Student performance on proof comprehension tests in transition-to-proof courses

Juan Pablo Mejía-Ramos Keith Weber
Rutgers – The State University of New Jersey

As part of a project aimed at designing and validating three proof comprehension tests for theorems presented in a transition-to-proof course, we asked between 150 and 200 undergraduate students in several sections of one of these courses to take long versions (20 to 21 multiple-choice questions) of these tests. While analysis of these data is ongoing, we discuss preliminary findings about psychometric properties of these tests and student performance on these proof comprehension measures.

Key words: Proof reading, Proof comprehension assessment, Transition to proof

Most advanced mathematics courses are taught in a "definition-theorem-proof" format, where textbooks and lecturers present the definitions of new concepts and then prove theorems about those concepts (Weber, 2004; see also Dreyfus, 1991; Mills, 2011). Underlying this widely used pedagogical format is the assumption that mathematics majors can learn a great deal by studying the proofs that their professors present. Yet as mathematicians and mathematics educators observe, students' understandings of the proofs that they read are rarely assessed in a meaningful way (e.g., Conradie & Frith, 2000; Cowen, 1991). This is due, in great part, to the dearth of valid assessments that measure proof comprehension (cf., Cowen, 1991; Weber, 2012).

In a recent project we are developing proof comprehension tests for three theorems that are commonly presented in undergraduate transition-to-proof courses. Our aim is to validate these multiple-choice tests and make them available for others to use in their own courses and research projects.

Literature Review

Prior to our work in the area, we are not aware of the existence of systematic ways of assessing students' comprehensions of proof in undergraduate mathematics courses. However, there have been three important contributions in this area in the research literature. Conradie and Frith (2000) directly addressed the issue of proof comprehension tests in undergraduate mathematics. In addition to stressing their importance, these researchers provided comprehension tests for two different proofs. While their items were intriguing and called mathematics educators' attention to an underrepresented area of research, we note that these tests seemed to be created in a somewhat *ad hoc* manner, and it was unclear how these items were generated or what specific skills or understanding each item was designed to assess. Yang and Lin (2008) developed a model of reading comprehension for geometry proofs (RCGP) that consisted of four levels of understanding: *surface* (i.e., understanding the meaning of terms and statements), *recognizing the elements* (i.e., knowing whether a statement was an axiom, assumption, definition, or deduction), *chaining elements* (i.e., seeing how new statements are deduced from previous ones), and *encapsulation* (i.e., viewing the proof as a whole to comprehend the higher level ideas in the proof). Yang and Lin developed specific assessment items to assess the first three levels of understanding of a given proof, but notably did not attempt to assess how well students encapsulated the proof.

While some of the ideas in Yang and Lin's model are pertinent to proofs at the undergraduate level, we argue that the model by itself is not sufficient to probe students' understanding of a proof in an advanced mathematics class. For instance, Yang and Lin did not attempt to assess if students achieved the highest level of understanding in their model, which consisted of viewing the proof as a whole to comprehend its higher-level ideas. While this type of understanding might not be a central concern for a high school geometry teacher, we contend that skills such as being able to summarize a proof or being able to flexibly apply the methods of a proof to prove a new theorem are crucial skills for students in advanced mathematics classes. Further, there are also logical nuances that are present in some undergraduate proofs that are not accounted for in Yang and Lin's assessment model, such as how should proof by contradiction or proof by cases be understood. From Yang and Lin's perspective, this is not important as such proofs are rare in high school geometry classes, but they are common in undergraduate mathematics classes.

To address this limitation, Mejia-Ramos et al. (2012) built upon Yang and Lin's model to develop an assessment model for proof comprehension that is more suitable to the context of undergraduate mathematics. The components of their assessment model can be separated into two groups. The first group concerns *local* understandings of the proof, meaning that these questions can be answered by focusing on a small number of statements within the proof. In general, these questions would be concerned with describing the logical structure or evaluating the validity of the proof, and are adaptations of the first three components of Yang and Lin's (2008) model for geometry proofs. These local types of assessment items are:

- *Meaning of terms and statements*: items of this type measure students' understanding of key terms and statements in the proof.
- *Logical status of statements and proof framework*: these questions assess students' knowledge of the logical status of statements in the proof and the logical relationship between these statements and the statement being proven.
- *Justification of claims*: these items address students' comprehension of how each assertion in the proof follows from previous statements in the proof and other proven or assumed statements.

The second group concerns *holistic* understandings of the proof. In contrast with local understandings, one would not be able to answer questions about holistic understandings of a proof by focusing on a small number of statements in a proof, but would have to be addressed by inferring the ideas or methods that motivated the proof in its entirety. The holistic understandings relate to the "encapsulation" level in Yang and Lin's (2008) model, and include four types of assessment items that address students' understanding of the proof as a whole:

- *Summarizing via high-level ideas*: these items measure students' grasp of the main idea of the proof and its overarching approach.
- *Identifying the modular structure*: items of this type address students' comprehension of the proof in terms of its main components/modules and the logical relationship between them.
- *Transferring the general ideas or methods to another context*: these questions assess students' ability to adapt the ideas and procedures of the proof to solve other proving tasks.
- *Illustrating with examples*: items of this type measure students' understanding of the proof in terms of its relationship to specific examples.

Lecturers, textbook writers and researchers can use Mejia-Ramos et al.'s (2012) model to generate open-ended items that measure students proof comprehension along dimensions that are valued by both mathematicians and mathematics educators. However, this way of using the model has two shortcomings. First, these open-ended questions can be time consuming to generate and grade, which may limit their utility in teaching and research situations involving a large number of test takers. Second, the validity and reliability of these questions has yet to be verified. The purpose of our current project is to address both shortcomings by designing and validating multiple-choice proof comprehension tests for three proofs from a transition-to-proof course.

Methods

Materials

We have developed comprehension tests for proofs of the following three theorems:

- *Theorem 1:* The set of prime numbers is infinite.
- *Theorem 2:* Every third Fibonacci number is even. That is, if we define the n^{th} Fibonacci number (denoted by f_n) in the usual way, then f_{3k} is even for every $k \in N$.
- *Theorem 3:* The open interval $(0,1)$ is uncountable.

Theorem 1 is a central theorem in transition-to-proof courses, being one of the first indirect proofs that students encounter. The proof is also often misunderstood by students, as there is frequently confusion as to whether the constant generated in this proof is a prime number (e.g., Hazzan & Zazkis, 2003). The proof of Theorem 2 is a typical example of the kind of proofs by induction presented in transition-to-proof courses. Proofs by induction are a central concept in these courses and is notoriously difficult for students (e.g., Dubinsky, 1987, 1989; Harel, 2001). Theorem 3 is a more advanced theorem with a more sophisticated proof method that is usually covered in transition-to-proof courses.

In order to generate the current version of the multiple-choice, proof comprehension tests, we followed the following procedure:

1. For each one of the three proofs, we first generated open-ended questions of each type of assessment item in Mejia-Ramos et al.'s (2012) model.
2. We then conducted task-based interviews with 12 mathematics majors who had recently completed a transition-to-proof course. These participants were asked to read the three proofs and answer the open-ended questions.
3. We observed the correct answers that the participants provided as well as common incorrect answers. These data were used as a basis to generate a larger set of multiple-choice questions, with at least one question for every dimension of understanding in Mejia-Ramos et al.'s (2012) proof assessment model.
4. We then sought feedback from mathematicians at our institution and an advisory board (which included a prominent mathematician and a leading researcher on proof comprehension a the undergraduate level) regarding the accuracy and appropriateness of our items.
5. Finally, we piloted these multiple-choice items with 12 mathematics majors to make sure our items had appropriate wording and choices.

To illustrate the type of items in our proof comprehension tests consider the proof used for Theorem 1:

Suppose the set of primes is finite. Let $p_1, p_2, p_3, \ldots, p_k$ be all those primes with $p_1 < p_2 < p_3 < \cdots < p_k$. Let n be one more than the product of all of them. That is, $n = p_1 \cdot p_2 \cdot p_3 \cdots p_k + 1$. Then n is a natural number greater than 1, so n has a prime divisor q. Since q is prime, $q > 1$. Since q is prime and $p_1, p_2, p_3, \ldots, p_k$ are all the primes, q is one of the p_i in the list. Thus, q divides the product $p_1 \cdot p_2 \cdot p_3 \cdots p_k$. Since q divides n, q divides the difference $n - p_1 \cdot p_2 \cdot p_3 \cdots p_k$. But this difference is 1, so $q = 1$. From the contradiction $q > 1$ and $q = 1$, we conclude that the assumption that the set of primes is finite is false. Therefore, the set of primes is infinite.

Item type	Open-ended items	Multiple-choice items
Meaning of terms and statements	1. Please give an example of a finite set and explain why it is finite. 2. Please give an example of a set that is infinite and explain why it is infinite.	Which of the following are examples of finite sets? Please select **all** that apply. a) The set with the following elements: 1, 2, and 3. b) The set of real numbers between -2 an 2 c) The set of all fractions $\frac{1}{r}$ where r is a natural number. d) The set of integers greater than 4.5 and smaller than 9999.
Justification of claims	1. Why is it valid to conclude that n is a natural number? 2. Why does n have to have a prime divisor? 3. Why exactly can one conclude that if q is prime, then $q>1$?	In the proof, why is it valid to conclude that n is a natural number? Please select **the best** option. a) Because the product and sum of natural numbers is a natural number. b) Because n is greater than 0. c) Because $1, p_1, p_2, \ldots, p_k$ are all integers. d) Because it is a given in the proof that n is a natural number.
Summarizing via high-level ideas	1. Summarize in your own words the main idea of this proof. 2. What do you think are the key steps of the proof? 3. Give a three-sentence description of how the proof established the theorem.	Which of the following options **best** summarizes the main idea of this proof? [1] a) The main idea of the proof is to show that if the set of primes were finite, one could find a formula for a new prime number that is not in that finite set, contradicting the assumption. b) The main idea of the proof is to assume that the set of prime numbers is finite and to construct a natural number that has a prime divisor equal to 1, which is impossible.
Transferring the general ideas or methods to another context	1. In the proof, we define $n = p_1 \cdot p_2 \cdots p_k + 1$. Would the proof still work if we instead defined $n = p_1 \cdot p_2 \cdots p_k + 31$? Why? 2. Define the set $S_k = \{2, 3, 4, \ldots, k\}$ for any $k > 2$. Using the method of this proof, show that for any $k > 2$, there exists a natural number greater than 1 that is not divisible by any element in S_k.	In the proof, we define $n = p_1 \cdot p_2 \cdots p_k + 1$ Would the proof still work if we instead defined $n = p_1 \cdot p_2 \cdots p_k + 31$? Please select **the best** option. a) Yes, because n will still be a prime number, so the contradiction will still hold. b) Yes, because 31 is a prime number, which means that q must still be 1. c) No, because this definition of n would not be necessarily prime. d) No, because in this case q could be 31, which does not lead to a contradiction.

Table 1. Examples of items used in the proof comprehension tests for Theorem 1.

[1] This item has two other foils that did not fit in the table/proposal.

Table 1 contains examples of open-ended and multiple-choice versions of some of the different types of items used in the test for this proof. The multiple-choice tests generated for theorems 1 and 2 contained 20 items each, while the test for Theorem 3 contained 21 questions.

Participants and procedure

The proof comprehension tests were distributed to students in several sections of an undergraduate transition-to-proof course in a large state university. Each of the five participating instructors allocated 40 minutes of class to distribute each test. On the day each test was distributed, students in the course received a packet that contained the theorem and its proof, instructions on the different types of items in the test, and all the multiple-choice questions (the order of the items in each section of the test was randomized). The test for Theorem 1 was distributed after instructors had introduced proofs by contradiction in class (approximately a third of the way into the term), the test for Theorem 2 was distributed once instructors had discussed the principle of mathematical induction (usually by the middle of the term), and the test for Theorem 3 was distributed by the end of term, after instructors had discussed the notion of the cardinality of sets.

A total of 201 students took the proof comprehension test for Theorem 1, 192 students took the test for Theorem 2, and 152 students took the test for Theorem 3. [2]

Preliminary Results

Analysis of this data set is on going. However, preliminary analyses suggest several interesting trends:

1. There is a strong correlation between students' performance on any two of the three proof comprehension tests,
2. The tests, even before excluding poor or uninformative items, show a high internal consistency.

Taken together, these results suggest that proof comprehension can be a meaningful single-dimensional construct. Ongoing analyses will explore the extent that this is the case. We will also discuss items that the large majority of students answered correctly and the items that most students answered incorrectly, which can provide some much needed baseline data on how well mathematics majors understand proof in a transition-to-proof course.

Questions for the audience

1. Do you have any suggestions for further analysis of the data?
2. How would you recommend that we disseminate these tests to mathematicians?
3. How might we improve the test design process for future iterations of these types of studies?

[2] The decreasing number of participating students was not only due to the regular reduction of class size as the term progresses. One of the participating instructors did not reach the topic of cardinality in class, which meant that we could not distribute the test for Theorem 3 in the two sections led by this instructor.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. DUE-1245626.

References

Conradie, J., & Frith, J. (2000). Comprehension tests in mathematics. *Educational Studies in Mathematics, 42*, 225–235.

Cowen, C. (1991). Teaching and Testing Mathematics Reading. *The American Mathematical Monthly, 98*(1), 50-53.

Dreyfus, T. (1991). Advanced mathematical thinking processes. In D. Tall (Ed.) *Advanced Mathematical Thinking* (pp. 215-230). Kluwer: The Netherlands.

Dubinsky, E. (1987). On teaching mathematical induction I. *Journal of Mathematical Behavior, 6*(1), 305-317.

Dubinsky, E. (1989). On teaching mathematical induction I. *Journal of Mathematical Behavior, 8*, 285-304.

Harel, G. (2001). The Development of Mathematical Induction as a Proof Scheme: A Model for DNR-Based Instruction. In S. Campbell & R. Zaskis (Eds.). *Learning and Teaching Number Theory* (pp. 185-212). Dordrecht, The Netherlands: Kluwer.

Hazzan., O. & Zazkis., R. (2003). Mimicry of proofs with computers: The case of Linear Algebra. *Intenational Journal of Mathematics Education in Science and Technology 34*(3), 385-402.

Mejia-Ramos, J. P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics, 79*(1), 3-18.

Mills, M. (2011). Mathematicians' pedagogical thoughts and practices in proof presentation. *Presentation at the 14th Conference for Research in Undergraduate Mathematics Education.* Available for download from: http://sigmaa.maa.org/rume/crume2011/Preliminary_Reports.html

Weber, K. (2004). Traditional instruction in advanced mathematics courses. *Journal of Mathematical Behavior, 23*(2), 115-133.

Weber, K. (2012). Mathematicians' perspectives on their pedagogical practice with respect to proof. *International Journal of Mathematics Education in Science and Technology, 43*(4), 463-482.

Yang, K.-L., & Lin, F.-L. (2008). A model of reading comprehension of geometry proof. *Educational Studies in Mathematics, 67*, 59–76.