**Mathematics Majors' Evaluation of Mathematical Arguments and Their Conceptions of Proof**

Keith Weber
Rutgers University
keith.weber@gse.rutgers.edu

## 1. Introduction

In the last two decades, there has been a tremendous growth in educational research on mathematical proof. Researchers have investigated many issues related to proof, including the epistemological nature of proof (e.g., Hanna, 1991), the role of proof in mathematics classrooms (e.g., Hersh, 1993), students' conceptions of proof (e.g., Harel & Sowder, 1998), students' difficulties with proof (e.g., Moore, 1994), and the construction of proofs (e.g., Weber, 2001; Martin & McCrone, 2003; Weber & Alcock, 2004). Several researchers have noted that one topic that has received relatively little attention is the reading of proof (Selden & Selden, 2003; Hazzan & Zazkis, 2003; Mamona-Downs, 2005; Alcock & Weber, 2005).

Educational research on the reading of proof has primarily focused on how individuals evaluate proofs for correctness. Several studies show that both students and teachers have difficulty with this task. Selden and Selden (2003) and Alcock and Weber (2005) found that undergraduate mathematics majors performed at chance level when asked to determine if a proof was correct. Martin and Harel (1989) demonstrated that many pre-service elementary teachers would accept a geometry proof as valid if it was in the standard two-column format and invalid if it was in paragraph form, regardless of its mathematical content. Knuth (2002) also found that in-service high school teachers would accept flawed arguments as proofs if they were in an appropriate format. In a separate line of research, I examined the ways in which mathematicians determined if an argument was a valid proof (Weber, 2008). This study revealed that proof validation was not a formal or exclusively deductive activity. Mathematicians sometimes sought to bridge a perceived gap in the proof not by constructing a sub-proof, but by an informal argument or the inspection of examples. Also, the mathematicians' conceptual knowledge of the domain they were studying influenced how they checked an argument for correctness.

The lack of research on how students *do* read mathematical arguments, as well as how they *should* read them, represents an important void. Advanced mathematics courses are typically taught in a definition-theorem-proof format (cf., Weber, 2004) in which students spend much of their time observing proofs that their professor presents for them and reading proofs in their textbooks. This practice is based upon the implicit assumption that students can learn mathematics by reading the proofs of others (Selden & Selden, 1995). Reform-oriented K-12 classrooms place a high value on having students attend to, evaluate, and critique the arguments of their teacher and classmates (e.g., NCTM, 2000) suggesting that the observation and evaluation of arguments plays an important role here too.

This paper reports a study in which 28 mathematics majors were observed reading and evaluating ten mathematical arguments. I will use this data to address four questions:

(1) What types of written arguments do mathematics majors find convincing?

(2) To what extent do mathematics majors have specific skills, as suggested by mathematics education researchers, needed to evaluate arguments?

(3) What relationship do the students perceive between understanding, conviction, and mathematical proofs?

(4) What do mathematics majors perceive their responsibility to be when they are reading a proof?

## 2. Related literature

### 2. 1. Students' standards of conviction

The way a student evaluates a mathematical argument is necessarily dependent on how he or she becomes convinced that a mathematical assertion is true. In an influential paper, Harel and Sowder (1998) introduce the notion of *proof schemes* as the ways in which students remove personal doubts about the truth of mathematical statements or attempt to convince others that a statement is true. Harel and Sowder (1998, 2007) contend that mathematicians have *deductive proof schemes*—mathematicians become convinced and persuade others of the truth of a mathematical assertion by deductive reasoning. However, research studies document that students often are convinced by non-deductive arguments. To Harel and Sowder (2007), a primary goal of mathematics instruction is leading students to adopt the same proof schemes as mathematicians.

Harel and Sowder argue that many students hold *empirical proof schemes*, meaning that they seek to verify that a general claim about a large number of mathematical objects is true by checking that the claim holds true for specific examples. Research studies document that students often behave in this way when they are asked to prove a general assertion (e.g., Harel & Sowder, 1998; Recio & Godino, 2001). Further, many students prefer example-based arguments to deductive arguments, even though some believe that deductive arguments would receive more credit from teachers (e.g., Segal, 200; Healy & Hoyles, 2000; Raman, 2002). Many mathematics educators believe that empirical proof schemes are deeply held beliefs that are not likely to change by direct instruction. Several researchers have designed and tested instruction whose sole goal was to lead students to modify or abandon their empirical proof schemes (e.g., Harel, 2001; Stylianides & Stylianides, 2008).

Harel and Sowder (1998) argue that some students hold *perceptual proof schemes*, and may be convinced that a statement is true by a diagram or graph. However, the extent to which this is true for mathematics majors is unclear. In a recent study, Inglis and Mejia Ramos (2009a) found that some mathematics majors hold the opposite view—that pictures can never be convincing and all arguments that rely on a picture are inappropriate.

Research also suggests that students are often not convinced by deductive arguments, viewing a correct proof merely as evidence that an assertion is true rather than establishing its necessity (e.g., Fischbein, 1982; Chazan, 1993). When students read proofs, they often focus on whether the article has the *appearance* of deductive arguments that they have seen in the past, such as the use of a two column format in geometry or the appearance of mathematical symbols (e.g., Martin & Harel, 1989), rather than whether the content of the argument makes sense.

While there have been many studies of students' proof schemes and conceptions of proof, few studies have investigated these issues with upper-level mathematics majors in a systematic matter. One question addressed in this paper is the extent to which the findings reported in this section hold true for mathematics majors.

2. 2. Skills in reading mathematical arguments and proofs

Research on how students *can* or *should* validate proofs has been limited. Selden and Selden (1995) emphasized that students' ability to discern the logical structure of informal mathematical statements is necessary both for constructing and validating proofs. These researchers also stressed that when one is reading a proof, one needs to determine if a legitimate *proof framework* (e.g., direct proof, proof by contradiction) is being employed (Selden & Selden, 1995, 2003). Weber and Alcock (2005) argued that when one considers whether a new assertion in a proof follows legitimately from previous statements, one first needs to determine what statements are used to support the new assertion and what general mathematical principle, or warrant, specifies how the new assertion can be deduced from these statements. In cases where this warrant is not stated, it must be inferred by the reader. Judging whether a new assertion follows validly from previous ones therefore involves judging whether a possibly inferred warrant is a valid mathematical principle that is acceptable to one's mathematical community. A subsequent research study illustrated how most students do not infer and check warrants when validating proofs (Alcock & Weber, 2005). The extent to which mathematics majors can evaluate proof frameworks and infer warrants will be investigated in this study.

2. 3. Responsibility in proof reading

Students enter mathematics classrooms with implicit understandings of what their responsibilities are and how they are expected to behave. The way students engage in a mathematical task is significantly influenced by these belief systems. For instance, Schoenfeld (1985) observes that if students believe their role as problem solvers is to quickly implement a procedure that they had just been taught, then they may not generate new representations or use heuristics when solving a problem, even if that problem invites them to do so.

There has been little research on what students perceive their responsibility to be when reading proofs. Herbst and Brach (2006) argue that high school geometry students believe their responsibility in writing proof is to illustrate to the teachers the extent that they can engage in logical reasoning. As a result, some students reject tasks in which they are asked to determine if a statement is true or choose a set of hypotheses that would necessitate a desired conclusion. If the teacher wants to measure students' logical abilities, then it is the teacher's responsibility to provide them with clearly specified statements to prove. This paper explores what mathematics majors perceive their responsibility to be when reading mathematical arguments.

## 3. Methods

### 3. 1. Participants

28 mathematics majors who had recently completed a transition-to-proof course agreed to participate in this study. All were mathematics majors in their sophomore or junior year. Each student was paid a small fee for his or her participation.

### 3. 2. Materials

Participants were asked to read the ten proofs in the Appendix of this paper. These arguments varied along the following dimensions:

- *The mathematical content.* Arguments came from elementary algebra (1, 2, and 10), elementary number theory (3, 5, and 7), and calculus (4, 6, 8, 9).

- *The format of the argument.* The arguments were presented in paragraph form (2, 4, 5, and 6), using a logical, symbolic format (1, 3, 7, 8, 9, and 10), and in some cases, accompanied with a diagram (2, 6).
- *The mode of argumentation.* Most arguments used deductive reasoning, but argument 5 was empirical and arguments 2 and 6 were perceptual.
- *The validity of the argument.* Arguments 1, 3, and 4 were valid proofs. Argument 5 was invalid because it relied on empirical reasoning. Argument 6 was invalid because it relied on perceptual reasoning. Argument 10 contained an algebraic error in the third line of the argument. The flaws in arguments 7, 8, and 9 are discussed below. Argument 2 relies on perceptual reasoning and is thus arguably invalid; however some might consider it a legitimate "proof without words" (e.g., Nelsen, 1993). Hence no judgment is made in this paper on its validity.

Some arguments were based on specific errors suggested by the literature. Argument 7 was taken from Selden and Selden's (2003) study on proof validation and assessed whether students could recognize if a proof used a valid proof framework (i.e., if it used appropriate assumptions and deduced an appropriate conclusion). Argument 8 also used an invalid assumption for writing a proof by cases. Argument 9 was similar to the task used in Alcock and Weber (2005) to determine if students would check if implications within the proof were warranted (i.e., if there was a valid mathematical principle for how a new statement could be deduced from previous ones).

The goal of this study was to examine the ways mathematics majors read arguments, not assess their mathematical content knowledge. As such, several measures were taken to minimize the possibility that the participants would have difficulty with an argument because of a lack of background knowledge. The arguments all were related to basic concepts from domains that the participants had previously studied. All of the arguments were straightforward and none contained subtle tricks. Twenty arguments were originally generated for the study. If any of the participants had difficulties with the content of the argument during pilot studies, that argument was not used in the study.

### 3. 3. Procedure

Each participant met individually with the author for a task-based interview. Participants were told they would be presented with ten arguments, one at a time, and that they would be asked to make three judgments on each argument. First, they were asked to rate on a five-point scale the extent to which they felt they understood the argument, where a 5 indicated that they understood the argument completely. Second, they were asked to rate how convinced they were by the argument using a five-point scale, where a 5 indicated they were completely convinced. Third, they were asked to decide whether the argument was a proof. They were given four choices: (1) The argument was a rigorous proof, (2) it was a non-rigorous proof (3) it did not meet the standards of a proof, or (4) they were not sure because they did not fully understand the argument. They were also permitted to opt for "other" if they did not feel comfortable with these four choices.

It was emphasized to the participants that their judgments on the arguments should come from what was contained in the argument, and not from their knowledge of whether the claim being proven was true or false. Specifically, participants were told that even if they knew the claim was true, if they found the argument to be unconvincing, then they should rate it as such. Participants were also informed that some of the arguments would be “good arguments” while others would be “flawed”. The participants were told that they should spend as much time as they liked while reading the arguments.

Participants were given each argument individually and asked to “think aloud” as they read the argument and made their judgments. If participants claimed they did not understand an argument or did not find it fully convincing (i.e., gave a mark less than a five for these two judgments), or did not find the argument to be a proof, they were asked why they gave it that mark.

After reading all ten arguments, participants were asked a series of open-ended questions about how they read arguments (e.g., “What are some of the things that you do when you read a mathematical argument?”) or about their perceptions of mathematical arguments (e.g., “What do you think makes a good mathematical argument?”).

Finally, if the participant had judged argument 8 to be valid, his or her attention was directed to the problematic line in the argument, “Either $f(x) \geq 0$ or $f(x) < 0$”. (The scope of $x$ was the interval $[a, b]$, so it is possible for a function to assume both positive and negative numbers on this interval). The participant was asked if he or she saw a specific problem with this line of the argument. Similarly, if the participant judged argument 9 to be valid, his or her attention was directed to the last two lines in the argument, in which the statement “ $f(x) \to \infty$ as $x \to \infty$” was deduced from the fact that $f(x)$ was increasing (increasing functions are not necessarily bounded above) and asked if they saw a problem with this inference.

3. 4. Analysis

Each participant’s justification for why an argument was not fully convincing or not a proof was grouped using the constant-comparative method (Strauss & Corbin, 1990). Each justification was given an initial description. Similar justifications were grouped together and given preliminary category names and definitions. New episodes were placed into existing categories when appropriate, but also used to create new categories or modify the names or definitions of existing categories. This process continued until a set of categories was formed that were grounded to fit the available data.

4. Results

The quantitative results from this study are presented in Table 1. Participants collectively and correctly judged arguments 1, 3, and 4 to be valid proofs in 74 of their 79 judgments[1], or 95% of the time. They collectively and incorrectly judged arguments 5, 6, 7, 8, 9, and 10 to be valid proofs in 74 out of 130 instances, or 57% of the time. For argument 2, which could arguably be considered a proof without words, 24 of the 28 participants (86%) of the participants believed the argument was a proof. The findings that students can correctly identify valid proofs but have trouble identifying the flaws in invalid arguments has also been reported in Knuth's (2002) study with in-service high school teachers.

*Empirical proof schemes.*

The participants in this study did not exhibit empirical proof schemes. 26 of the 28 participants believed argument 5, which relied on examples, was not a proof. Only one participant found the argument to be completely convincing, while 22 participants gave the argument a score of 3 or less for being convincing. The average score for convincingness was 2.61, easily the lowest of any of the arguments used in the study.

[1] Instances in which participants marked "not sure" were not included in this tally.

After reading the arguments, participants were asked to name their favorite two arguments and their least favorite argument. None of the 28 participants named argument 5 as one of their two favorite arguments while 11 named argument 5 as their least favorite argument.

**Table 1. Summary of results**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Understanding | | | | | | | | | | |
| 5 | 23 | 26 | 22 | 14 | 25 | 22 | 27 | 22 | 20 | 28 |
| 4 | 2 | 1 | 4 | 13 | 2 | 4 | 1 | 3 | 4 | 0 |
| 3 or less | 3 | 1 | 2 | 1 | 1 | 2 | 0 | 3 | 4 | 0 |
| Conviction | | | | | | | | | | |
| 5 | 22 | 22 | 16 | 16 | 1 | 11 | 14 | 15 | 20 | 12 |
| 4 | 2 | 4 | 5 | 10 | 5 | 7 | 2 | 3 | 4 | 2 |
| 3 or less | 2 | 2 | 7 | 2 | 22 | 10 | 12 | 9* | 4 | 16 |
| Rigorous proof | 15 | 10 | 17 | 23 | 0 | 3 | 12 | 15 | 11 | 11 |
| Non-rigorous | 9 | 14 | 7 | 4 | 2 | 9 | 4 | 2 | 8 | 1 |
| Not a proof | 0 | 4 | 4 | 0 | 26 | 14 | 12 | 9 | 5 | 16 |
| Unsure | 4 | 0 | 0 | 1 | 0 | 2 | 0 | 1 | 2 | 0 |
| Other | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 |

*- One student declined to rate argument 8 for conviction

## 4.1. Students' proof schemes

The data above is not consistent with research that reveals that younger students are convinced by examples and prefer example-based arguments to deductive ones (e.g., Healy & Hoyles, 2000; Segal, 2000). One hypothesis is that these participants learned in their transition-to-proof course that it was inappropriate to use examples in mathematical argumentation, which they may have assimilated as a rule without understanding of the limitations of empirical reasoning. However, this does not appear to be the case. When asked why the empirical argument was not fully convincing, in 26 cases the participants responded by articulating the limitations of empirical reasoning. Two representative responses of this are provided below:

P11: I am not convinced by this argument. The argument is wrong. It's a terrible argument, although it kind of shows a pattern that does make sense, so say I'm partially

convinced by the argument. But I would say this is not at all a rigorous proof, or proof at all. Does not meet the standards of a proof.
I: Why would you say it's not a proof?
P11: Because this only shows that 26, doesn't show any above that. You can't just make an assumption based on the first 26 numbers. Not even 26 numbers. It's a pretty small sample, since the real number line's infinitely long.

P2: I understood what they were saying. I'm not convinced since there's a lot more prime numbers then the ones that they presented. So it's not really a proof because it's a set that goes a lot larger than what they gave.

The data in this study do not address how students came to see the limitations of empirical reasoning. However, P27's responses below are suggestive of how this might be learned.

P27: Are you convinced by this argument? I'm going to go with a 3, because it shows only up to 26, and I feel like if you're going to use specific examples, it doesn't prove the claim is true for every single integer.
I: So there could be one…
P27: I mean it's kind of like an assumption in the proof, so because from 4 to 26 there's at least one set of primes that add up to that even integer… it's kind of like an assumption that yeah, ok, it happened for this case, so it's going to happen later on, and that might not be true by this proof.

In this excerpt, P27 clearly shows an understanding of why argument 5 is not a valid proof. However, later in the interview, she reveals that this is the way she would try to prove the statement before she took her introduction to proof course.

P27: I probably, in the beginning, when I was taking [the transition-to-proof course], I would have proved it like that, and then my professor probably would have murdered my answer. He would have said that my answer only proves from 4 to 26. Technically you only prove it from 4 to 26, so I mean I probably would have done that initially, but I don't think it proves it for all.

For this student, simply having the professor repeatedly explain to her why her arguments were wrong was sufficient for P27 to abandon her empirical proof scheme. The extent that this would work for mathematics majors is an open question, but arguably plausible since nearly all of the 28 participants in this study could articulate the weaknesses of empirical reasoning.

It is interesting to speculate why the participants in this study rejected empirical reasoning when many other studies found that students held empirical proof schemes (e.g., Harel & Sowder, 1998; Healy & Hoyles, 2000; Segal, 2000; Recio & Godino, 2001). The methods used in this study differed from other studies in two ways—the population being studied and the way that proof schemes were evaluated. First, this study examined mathematics majors in their sophomore or junior years. It is possible that mathematics majors are less likely to hold empirical proofs schemes that students who study other disciplines or that their experience in their undergraduate mathematics classes may have led these mathematics majors to refine their proof schemes. The interview excerpts with P27 provide suggestive evidence that the latter possibility may be true.

The other way that this study differed from many previous studies is that the latter inferred students' proof schemes from the arguments that students *produced* (Harel & Sowder, 1998; Recio & Godino, 2001) while I inferred students proof schemes from the arguments that students *evaluated*. Vinner (1997) warns that it is often inappropriate to infer students' beliefs from the wrong answers they produce as students might be providing these incorrect answers not because they believe they are right, but for other reasons, such as obtaining partial credit on a test or pleasing the interviewer. If students are asked to prove something and they don't know how to begin, they might plausibly show the claim works for specific examples for these reasons. After all, checking that a claim holds in specific cases increases the likelihood that the claim is correct. This does not necessarily imply that students believe that inspecting examples is an appropriate way to seek conviction.

*Perceptual proof schemes.*

Many of the participants in this study did show evidence of holding perceptual proof schemes. Argument 2 relied in a critical way on a drawing. Nonetheless, 24 of the 28 participants judged the argument to be a proof and 10 of these participants believed the proof was fully rigorous. For argument 6, 12 of the participants judged the argument to be a proof, although only three thought it was fully rigorous. Of the 14 participants who believed argument 6 did not meet the standards of proof, nine cited the use of a graph as a reason for rejecting the argument. The remaining five cited the fact that only part of the graph was shown; as a result, this argument was implicitly relying on empirical reasoning. Only four of the nine participants who rejected argument 6 because a graph was used cited specific limitations of using a graph to prove.

This result also contradicts a result from the literature. Inglis and Meija-Ramos (2009a) found that mathematics majors are less convinced by pictures than mathematicians because they have been taught the motto that "pictures do not prove". Yet 24 of the participants in this study thought argument 2 was a proof and 22 found it fully convincing. One possible reason for these differing results might be found in another study by Inglis and Mejia Ramos (2009b) which revealed that mathematicians find diagrammatic arguments more convincing if accompanying text explained the arguments. The perceptual arguments in this study had this feature while the one used in Inglis and Mejia-Ramos (2009a) did not.

4. 2. Skills in reading proofs

*Recognizing valid proof frameworks*.

Arguments 7 and 8 were designed to see whether the participants would attend to whether or not these arguments had a valid structure. In argument 7, the participants were shown a purported proof of the statement, "If $n^2$ is divisible by 3, then $n$ is divisible by 3". However this argument began by assuming the conclusion ($n$ is divisible by 3) and then deducing the hypothesis ($n^2$ is divisible by 3). Similar to the Selden and Selden (2003) study in which this argument was used, roughly half the participants (14 of 26) judged this to be a valid proof. Four of the fourteen who judged the argument to be a proof did recognize that the argument had an invalid proof structure. A representative excerpt from one student who judged the argument to be a proof is provided below:

P28: So shouldn't you… I feel like it should start off with, you know, assume $n^2$ is divisible by 3, then relate it to "n is divisible by 3", instead of saying… Like you assume what you are trying to prove, but… Wait, if n is divisible by 3, then… Yeah, I agree with the claim[…] So, I think the argument is right. I don't know if the formatting matters. Once again I'm kind of biased because of how I was taught. So convincing? I'm going to go with a 4 instead of a 5. Only because of particular preference but I don't think it's wrong.

This result replicates Selden and Selden's (2003) finding that many students do not check to see if arguments employ legitimate proof frameworks. Further, this extends the result by showing that even if students notice an unusual proof format is being used, they might not think this is important. In the excerpt above, P28 dismisses this concern as "formatting" which might not matter.

Argument 8 also is a purported proof with an invalid structure. The argument is a proof by cases. However the cases considered are not exhaustive. The statement, "either $f(x)\geq 0$ or $f(x)<0$" is not a valid assumption, since the scope of $x$ in this statement is the non-trivial interval $[a, b]$ and $f(x)$ could assume both positive and negative values on this interval. 17 of the 24 participants who made a judgment on this argument accepted it as a valid proof. However, when they were asked to focus on the specific invalid statement after they had read the arguments, only three of these 17 students could find a problem with it. Hence many of the participants may have lacked the content knowledge, but not necessarily the skills at reading arguments, to evaluate this argument correctly.

*Inferring warrants*.

The first seven lines of argument 9 consist of a mostly correct demonstration that *f(x)=ln x* is an increasing function. The eighth line of the proof states, "Therefore, $f(x) \to \infty$ as $x \to \infty$." This is not a valid conclusion because increasing functions do not necessarily diverge. 19 of the 24 participants who judged argument 9 believed it was a valid proof. After they read all the arguments, these 19 participants' attention was drawn to the last two lines of this proof and they were asked if they thought these lines were valid. At this point, 13 of the participants were able to explain why they were not. Two representative responses are provided below:

P4: Oh, I guess it's not true. It could approach uh, that's true. It could just approach an asymptote. I didn't, okay, I didn't consider that.

P18: Yeah, because it could just be getting bigger and bigger and bigger, but still converging to a specific number. Like if f(a)… if f(b) was 1, f(a) could be 1.5, and then your next f(a) could be 1.75 and then it could be 1.8, and then converging to like 2 or something.

Hence these participants had the mathematical knowledge to know that increasing sequences do not necessarily diverge. However, none of these 13 participants attempted to produce a reason for how the last line of the argument followed from previous assertions when reading the proof. One participant, P20, offers an explanation for why she did not check to be sure the last line of the proof followed from previous assertions.

P20, and perhaps other students, had a false sense of confidence because the argument until that point had been correct and made sense.

P20: The therefore. Yeah, like, I was reading all of [the lines of the proof] and I got it. But then, like, all of a sudden like this popped up. And I was reading it and like I believe like all of it was true and then the therefore popped up. So I believe that it was a part of it so then I thought it was true.

The finding that many mathematics majors will fail to recognize that an argument implicitly uses invalid mathematical principles but can do so if their attention is specifically focused on this aspect of the argument replicates the findings of Alcock and Weber (2005). It also challenges Reid and Inglis' (2005) claim that most students will naturally do this when reading mathematical arguments.

4. 3. Perceived relationship between understanding, conviction, and proof

Ideally students would not be sure an argument was a proof if they did not fully understand it. Likewise, they would not be sure an argument was a proof if they did not find it fully convincing. Further, if a student found a significant flaw in an argument, then he or she would recognize the argument as invalid. I found that mathematicians behaved in this way in a study in which I observed them evaluating the proofs of others (Weber, 2008). However the data in this study suggests this is not the case with mathematics majors.

*Judging an argument that is not fully understood to be a proof*

12 of the 28 participants judged an argument that they did not fully understand to be a proof. There were a total of 21 instances of this. 11 of these instances occurred for argument 4, which demonstrated that the equation $4x^3 - x^4 = 30$ has no solutions because the function $f(x) = 4x^3 - x^4$ had a global maximum value of 27. Participants found many

aspects of the argument confusing, but nonetheless thought the argument was a proof.

Two representative excerpts are provided below:

P7: Because f of x is a polynomial of degree 4 and the coefficient of x to the fourth is negative, f of x is continuous and will approach negative infinity as x approaches infinity or negative infinity. That whole sentence kind of confused me. But everything else makes sense. So, I will give that a four [in terms of understanding]. Um, am I convinced by this argument? I give it a five. Is it a proof? Yes.

I: What didn't you understand about it?
P16: I forgot what it means to find a global maximum. So I'm not quite sure. But everything else looks right. So I also give it a four for being convincing. Do I consider the argument to be a proof? Yes. And it's rigorous.

*Judging an argument to be a proof that is not fully convincing*

21 of the 28 participants judged an argument to be a proof that they did not find fully convincing. There were a total of 39 instances of this. In 10 of these instances, the participant was not sure about a particular step within the proof. For instance, when reading argument 3 in support of the statement, $n^3 - n$ was divisible by 6, P1 responded:

P1: Since n cubed minus n is even and divisible by 3, n cubed minus n is divisible by 6. I don't remember all of the rules for everything for all of those division things, so that's holding me back right now. So I guess I understand the argument for the most part. And I guess I can say that I am pretty much convinced […] I forgot what it means, the last part, like what numbers are divisible by three, divisible by 6, I forgot those rules.

In some sense, these participants seemed to be saying the argument was a proof conditionally, on the assumption that the statements that they were unsure about were true and followed validly from the previous assertions. Although such behavior is understandable, the mathematicians that I studied reading arguments did not do this. If they were unable to validate a single assertion in the proof, they would spend extended periods of time trying to justify why the assertion was legitimate. If they were unable to do so, they would not make a judgment on that proof. In contrast, the students in this study rarely spent more than two minutes reading an argument. They simply assumed the statement would be true.

In 13 instances, the participants could not spot anything particularly wrong with the proof but could not follow the general logic in the proof. This happened fairly often with argument 4 and is illustrated with P19's response below:

P19: For convincing, hmm, I'll give it a 4. Is it a proof? Yes, they've shown that [referring to $4x^3 – x^4$] has a maximum of 27 and is decreasing after that so it could never be equal to 30. Yes, it's a rigorous proof.
I: What didn't you find convincing about the argument?
P19: The whole, I didn't remember how you found global maximums. Each of the steps made sense though. I think it's right, but I'm not sure.

There were 6 instances in which the participants had a general sense that the argument was flawed, but could find no specific flaw within it. Consider the excerpt below:

P9: It's just not right, I think [...] I couldn't find something wrong here but for some reason it doesn't convince me. How about I put as a three? Neutral?
I: Okay. Even though you can't find anything wrong, you just don't find it convincing?
P9: I have to think about this one […] I think it is right, this proof. But I don't know. Let me just put I consider this argument to be a proof although not fully rigorous.

*Judging arguments with significant flaws to be proofs*.

10 of the 28 participants judged an argument to be a proof after they located a significant flaw in the proof. This occurred a total of 16 times. As discussed in section 4. 2, four of the participants judged argument 7 to be a proof, even though they recognized that it had an invalid proof framework. Six of the participants judged empirical or perceptual arguments to be proofs, even though they were uncertain whether this type of reasoning was legitimate in a proof and they did not find these arguments to be fully convincing. Three of the participants who judged argument 9 to be a proof commented that the last line of the proof did not follow from the previous inferences in the proof. In

general, some participants appeared to believe that if most of the steps in a proof were valid, then the proof in its entirety was valid. Likewise, if they could verify that most of the steps in a proof were correct, they could act on the assumption that the remaining steps that they could not validate were also correct. These participants did not seem aware that a single invalid statement could make the entire proof invalid; as such, they needed to check every step when reading an argument or proof.

4. 4. Students' responsibility when reading proofs

When participants were reading an argument, they focused predominantly on the local logical details of that argument. Specifically, they sought to determine what logical rule was being used to deduce each new assertion within that proof. Rarely did they consider the *semantic content* of the proof by using their own knowledge of the concepts in the proof to see if the proof made sense. There were only four instances of the participants drawing a diagram or graph. Many of the participants remarked that a graph would be useful for interpreting argument 4, yet most did not attempt to draw one. Similarly there were only five instances in which participants considered specific examples to help make sense of what was being asserted within a proof. In argument 3, $n^3 - n$ was factored into $(n-1)n(n+1)$ and from this, it was argued that $n^3 - n$ must be even, a multiple of 3, and hence a multiple of 6. Many participants indicated that they found these steps confusing and could not see why they were true. Only a few of them used specific examples to see if such assertions might be true. This can be contrasted with mathematicians who regularly use examples when reading these types of arguments (Weber, 2008).

Some participants indicated that they believed it was the author of a proof's responsibility to explicitly list all of the logical details of a proof. When asked what made a good logical argument, 11 participants stated that a logical explanation should accompany every step. Consider the two representative excerpts below:

P9: It's got to be really detailed, I think. You have to tell every detail. Every step, it is very clear. And have some kind of like form, ya know. I like doing things step by step. Like in a certain way. Very, uh, very formal way, ya know?
I: So you like having every detailed spelled out as much as possible?
P9: Yeah, yeah, yeah.

P23: The argument should be clear. You shouldn't reach a point in the proof where you say, "Whoa! Where did that come from?". The argument should tell you where it came from. Which theorem you used, or whatever.

This might explain, in part, why participants spent such a short time reading the arguments in this study. If they were confused by a particular argument, they considered that a fault of the argument, not of their own understanding.

## 5. Discussion

This paper presents a number of interesting findings about the ways in which mathematics majors read and evaluate proofs. First, most of the participants in this study did not exhibit an empirical proof scheme. In light of previous research, this is a surprising finding that would be useful to replicate. If this finding generalizes to mathematics majors at other universities, this has an important consequence for collegiate mathematics education research. A significant goal of contemporary mathematics education research is to have students be convinced by deductive arguments but not empirical ones (e.g., Harel & Sowder, 2007) and several researchers have designed interventions to lead students to do exactly that (e.g., Harel, 2002; Stylianides & Stylianides, 2008). However, the data suggests that this might not be necessary with mathematics majors. Completing a traditional transition-to-proof course might be sufficient to accomplish this.

Many of the participants in this study did obtain conviction from perceptual reasoning. Perhaps this is not a problem, as some mathematicians and philosophers also apparently gain high levels of conviction from graphs and diagrams (e.g., Nelsen, 1993; Giaquinto, 2007; Kulpa, 2009). However, students should recognize that perceptual arguments are (usually) not accepted as proofs by the mathematical community. For an argument to be considered a mathematical proof, it not only has to be convincing, but also use deductive reasoning so that the proven statement is deduced logically from accepted facts.

This study replicates the findings of other studies that illustrate that students lack particular proof validation skills (Selden & Selden, 2003; Alcock & Weber, 2005). While it is important to teach these skills, it would also be beneficial to explain why these skills are necessary to employ. Some participants appeared to believe that what the author chose as the assumptions and the conclusion of a proof were a minor detail. Others acted as if an isolated error within a long chain of argumentation was not significant. If mathematics majors do hold such beliefs, they likely would see no reason to implement these skills.

In a recent unpublished study in which I interviewed mathematics professors about their pedagogical practice, these professors indicated to me that they expected students to spend a lengthy time outside of class studying the proofs that were presented to them. One professor suggested that his difficult proofs might take students as long as two hours to understand. The data from this study suggest that students believe a proof should be transparently obvious. Many indicated that *every* logical detail should be explicitly stated and almost none of the participants spent more than two minutes reading an argument, even ones they found confusing and did not fully understand. If participants do not believe that there is a lot to be gained from understanding a proof but these gains take time to be achieved, it is doubtful they will see the benefits of learning the complex processes involved in proof reading and validation.

**References**

Alcock, L. & Weber, K. (2005). Proof validation in real analysis: Inferring and evaluating warrants. *Journal of Mathematical Behavior*, 24(2), 125-134.

Chazan, D. (1993). High school geometry students' justifications for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, 24, 359-387.

Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics*, 3(2), 9-18.

Giaquinto, M. (2007). *Visual thinking in mathematics: An epistemological study*. Oxford University Press: Oxford.

Hanna, G. (1991). Mathematical proof. In D. Tall (ed.) *Advanced Mathematical Thinking.* Dordrecht: Kluwer.

Harel, G. (2001). The development of mathematical induction as a proof scheme: A model for DNR-based instruction. In S. Campbell & R Zazkis (eds.) *Learning and teaching number theory*. Ablex Publishing Company: New Jersey.

Harel, G. & Sowder, L. (1998). Students proof schemes. *Research in collegiate mathematics education*, 3, 234-282.

Harel, G. & Sowder, L. (2007). Towards a comprehensive perspective on proof. In F. Lester (ed.) *Second handbook of research on mathematical teaching and learning*. NCTM: Washington, DC.

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

Healy, L. & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31, 396-428.

Herbst, P. & Brach, C. (2006). Proving and doing proofs in high school geometry: What is it that is going on for the students. *Cognition and Instruction*, 24, 73-122.

Hersh, R. (1997). *What is mathematics, really?* Oxford: Oxford University Press.

Inglis, M. & Mejia-Ramos, J.P. (2009a). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction*, 27, 25-50.

Inglis, M. & Mejia-Ramos, J.P. (2009b). On the persuasiveness of visual arguments in mathematics. *Foundations of Science,* 14, 97-110.

Knuth, E. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education*, 33(5), 379-405.

Kulpa, Z. (2009). Main problems of diagrammatic reasoning: Part 1. The generalization problem. To appear in *Foundations of Science*.

Mamona-Downs, J. & Downs, M. (2005). The identity of problem solving. *Journal of Mathematical Behavior*, 24(3/4), 385-401.

Martin, T.S. & McCrone, S. (2003). Classroom factors related to geometric proof construction ability. *The Mathematics Educator*, 7, 41-51.

Martin, W. G. & Harel, G. (1989). Proof frames of pre-service elementary teachers. *Journal for Research in Mathematics Education*, 20(1), 41-51.

Moore, R.C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27, 249-266.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.

Nelsen, R.B. (1993). *Proofs without words*. MAA: Washington, DC.

RAND Mathematics Study Panel (2003). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education*. Santa Monico, CA: RAND corporation.

Raman, M. (2002). *Proof and justification in collegiate calculus*. Unpublished doctoral dissertation. University of California. Berkeley.

Recio, A.M. & Godino, J.D. (2001). Institutional and personal meanings of proof. *Educational Studies in Mathematics*, 48(1), 83-99.

Reid, D. & Inglis, M. (2005). Talking about logic. *For the Learning of Mathematics*, 25(2), 24-25.

Schoenfeld, A. (1985). *Mathematical problem solving*. Orlando: Academic Press.

Segal, J. (2000). Learning about mathematical proof: Conviction and validity. *Journal of Mathematical Behavior*, 18, 191-210.

Selden, A. & Selden, J. (2003). Validations of proofs written as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, 36(1), 4-36.

Selden, J. & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics*, 29(2), 123-151.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. London: SAGE.

Stylianides, A. & Stylianides, G. (2008). Enhancing undergraduates students' understanding of proof. In *Proceedings of the 11th Conference for Research in Undergraduate Mathematics Education*. Last downloaded March 18, 2009: http://www.rume.org/crume2008/Stylianides_LONG(21).pdf.

Vinner, S. (1997). The pseudo-conceptual and pseudo-analytic thought processes in mathematical learning. *Educational Studies in Mathematics*, 34, 97-129.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, 48(1), 101-119.

Weber, K. (2004). Traditional instruction in advanced mathematics: A case study of one professor's lectures and proofs in an introductory real analysis course. *Journal of Mathematical Behavior*, 23, 115-133.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, 39, 431-459.

Weber, K. & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics*, 56, 209-234.

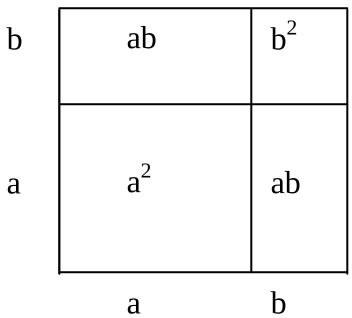
Weber, K. & Alcock, L. (2005). Using warranted implications to understand and validate proofs. *For the Learning of Mathematics*, 25(1), 34-38.

## Appendix- Arguments used in this study

**Argument 1: Valid, deductive algebra**
*Claim:* $(a + b)^2 = a^2 + 2ab + b^2$
*Argument:*
$(a + b)^2 = (a + b)(a + b)$
$(a + b)(a + b) = a(a + b) + b(a + b)$
$a(a + b) = a^2 + ab$
$b(a + b) = ba + b^2$
So $(a + b)(a + b) = a^2 + ab + ba + b^2 = a^2 + 2ab + b^2$

---

**Argument 2: Perceptual, algebra**
*Claim:* $(a + b)^2 = a^2 + 2ab + b^2$
*Argument:*
Consider the diagram below:




The length and width of the square are each (a+b), so the area of the diagram is $(a+b)(a+b) = (a+b)^2$.
The area can also be found by adding the areas of the four cells of the square whose areas are $a^2$, ab, ab, and $b^2$, which is $a^2 + 2ab + b^2$.
So $(a+b)^2 = a^2 + 2ab + b^2$.

---

**Argument 3: Valid, deductive number theory**
*Claim:* For all natural numbers $n$, $n^3 - n$ is divisible by 6.
*Argument.*
$n^3 - n = n(n^2 - 1) = n(n+1)(n-1)$.
Either $n$ is even or $n+1$ is even.
Since both numbers are factors of $n^3 - n$, $n^3 - n$ is even.
Because $n$-1, $n$, and $n$+1 are three consecutive numbers, one of them is divisible by 3.
So $n(n+1)(n-1)=n^3 - n$ is divisible by 3.
Since $n^3 - n$ is even and divisible by 3, $n^3 - n$ is divisible by 6.

---

**Argument 4: Valid, deductive calculus**
*Claim*. The equation, $4x^3 - x^4 = 30$, has no real solutions.
*Argument.* Consider the function, $f(x) = 4x^3 - x^4$. Because $f(x)$ is a polynomial of degree 4 and the coefficient of $x^4$ is negative, $f(x)$ is continuous and will approach $-\infty$ as $x$ approaches $\infty$ or $-\infty$. Hence, $f(x)$ must have a global maximum. The global maximum will be a critical point. $f'(x) = 12x^2 - 4x^3$. If $f'(x) = 0$, then $x = 0$ or $x = 3$. $f(0) = 0$. $f(3) = 27$. Since $f(3)$ is the greatest $y$-value of $f$'s critical points, the global maximum of $f(x) = 27$. Therefore $f(x) \neq 30$ for any real number $x$. $4x^3 - x^4 = 30$ has no real solutions.

---

**Argument 5: Invalid, empirical, prose number theory**
*Claim.* Any even integer greater than two can be written as the sum of two primes.
*Argument*. Consider the following table:

| Even | Sum of two primes |
|---|---|
| 4 | 2+2 |
| 6 | 3+3 |
| 8 | 3+5 |
| 10 | 3+7, 5+5 |
| 12 | 5+7 |

```
14          3+11,  7+7
16          3+13,  5+11
18          5+13,  7+11
20          3+17,  7+13
22          3+19,  5+17,  11+11
24          5+19,  7+17,  11+13
26          3+23,  7+19,  13+13
```

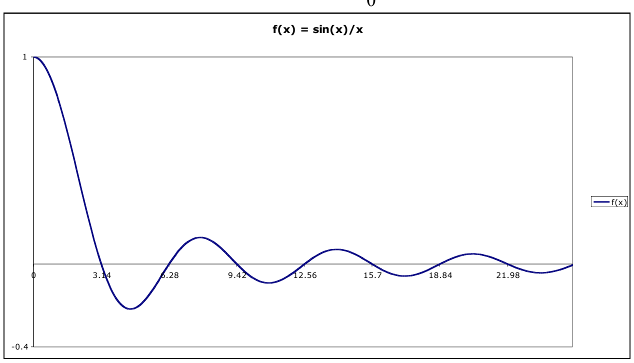
First, note that each even number between 4 and 26 can be written as the sum of two primes. Second, note that the number of pairs of primes that work appears to be increasing. For 4, 6, 8, and 12, there is only one prime pair whose sum is that number. For 22, 24, and 26, there are three prime pairs whose sum is that number. Every even number greater than 2 will have at least one prime pair whose sum is that number. For large even numbers, there will be many prime pairs that satisfy this property.

**Argument 6: Invalid, perceptual calculus**

*Claim.* $\int_0^\infty \frac{1}{x}\sin x\,\mathrm{dx} > 0$

*Argument.* The graph of $f(x) = \frac{1}{x}\sin x$ is given below.

$\int_0^\infty \frac{1}{x}\sin x\,\mathrm{dx} > 0$ means that $f(x) = \frac{1}{x}\sin x$ has more area above the *x*-axis than below it.
To show this, note that it is clear from the graph that the first positive region—between 0 and π (about 3.14)—has more area than the first negative region—between π and 2π (between 3.14 and 6.28). The second positive region has more area than the second negative region. The third positive region has more area than the third negative region. Since each positive region has a greater area than the negative region to the right of it, the overall area of $\int_0^\infty \frac{1}{x}\sin x\,\mathrm{dx}$ will be positive.




**Argument 7: Invalid, deductive number theory (invalid proof structure)**
*Claim:* If $n^2$ is divisible by 3, then $n$ is divisible by 3.
*Argument*.
We need to show that $n$ is divisible by 3.
If $n$ is divisible by 3, then there exists an integer $k$ such that $n = 3k$.
$n^2 = (3k)^2 = 9k^2$.
So $n^2$ is divisible by 9.
All numbers divisible by 9 are also divisible by 3.
So if $n^2$ is divisible by 3, then $n$ is divisible by 3.

**Argument 8: Invalid, deductive calculus (invalid assumption, first line)**
*Claim:* Let *f(x)* be a real valued function, *a* and *b* be real numbers, and *b*> *a*.

$$\int_a^b |f(x)|dx \ge \int_a^b f(x)dx$$

*Argument.* (Proof by cases).
Either $f(x) \geq 0$ or $f(x) < 0$.
Case 1: $f(x) \geq 0$.
If $f(x) \geq 0$, then $|f(x)| = f(x)$.
Thus, $\int_a^b |f(x)| dx = \int_a^b f(x) dx$.
Case 2: $f(x) < 0$.
If $f(x) < 0$, then $\int_a^b f(x) dx \leq 0$.

Since $|f(x)| > 0$, then $\int_a^b |f(x)| dx \geq 0$.

So $\int_a^b |f(x)| dx \geq 0 \geq \int_a^b f(x) dx$.

Thus, $\int_a^b |f(x)| dx \geq \int_a^b f(x) dx$.

---

**Argument 9: Invalid, deductive algebra (invalid warrant, last line)**
*Claim.* Let $f(x) = \ln x$. Then $f(x) \to \infty$ as $x \to \infty$.
*Argument.*
Let $a$ and $b$ be positive real numbers with $a > b$.
Dividing both sides by $b$ gives:
$a/b > 1$ (since $b$ is positive).
$\ln(a/b) > 0$ (since $\ln x > 0$ when $x > 1$)
$\ln(a) - \ln(b) > 0$ (by the rules of logarithms)
$\ln(a) > \ln(b)$
Hence, for positive reals $a$ and $b$, if $a > b$, then $f(a) > f(b)$.
Therefore, $f(x) \to \infty$ as $x \to \infty$.

---

**Argument 10: Invalid, deductive algebra (computation error, line 3)**
*Claim.* For all real numbers $x$, $x^2 + 12x + 28 > 0$.
*Argument:*
$x^2 + 12x + 28 =$
$(x^2 + 12x + 24) + 4 =$
$(x + 6)^2 + 4$
Since $(x + 6)^2$ is a perfect square, $(x + 6)^2 \geq 0$ for all real numbers $x$.
Hence $(x + 6)^2 + 4 \geq 4 > 0$.
So $x^2 + 12x + 28 > 0$.