Significance	0,79	0,04	Valid	
4			12,722	Significance	0,76	0,08	Valid	
5	P	Formal Proof	11,434	Significance	0,3	0,128	Valid	0,749
6			4,659	Significance	0,49	0,125	Valid	
7		Empirical Proof	10,981	Significance	0,57	0,085	Valid	
8			9,093	Significance	0,78	0,086	Valid	

Based on Table 3, in terms of the t-value test, all test items are significant in supporting the reasoning and mathematical proof test constructs, with the highest support by item 5 and the lowest by item 6. A construct has good reliability if the value of Construct Reliability (CR) is 0.70 (Rosli et al., 2021; Hair et al., 2010). It means that the constructs contained in the test items could measure the ability of reasoning and mathematical proof. Good test reliability also shows that the measurement results obtained from this test are consistent.

Construct validity can be defined as an effort to measure how far the items can measure what they want to measure following the previously defined concept. The criterion is valid in the CFA analysis or can be said to be valid if the  $t$ -value  $> 1.96$  has a loading factor  $> 0.5$  for a sample size of more than 300 (Hair et al., 2010). Therefore, all items on the current study fulfil construct validity.

In addition to fulfilling good construct validity, it also indicates fulfillment of unidimensionality in the reasoning and proof test. One of the most commonly used psychometric models to measure a single construct is the Unidimensional Item Response Theory (UIRT). It recognizes the potential presence of multiple dimensions in a test (Strachan et al., 2022). It also means that the developed reasoning and proof test integrated with GeoGebra consists of 2 subsets of questions. Each of these questions measures one ability: reasoning or proof ability. In addition, the fulfillment of unidimensionality can be shown visually on the Eigenvalue graph with factor analysis using SPSS version 20, as shown in Figure\_5. It shows the eigenvalue of one factor, which is dominant compared to the Eigenvalue of the other factors. It means that the unidimensional assumption on the test is said to have been fulfilled.



**Figure 5.** The graph of Eigenvalue on mathematical reasoning and proof test

The developed test instrument's constructs include two frameworks of reasoning competence and two frameworks of proof; the first includes creative mathematical reasoning (CMR) and imitative reasoning (IR). Meanwhile, the latter is comprised of formal proof (FP) and empirical proof (EP). As a result, we discovered that the developed test instrument met the construct reliability criteria (Table 3). Construct reliability measures how well variables underlying constructs served in structural equation modeling (Zinbarg et al., 2005). Construct reliability could be determined after construct validity has been established using confirmatory factor analysis. Construct reliability is determined using the factor loading analysis (Geldhof et al., 2014). According to Gefen et al (2000), a construct reliability coefficient greater than 0.70 is appropriate. A high coefficient denotes a high level of internal consistency. It would only be conceivable if every variable measured the same latent construct consistently.

## Conclusion

The second-order CFA on the reasoning and proof instrument integrated with GeoGebra shows that the reasoning and proof scale is valid and reliable. Therefore, the instrument could

be used to measure reasoning and proof ability among high school students. The evidence of construct validity of the developed instrument was based on four latent variables, namely CMR (Creative Mathematical Reasoning), IR (Imitative Reasoning), FP (Formal Proof), and EP (Empirical Proof). The instrument comprises eight items from the CMR, IR, FP and EP. The loading factor has a significant effect as unidimensional on the latent variable derived from the CFA analysis. It can be seen from the  $t$ -value  $> 1.96$  and the loading factor value  $> 0.5$ . However, the limited number of items and the tight control of time allow the scores obtained by the respondents to be influential. For further research, this instrument might be utilized to describe the profile of high school students in solving mathematical reasoning and proof tests.

## References

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131-142. Doi: [10.1177/0013164485451012](https://doi.org/10.1177/0013164485451012)
- Akkurt, Y. Y., & Durmus, S. (2022). Tracing proof schemes: Some patterns and new perspectives. *Journal of Research and Advances in Mathematics Education*, 7(1), 1-16. Doi: [10.23917/jramathedu.v7i1.15740](https://doi.org/10.23917/jramathedu.v7i1.15740)
- Albano, G., & Dello Iacono, U. (2019). GeoGebra in e-learning environments: A possible integration in mathematics and beyond. *Journal of Ambient Intelligence and Humanized Computing*, 10(11), 4331-4343. Doi: [10.1007/S12652-018-1111-X](https://doi.org/10.1007/S12652-018-1111-X)
- An, X., & Yung, Y. F. (2014). Item response theory: What it is and how you can use the IRT procedure to apply it. *SAS Institute Inc. SAS364-2014*, 10(4), 1-14.
- Balacheff, N. (1991). The benefits and limits of social interaction: The case of mathematical proof. In A.J. Bishop, S. Mellin-Olsen, & J. van Dormolen. (Eds.), *Mathematical knowledge: Its growth through teaching* (pp. 173-192). Springer. Doi: [10.1007/978-94-017-2195-0\\_9](https://doi.org/10.1007/978-94-017-2195-0_9)
- Baylis, J. (1983). Proof — the essence of mathematics. *International Journal of Mathematical Education in Science and Technology*, 14(4), 409–414. Doi: [10.1080/0020739830140403](https://doi.org/10.1080/0020739830140403)
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. Doi: [10.1037/0033-2909.107.2.238](https://doi.org/10.1037/0033-2909.107.2.238)
- Blanton, M. L., & Stylianou, D. A. (2014). Understanding the role of transactive reasoning in classroom discourse as students learn to construct proofs. *The Journal of Mathematical Behavior*, 34, 76-98. Doi: [10.1016/J.JMATHB.2014.02.001](https://doi.org/10.1016/J.JMATHB.2014.02.001)
- Botana, F., Hohenwarter, M., Janičić, P., Kovács, Z., Petrović, I., Recio, T., & Weitzhofer, S. (2015). Automated theorem proving in GeoGebra: Current achievements. *Journal of Automated Reasoning*, 55(1), 39-59. Doi: [10.1007/s10817-015-9326-4](https://doi.org/10.1007/s10817-015-9326-4)
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). The Guilford Press.
- Buchbinder, O., & McCrone, S. (2022). Guiding principles for teaching mathematics via reasoning and proving. In J. Hodgen, E. Geraniou, G. Bolondi, & F. Ferretti. (Eds.). *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)*. Free University of Bozen-Bolzano and ERME. <https://hal.archives-ouvertes.fr/hal-03746878v2>
- Bray, A., & Tangney, B. (2017). Technology usage in mathematics education research – A systematic review of recent trends. *Computers and Education*, 114, 255–273. Doi: [10.1016/j.compedu.2017.07.004](https://doi.org/10.1016/j.compedu.2017.07.004)
- Chotimah, S., Wijaya, T. T., Aprianti, E., Akbar, P., & Bernard, M. (2020). Increasing primary school students' reasoning ability on the topic of plane geometry by using hawgent dynamic mathematics software. *Journal of Physics: Conference*, 1-8. Doi: [10.1088/1742-6596/1657/1/012009](https://doi.org/10.1088/1742-6596/1657/1/012009)
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412-1427. Doi: [10.1037/pas0000626](https://doi.org/10.1037/pas0000626)

- Dickerson, D. S., & Doerr, H. M. (2014). High school mathematics teachers' perspectives on the purposes of mathematical proof in school mathematics. *Mathematics Education Research Journal*, 26(4), 711-733. Doi: [10.1007/s13394-013-0091-6](https://doi.org/10.1007/s13394-013-0091-6)
- Doruk, M. (2019). Preservice mathematics teachers' determination skills of the proof techniques: The case of integers. *International Journal of Education in Mathematics, Science and Technology*, 7(4), 335-348.
- Fiangga, S. (2014). Tangram game activities, helping the student's difficulty in understanding the concept of area conservation. *Proceedings of International Conference on Research, Implementation, and Education of Mathematics and Sciences* (pp. 453-460). UNY. Doi: [10.13140/RG.2.1.3479.4965](https://doi.org/10.13140/RG.2.1.3479.4965)
- Fischer, G., Lemke, A. C., McCall, R., & Morch, A. I. (2020). Making argumentation serve design. In T. P. Moran & J. M. Carroll (ed.), *Design rationale* (pp. 267-293). CRC Press. Doi: [10.1201/9781003064053-12](https://doi.org/10.1201/9781003064053-12)
- Fu, Y., Qi, C., & Wang, J. (2022). Reasoning and proof in algebra: the case of three reform-oriented textbooks in China. *Canadian Journal of Science, Mathematics and Technology Education*, 22(1), 130-149. Doi: [10.1007/s42330-022-00199-1](https://doi.org/10.1007/s42330-022-00199-1)
- Gefen, D., Straub, D. & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*, 4(7), 1-77. Doi: [10.17705/1CAIS.00407](https://doi.org/10.17705/1CAIS.00407)
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72-91. Doi: [10.1037/a0032138](https://doi.org/10.1037/a0032138)
- Ginting, M. S., Prahmana, R. C. I., Isa, M., & Murni, M. (2018). Improving the reasoning ability of elementary school student through the indonesian realistic mathematics education. *Journal on Mathematics Education*, 9(1), 41-54. Doi: [10.22342/jme.9.1.5049.41-54](https://doi.org/10.22342/jme.9.1.5049.41-54)
- Gutierrez, A., Pegg, J., & Lawrie, C. (2004). Characterization of Students' Reasoning and Proof Abilities in 3-Dimensional Geometry. In A. B. Fuglestad & M. J. Høines (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (pp.511-518). PME.
- Green, B. F. (1964). Intelligence and computer simulation. *Transactions of the New York Academy of Sciences*, 27(1), 55-63. Doi: [10.1111/j.2164-0947.1964.tb03486.x](https://doi.org/10.1111/j.2164-0947.1964.tb03486.x)
- Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester Jr (Ed.). *Second Handbook of research on mathematics teaching and learning* (pp. 805-842). Information Age Publishing.
- Harrington, D. (2009). *Confirmatory factor analysis*. Oxford University Press.
- Haylock, D., Thangata, F. (2007). *Key concepts in teaching primary mathematics*. Sage Publication. Doi: [10.4135/9781446214503](https://doi.org/10.4135/9781446214503)
- Hair, J., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., (2006). *Multivariate data analysis* (6th ed.). Prentice-Hall
- Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010). *Multivariate Data Analysis* (7th ed.). Prentice-Hall.
- Hemmi, K., Lepik, M., & Viholainen, A. (2013). Analysing proof-related competences in Estonian, Finnish and Swedish mathematics curricula—towards a framework of developmental proof. *Journal of Curriculum Studies*, 45(3), 354-378. Doi: [10.1080/00220272.2012.754055](https://doi.org/10.1080/00220272.2012.754055)
- Hidayat, W., & Prabawanto, S. (2018). Improving students' creative mathematical reasoning ability students through adversity quotient and argument driven inquiry learning. *Journal of Physics: Conference Series*, 948(1). Doi: [10.1088/1742-6596/948/1/012005](https://doi.org/10.1088/1742-6596/948/1/012005)
- Hohenwarter, M., Hohenwarter, J., Kreis, Y., & Lavicza, Z. (2008). Teaching and learning calculus with free dynamic mathematics software GeoGebra. In G Kaiser (ed.), *Proceedings of the 11th International Congress on Mathematical Education* (pp. 1-9). ICMI <https://orbi.lu.uni/bitstream/10993/47219/1/ICME11-TSG16.pdf>
- Jöreskog, K. & Sörbom, D. (2006). *LISREL 8.80 for Windows*. Computer Software. Scientific Software International, Inc.
- Karadag, Z., & McDougall, D. (2011). Geogebra as a Cognitive Tool. In L. Bu & R. Schoen (Eds.), *Model-Centered Learning: Modeling and Simulations for Learning and Instruction* (vol 6). Sense Publishers. Doi: [10.1007/978-94-6091-618-2\\_12](https://doi.org/10.1007/978-94-6091-618-2_12)
- Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. SAGE. Doi: [10.4135/9781483385693](https://doi.org/10.4135/9781483385693)

- Lacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, 20(1), 90-98. Doi: [10.1016/j.jcps.2009.09.003](https://doi.org/10.1016/j.jcps.2009.09.003)
- Lepmann, T., & Albre, J. (2008). Some possibilities of teaching geometry with GeoGebra. *Koolimatematika*, 35, 52-57.
- Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, 67(3), 255-276. Doi: [10.1007/s10649-007-9104-2](https://doi.org/10.1007/s10649-007-9104-2)
- Mainali, B. R., & Key, M. B. (2012). Using dynamic geometry software GeoGebra in developing countries: A case study of impressions of mathematics teachers in Nepal. *International Journal for Mathematics Teaching and Learning*, 1-16. <http://www.cimt.plymouth.ac.uk/Journal/mainali.pdf>
- Maoto, S., Masha, K., & Mokwana, L. (2018). Teachers' learning and assessing of mathematical processes with emphasis on representations, reasoning and proof. *Pythagoras*, 39(1), 1-10. Doi: [10.4102/pythagoras.v39i1.373](https://doi.org/10.4102/pythagoras.v39i1.373)
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10(1), 85-110. Doi: [10.1146/annurev-clinpsy-032813-153700](https://doi.org/10.1146/annurev-clinpsy-032813-153700)
- Mejía-Ramos, J. P., & Inglis, M. (2009). Argumentative and proving activities in mathematics education research. In F-L. Lin, F-J. Hsieh, G. Hanna, & M. de Villiers (Eds.), *Proceedings of the ICMI study 19 conference: Proof and proving in mathematics education* (Vol. 2, pp. 88-93). National Taiwan Normal University.
- Mthethwa, M., Bayaga, A., Bossé, M. J., & Williams, D. (2020). GeoGebra for learning and teaching: A parallel investigation. *South African Journal of Education*, 40(2), 1-12. Doi: [10.15700/saje.v40n2a1669](https://doi.org/10.15700/saje.v40n2a1669)
- Mumu, J., & Tanujaya, B. (2019). Measure reasoning skill of mathematics students. *International Journal of Higher Education*, 8(6), 85-91. Doi: [10.5430/ijhe.v8n6p85](https://doi.org/10.5430/ijhe.v8n6p85)
- NCTM. (2000). *Principles and standards for school mathematics*. NCTM.
- NCTM. (2009). *Focus in high school mathematics: Reasoning and sense making*. NCTM.
- Noto, M. S., Priatna, N., & Dahlan, J. A. (2019). Mathematical proof: The learning obstacles of preservice mathematics teachers on transformation geometry. *Journal on Mathematics Education*, 10(1), 117-126. Doi: [10.22342/jme.10.1.5379.117-126](https://doi.org/10.22342/jme.10.1.5379.117-126)
- OECD. (2019). *PISA 2018 results: Combined executive summaries* (Vols. I, II & III). OECD. [https://www.oecd.org/pisa/Combined\\_Executive\\_Summaries\\_PISA\\_2018.pdf](https://www.oecd.org/pisa/Combined_Executive_Summaries_PISA_2018.pdf).
- Reiss, K. M., Heinze, A., Renkel, A., & Grob, C. (2008). Reasoning and proof in geometry: Effects of a learning environment based on heuristic worked-out examples. *ZDM Mathematics Education*, 40(3), 455-467. Doi: [10.1007/s11858-008-0105-0](https://doi.org/10.1007/s11858-008-0105-0)
- Rosli, M. S., Saleh, N. S., Alshammari, S. H., Ibrahim, M. M., Atan, A. S., & Atan, N. A. (2021). Improving questionnaire reliability using construct reliability for researches in educational technology. *Int. J. Interact. Mob. Technol.*, 15(4), 109-116. Doi: [10.3991/ijim.v15i04.20199](https://doi.org/10.3991/ijim.v15i04.20199)
- Saeedullah, & Akbar, R. A. (2021). Developing a test to measure mathematical reasoning among high school students. *Journal of Science Education*, 3(1), 1-13. <http://journal.aiou.edu.pk/journal1/index.php/jse/article/view/1357>
- Sari, D. P., & Mahendra. (2017). Developing instrument to measure mathematical reasoning ability. *The Proceedings of International Conference on Mathematics and Science Education* (pp. 30-33). Atlantis Press. Doi: [10.2991/icmsed-16.2017.7](https://doi.org/10.2991/icmsed-16.2017.7)
- Sari, Y. M., Kartowagiran, B., & Retnawati, H. (2020). Mathematics teachers' challenges in implementing reasoning and proof assessment a case of Indonesian teachers. *Universal Journal of Educational Research*, 8(7), 3287-3293. Doi: [10.13189/ujer.2020.080759](https://doi.org/10.13189/ujer.2020.080759)
- Savalei, V. (2017). *Reconstructing fit indices in SEM with categorical data: Borrowing insights from nonnormal data*. Paper presented at the Annual Meeting of the Society of Multivariate Experimental Psychology. Minneapolis, MN.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. psychology press.
- Seah, R., & Horne, M. (2020). The construction and validation of a geometric reasoning test item to support the development of learning progression. *Mathematics Education Research Journal*, 32(4), 607-628. Doi: [10.1007/s13394-019-00273-2](https://doi.org/10.1007/s13394-019-00273-2)



- Sevimli, E. (2018). Undergraduates' propositional knowledge and proof schemes regarding differentiability and integrability concepts. *International Journal of Mathematical Education in Science and Technology*, 49(7), 1052-1068. Doi: [10.1080/0020739X.2018.1430384](https://doi.org/10.1080/0020739X.2018.1430384)
- Shpiro, A. (2014). Unfamiliar properties of familiar shapes. *The College Mathematics Journal*, 45(5), 371-375. Doi: [10.4169/college.math.j.45.5.371](https://doi.org/10.4169/college.math.j.45.5.371)
- Steiger J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180. Doi: [10.1207/s15327906mbr2502\\_4](https://doi.org/10.1207/s15327906mbr2502_4)
- Strachan, T., Cho, U. H., Ackerman, T., Chen, S. H., de la Torre, J., & Ip, E. H. (2022). Evaluation of the linear composite conjecture for unidimensional IRT scale for multidimensional responses. *Applied Psychological Measurement*, 46(5), 347-360. Doi: [10.1177/01466216221084218](https://doi.org/10.1177/01466216221084218)
- Stylianides, G. J. (2009). Reasoning-and-proving in school mathematics textbooks. *Mathematical Thinking and Learning*, 11(4), 258-288. Doi: [10.1080/10986060903253954](https://doi.org/10.1080/10986060903253954)
- Stylianides, G. J., & Stylianides, A. J. (2008). Proof in school mathematics: Insights from psychological research into students' ability for deductive reasoning. *Mathematical Thinking and Learning*, 10(2), 103-133. DOI: [10.1080/10986060701854425](https://doi.org/10.1080/10986060701854425)
- Stylianides, G. J., & Stylianides, A. J. (2017). Research-based interventions in the area of proof: The past, the present, and the future. *Educational Studies in Mathematics*, 96(2), 119-127. Doi: [10.1007/s10649-017-9782-3](https://doi.org/10.1007/s10649-017-9782-3)
- Stylianides, A. J. (2019). Secondary students' proof constructions in mathematics: The role of written versus oral mode of argument representation. *Review of Education*, 7(1), 156-182. Doi: [10.1002/rev3.3157](https://doi.org/10.1002/rev3.3157)
- Sumpter, L. (2013). Themes and interplay of beliefs in mathematical reasoning. *International Journal of Science and Mathematics Education*, 11(5), 1115-1135. Doi: [10.1007/s10763-012-9392-6](https://doi.org/10.1007/s10763-012-9392-6)
- Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, 11, 245. Doi: [10.5116/ijme.5f96.0f4a](https://doi.org/10.5116/ijme.5f96.0f4a)
- Thompson, D. R., Senk, S. L., & Johnson, G. J. (2012). Opportunities to learn reasoning and proof in high school mathematics textbooks. *Journal for Research in Mathematics Education*, 43(3), 253-295. Doi: [10.5951/jresmetheduc.43.3.0253](https://doi.org/10.5951/jresmetheduc.43.3.0253)
- Xia, Y., Yang Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51, 409-428. Doi: [10.3758/s13428-018-1055-2](https://doi.org/10.3758/s13428-018-1055-2)
- Ziatdinov, R., & Valles Jr, J. R. (2022). Synthesis of modeling, visualization, and programming in GeoGebra as an effective approach for teaching and learning STEM topics. *Mathematics*, 10(3), 398. Doi: [10.3390/math10030398](https://doi.org/10.3390/math10030398)
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega$ H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123-33. Doi: [10.1007/s11336-003-0974-7](https://doi.org/10.1007/s11336-003-0974-7)