Assessment of Students' Mathematical Proof Comprehension: Gender and Year Level Background

To cite this article: Mohamad Waluyo and Tibor Vidákovich 2021 *J. Phys.: Conf. Ser.* **1776** 012017

View the article online for updates and enhancements.

Assessment of Students' Mathematical Proof Comprehension: Gender and Year Level Background

Mohamad Waluyo[1,2], Tibor Vidákovich[1]

[1] Doctoral School of Education, University of Szeged, Hungary
[2] Mathematics Education, Universitas Muhammadiyah Surakarta, Indonesia

Corresponding author: Mohammad.Waluyo@ums.ac.id

Abstract. The ability to understanding proof is essential before students can write their proof. A literature study shows a few interests in investigating the students' knowledge from reading proof. This study aims to assess students' comprehension when they were reading proof using a developed instrument based on the previous study on the same field. Three groups of students from a different level of years from a preservice teacher training program in an Indonesian university are taken as the participants. This paper reports the preliminary finding of students' comprehension score comparison by gender and year level. Further analysis also involved the correlation of students' basic knowledge to their comprehension scores.

Keywords: *Mathematics Proof, Proof Comprehension, Gender, Undergraduate Mathematics*

1. Introduction

The ability to write proofs is pointed out by [1] as a primary goal and being emphasized in advance mathematics courses. However, university students are often facing difficulties in writing a well and valid proof [1–3] in several mathematical concepts such as abstraction such as real analysis [4], abstract algebra, and number theory [5, 6]. These difficulties are mainly connected to how and where students learn to write a proof. Two significant possible sources influencing students' ability to write are mathematics classrooms and textbooks.

It is widely approved that mathematics in university is largely taught in a 'definition theorem-proof' format, in which lectures mostly begin with introducing definitions of a concept, presenting theorems about the concept, and then providing proofs of the theorems. Textbooks for advanced mathematics courses deliver material in the same pattern. An important assumption based on this pedagogical practice is that the students are able to learn mathematics concepts from the proofs that they have studied. Still, as mathematicians or mathematics educators observe, students' comprehension of the proofs that they studied is rarely assessed in a purposeful and significant way [7, 8]. In this report, we describe the process of how we developed a test to measure students' understanding of proofs and elaborate on the background variable that may affect it, including grades and gender.

2. Theoretical Framework

2.1. Studies on Mathematics Proof Reading

Research on mathematical proof has been carried out by many researchers, practitioners, and even mathematics education teachers. From the literature review, it has been found that many studies on mathematical proof emphasize the ability of students to construct proof constructs rather than the students' ability to read, understand the contents, and the flow of the author's thinking in constructing mathematical proofs [9]. From literature studies that examine proof comprehension, research focuses on investigating whether students are able to distinguish valid evidence or not. In other words, these studies investigate how students evaluate evidence such as [10–13], be it its validity or other things related to it. From the research that has been done above, it can be concluded that students do not understand much and learn from the evidence they read. A number of studies on mathematics reading comprehension were also carried out on students and teachers as benchmarks in understanding mathematical proof, whether it is valid or not, such as [14–17].

2.2. Studies on Evaluation of Mathematics Proof Comprehension

So far, many students' proof comprehension evaluations still use open items. Some researchers, such as [18–20] use this open item model to measure students' understanding of reading a form of proof. Of course, research with open items can also provide an overview of students' understanding of mathematical evidence, but there are several shortcomings of this open item. The first is a matter of time. The open item model will require more time to evaluate, especially if it involves a large number of samples. The second problem is about the validity and reliability of the test, which cannot be well-measured. In dealing with this, several researchers have tried to develop a closed item model that can be used to measure students' proof comprehension extensively.

In general, there are two dominant contributions recorded in the literature study by Waluyo [21]. The first contribution is from the research of Yang & Lin [22], which initially developed a model to evaluate students' understanding of reading proof in Geometry (RCGP). The model they have developed consists of four stages: surface, recognize the element, chaining element, and encapsulation. The *surface* stage means that students can understand the meaning of a term and/or basic statement in proof. For example, they are able to understand the term "right-angle," which is a 90-degree angle in geometry. The stage *recognizes the element* means that students are able to understand an assumption, definition, or axiom. For example, students can understand the sentence "line *g* and line *l* are parallel to each other." Here students must understand the sentence and know its consequences in geometry. The *chaining element* stage means that students are able to understand a statement related to the previous information. The last step of *encapsulation* means that students can understand the whole idea of proof from beginning to end. From the four stages, then Yand and Lin developed items that could be used to measure students' level of understanding when reading proof in a specific area in mathematics.

A second contribution comes from the Mejia-Ramos study [23]. This contribution is an extension of the Yang and Lin model that was previously used to evaluate the proof comprehension abilities of senior high school students. Mejia-Ramos developed Yang and Lin's model for evaluating reading proof comprehension in advanced mathematics by breaking down the previous model into seven types of items, which can be grouped into two groups, namely *local* and *holistic*. *Local* understanding includes students' understanding of the components of a mathematical proof such as terms, axioms, and statements and the consequences of that statement. If this *local* understanding is reflected in the Yan and Ling model, this understanding is analogous to the *surface* stage and *recognize the element*. Meanwhile, the *holistic* understanding is an understanding of the idea and the general framework of evidence. The analogy of this understanding in the previous model is *chaining elements* and *encapsulation*.

3. Methods

We have developed comprehension tests for proofs of the following two theorems

Theorem 1: The summation of three even numbers cannot produce an odd number.
Theorem 2: n is odd if and only if $3n^2 + 8$ is odd

The proof that students try to understand in this test is formal proof with deductive logic using basic knowledge about odd and even numbers and then using algebraic properties to make a derivative statement. The proof scheme that students try to understand in theorem 1 is proof with contradiction, while in the proof of theorem 2 is proof of *if and only if* statement using the direct proof and contradictions scheme. The selection of the theorem considers that this theorem is relatively simple that allows it to be given to various levels of education. For secondary school education, researchers only use theorem-1 thinking that secondary school students do not commonly use theorem like theorem-2, which contains operators *if and only if*. Whereas for the undergraduate level, researchers provide the two theorems with the assumption that the theorems are quite simple and do not need the requirement of the theoretical basis of certain subjects (e.g., number theory or abstract algebra).

The method in this study consisted of three main phases aimed at developing multiple-choice tests that could be easily administered and graded. The first phase was to generate the questions and choices of items of the multiple-choice test. This phase aimed to develop each proof comprehension test that began by generating open-ended then generating multiple-choice questions versions of each type of assessment item in the model (can be seen in Table 1). In the next phase, we modified items from the test that was mathematically incorrect, ambiguous, or items that were not reflecting students' understanding of a particular proof. This phase was conducted by asking other math educators and mathematicians to review the multiple-choice tests generated. The final phase was administering the test for a small sample for validating the test, then administering it to a larger sample scale. The objective of this phase was to verify that these tests had good reliability and to identify items that can be removed to generate a final multiple-choice test.

Table 1. The aspects of proof comprehension used in the instrument

Aspect	Object of comprehension
Mean of terms	• Know the definition of terms • Give examples to a specific term in the proof
Logical Status and Proof framework	• Know the aim of a sentence in the proof • Identify the framework proof type
Justification and claims	• Make explicit an implicit warrant in the proof • Identify the specific data supporting a given claim • Identify the specific claims that are supported by a given statement
Summarizing high-level ideas	• select a good summary of the proof • Identify a good summary of a key sub-proof in the proof
Identify the modular structure	• Identify the logical relation between modules of a proof
Transferring to another context	• Appreciate the scope of the method
Illustrating example	• Illustrate a sequence of inferences with a specific example

3.1. Participants

The research is conducted in Indonesia. The participants for administering the test in a small sample were 53 students of undergraduate level who are taking a preservice teacher training program in mathematics. These students sit for the test twice in a short period of time to check the test-retest reliability of the instrument. After tested in a small group of samples, the instrument was tested for undergraduate level students from the same program of teacher training, accounted for 141 first-year, 118 second-year, and 100 third-year students. The test was administered during the first week of the fall semester of 2019.

4. Results

Instrumen consists of 25 multiple choice tests divided into two theorems. The score for the correct answer is one otherwise, it will get a score 0. Thus maximum score available is 25. In the instrument trial for a small sample of 53 students, we use a one-week test-retest reliability method as one of the most common forms of reliability using a correlation between the two sets of scores [24]. The correlation between the first and the second administration is 0.72 and is considered as acceptable reliability.

4.1. Differences across the year level

After obtaining an acceptable reliability value from the developed instrument, the next step is administering the test for a larger sample of the first, second, and third year of undergraduate preservice teacher training students. The description of the result from the test administration can be seen in table 2. The test consists of 2 main passage of theorem. The mean score for theorem 1 is higher than theorem 2 (see figure 1). The theorem-1 only use one direction scheme of proving whereas theorem-2 use *if and only if* proof scheme which contains 2 part of proofs. The overall score of proof comprehension of preservice teacher training students is slightly increasing over the year.

Table 2. The description of the score among proof comprehension in theorem 1, 2, and the overall

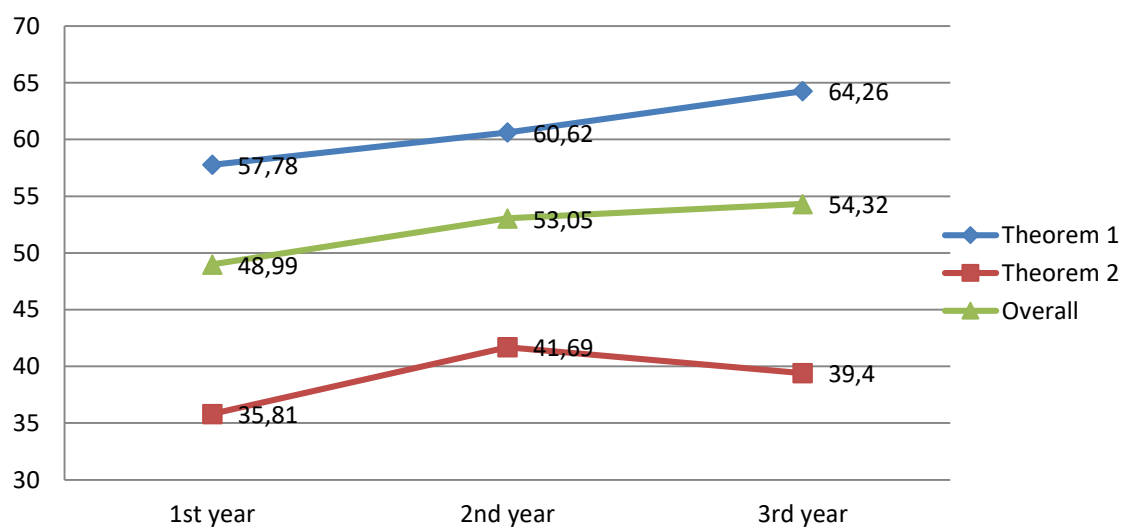| Measure	Year 1			Year 2			Year 3		
	M	SD	n	M	SD	n	M	SD	n
Theorem 1	57,78	13,08	141	60,62	15,12	118	64,26	13,14	100
Theorem 2	35,81	16,52	141	41,69	15,92	118	39,40	20,34	100
Overall	48,99	11,36	141	53,05	11,80	118	54,32	12,00	100

Figure 1. The development of proof comprehension of undergraduate students

Statistical tests of the overall result across the year of students show a significant difference among the three students' group levels. The analysis of variance showed a main effect year level on the overall score of proof comprehension, $F(2, 356) = 7.062$, $p = .001$. Posthoc analyses using Tukey's indicated that the score of proof comprehension for the first year is the lowest compared to the other groups, while there is no significant difference between the score of the second and third year in proof comprehension. If we only focused on theorem 1, the result can be extended to senior high school students. The result shows the consistency that the university student perform better to comprehend the theorem than senior high students.
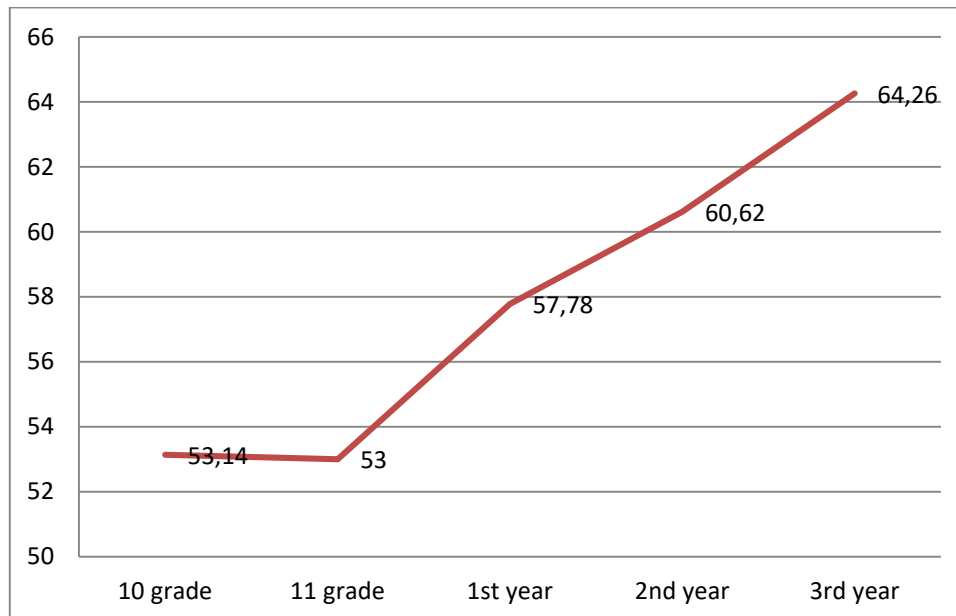

Figure 2. The development of proof comprehension from senior high school until university level

4.2. Gender Differences

The proportion of male-female at the teacher training program is not ideal as it has a ratio of 85:15 with a bigger number of female students. An independent-samples t-test indicated that proof comprehension scores were not significantly difference for female students ($M = 51.61$, $SD = 11.47$) and men ($M = 52.96$, $SD = 14.09$), $t(66) = -0.669$, $p > .005$. Levene's test indicated unequal variances ($F = 5.35$, $p = .021$), so degrees of freedom were adjusted from 357 to 66. This result indicates that proof comprehension of males and females is equal, but if we investigate further the development of the skill separately, we will discover another perspective. There are a interaction between variable gender and year toward the competence ($F(2, 353) = 7.042$, $p = .001$). In the first year, male students have a lower ability than the female ones, but in the third year, male students achieve higher than females.

4.3. Correlation between basic knowledge, logic and the proof matter

Those, as mentioned earlier on the outline of the instrument, basic knowledge, and logic, have some portion on the test. The objective was to investigate the contribution of these aspects to their understanding of particular proof. The correlation of the basic knowledge to the overall score was moderate (0.511), and its correlation to proof scheme understanding is quite small. Almost similar to basic knowledge, logical status also has a moderate contribution to the overall score and only has a small correlation to the proof understanding. It is still become an interesting area of research to remember that such a result is also indicated by Lin & Yang [25]. They use multiple regression to explain what influences the proof comprehension but for geometry subject. Their result is that the contribution of

these variables was varied for different grades around 10 to 40 percent explaining the variance of proof comprehension.
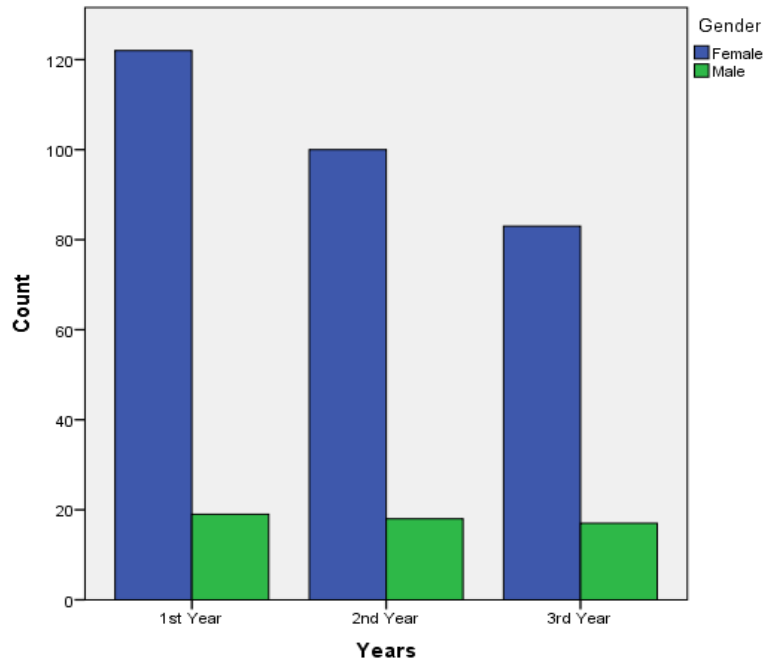

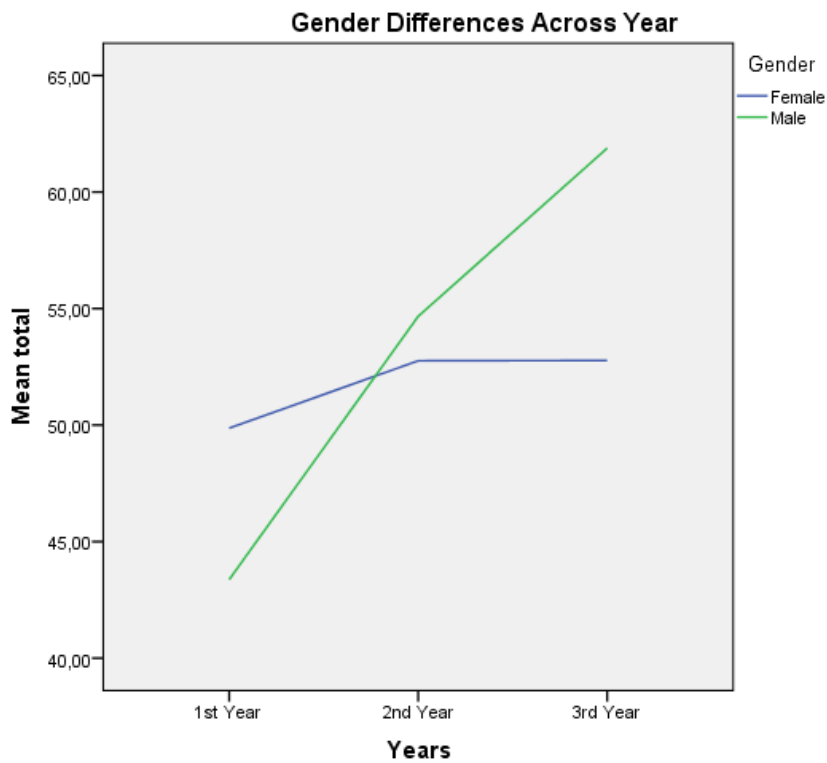**Figure 3**. The ratio of male and female students of the sample taken

Figure 4. The development of proof comprehension across gender and year

5. Conclusion

This pilot study is the prior study of proving ability. This study comes up with four main conclusions. Firstly, Several phases involved in the development of the instrument of proof comprehension produce acceptable reliability by test-retest reliability. Secondly, the higher the school level of the student, their ability to understanding proof becomes better, even though students of the second year or higher do not show a significant difference in proof comprehension ability. The third conclusion is that there are no gender differences in proof comprehension. However, there is an interaction between the variable gender and school level even though the reason what makes such effect still undetermined. The last conclusion is that the contribution of basic knowledge and logical reasoning has not shown the significant value, as explained in the previous study, that the contribution is relatively low.

6. References

[1] Weber K. Student difficulty in constructing proofs: The need for strategic knowledge. *Educ Stud Math* 2001; 48: 101–119.

[2] Stavrou SG. Common Errors and Misconceptions in Mathematical Proving by Education Undergraduates. *Issues Undergrad Math Prep Sch Teach*; 1, https://eric.ed.gov/?id=EJ1043043 (2014, accessed 30 August 2019).

[3] Sari CK, Waluyo M, Ainur CM, et al. Logical errors on proving theorem. *J Phys Conf Ser* 2018; 948: 012059.

[4] Zazkis D, Weber K, Mejía-Ramos JP. Two proving strategies of highly successful mathematics majors. *J Math Behav* 2015; 39: 11–27.

[5] Lew K, Zazkis D. Undergraduate mathematics students' at-home exploration of a prove-or-disprove task. *J Math Behav* 2019; 54: 100674.

[6] Inglis M, Mejia-Ramos JP, Simpson A. Modelling mathematical argumentation: the importance of qualification. *Educ Stud Math* 2007; 66: 3–21.

[7] Conradie J, Frith J. Comprehension tests in mathematics. *Educ Stud Math* 2000; 42: 225–235.

[8] Cowen C. Teaching and testing mathematics reading. *Am Math Mon* 1991; 98: 50–53.

[9] Mejía-Ramos JP, Inglis M. Argumentative and proving activities in mathematics education research. In: Lin F-L, Hsieh F-J, Hanna G, et al. (eds) *Proceedings of the ICMI Study 19 conference: Proof and Proving in Mathematics Education*. Taipei, Taiwan, 2009, pp. 88–93.

[10] Martin WG, Harel G. Proof frames of preservice elementary teachers. *J Res Math Educ* 1989; 20: 41–51.

[11] Segal J. Learning about mathematical proof: Conviction and validity. *J Math Behav* 1999; 18: 191–210.

[12] Healy L, Hoyles C. A Study of Proof Conceptions in Algebra. *J Res Math Educ* 2000; 31: 396.

[13] Weber K. Mathematics Majors' Perceptions of Conviction, Validity, and Proof. *Math Think Learn* 2010; 12: 306–336.

[14] Knuth EJ. Proof as a Tool for Learning Mathematics. *Math Teach* 2002; 95: 486–490.

[15] Selden A, Selden J. Validations of Proofs Considered as Texts: Can Undergraduates Tell Whether an Argument Proves a Theorem? *J Res Math Educ* 2003; 34: 4.

[16] Alcock L, Weber K. Proof validation in real analysis: Inferring and checking warrants. *J Math Behav* 2005; 24: 125–134.

[17] Powers R, Craviotto C, Grassl R. Impact of proof validation on proof writing in abstract algebra. *Int J Math Educ Sci Technol* 2010; 41: 501–514.

[18] Alcock L, Hodds M, Roy S, et al. Investigating and Improving Undergraduate Proof Comprehension. *Not Am Math Soc* 2015; 62: 742–752.

[19] Hodds M, Alcock L, Inglis M. Self-Explanation Training Improves Proof Comprehension. *J Res Math Educ* 2014; 45: 62–101.

[20] Weber K, Fuller E, Mejia-Ramos JP, et al. Do generic proofs improve proof comprehension? *15th Annual Conference on Research in Undergraduate Mathematics Education*.

[21] Waluyo M, Vidakovich T, Ishartono N, et al. A Review of Assessing Mathematical Proving Ability. In: *Proceedings of the Proceedings of the 4th Progressive and Fun Education*

International Conference, Profunedu 2019, 6-8 August 2019, Makassar, Indonesia. EAI. Epub ahead of print 2019. DOI: 10.4108/eai.7-8-2019.2288436.

[22] Yang K-L, Lin F-L. A model of reading comprehension of geometry proof. *Educ Stud Math* 2008; 67: 59–76.

[23] Mejia-ramos JP, Fuller E, Weber K, et al. An assessment model for proof comprehension in undergraduate mathematics. 2012; 3–18.

[24] Gliner JA, Morgan GA, Leech NL. *Research Methods in Applied Settings*. Routledge. Epub ahead of print 23 June 2009. DOI: 10.4324/9780203843109.

[25] Lin F-L, Yang K-L. The Reading Comprehension of Geometric Proofs: The Contribution of Knowledge and Reasoning. *Int J Sci Math Educ* 2007; 5: 729–754.