How do mathematicians determine if an argument is a valid proof?

Keith Weber

Rutgers University

Graduate School of Education

10 Seminary Place

New Brunswick, NJ 08901

U.S.A.

khweber@rci.rutgers.edu

(732) 932-7496 x8123

Mathematicians' validation of proofs

How do mathematicians determine if an argument is a valid proof?

Abstract. The purpose of this paper is to investigate the mathematical practice of proof validation—that is, the act of determining whether an argument constitutes a valid proof. The results of a study with eight mathematicians are reported. The mathematicians were first observed as they read purported mathematical proofs and made judgments about their validity and were then asked reflective interview questions about their validation processes and their views on proving. The results suggest that mathematicians use several different modes of reasoning in proof validation, including formal reasoning and the construction of rigorous proofs, informal deductive reasoning, and example-based reasoning. Conceptual knowledge plays an important role in the validation of proofs. The practice of validating a proof depends upon whether a student or mathematician wrote the proof and in what mathematical domain the proof was situated. Pedagogical and epistemological consequences of these results will be discussed.

Key words: Mathematicians; Mathematical practice; Mathematics education; Number Theory; Proof; Proof validation

How do mathematicians determine if an argument is a valid proof?

1. Introduction

In the last two decades, there has been a tremendous growth in educational research on mathematical proof. Research in this area has examined epistemological issues on proof, including the character of proof (e.g., Downs & Mamona-Downs, 2005), the purposes of proof (e.g., deVilliers, 1990), and the role of proof in the mathematics classroom (e.g., Hanna, 1990; Hersh, 1993). Other research has focused on psychological issues concerning proof, including students' and teachers' conceptions of proof (e.g., Harel & Sowder, 1998; Knuth, 2002; Williams, 1980), students' and mathematicians' construction of proofs (e.g., Weber, 2001; Weber & Alcock, 2004), and students' difficulties with proof (e.g., Moore, 1994; Weber, 2001). Still other research has investigated pedagogical actions associated with proof, including pedagogical techniques for teaching proof (e.g., Alibert & Thomas, 1991; Rowland, 2002) and the teachers' role in proof-related activity (e.g., Herbst, 2002; Blanton, Stylinaou, & David, 2003). Several researchers have noted that one area that has received comparatively little attention is the reading of mathematical proof (e.g., Hazzan & Zazkis, 2003; Selden & Selden, 2003; Mamona-Downs & Downs, 2005). In particular, more research is needed on the practice of and processes involved in determining whether an argument constitutes a valid proof. The purpose of this paper is to address this issue.

Determining whether a purported proof is a valid proof is an important mathematical activity. Selden and Selden (2003) define the act of judging whether an argument constitutes a valid proof as *proof validation*, and argue that the ability to reliably validate proofs is a crucial skill for all involved in proof-oriented mathematics, including mathematicians, mathematics majors, and teachers of mathematics, particularly secondary mathematics teachers who teach courses in which deductive arguments are emphasized (Selden & Selden, 2003). Undergraduate mathematics majors need to validate proofs reliably both to check the validity of the mathematical arguments that they produce and to extract conviction from the proofs that are presented to them in their lectures and textbooks (Selden & Selden, 1995). Secondary teachers need to validate proofs in order to have justification and proof play prominent roles in their mathematics classrooms. In the *Principles and Standards for School Mathematics*, the NCTM (2000) asserts that "with [the teacher's] guidance, students should develop high standards for accepting explanations" and that students should "formulate and critique explanations so their classrooms become communities of inquiry" (p. 346). To help accomplish this, the NCTM recommends that teachers should discuss the logical structure of students' arguments and assist students in critiquing each others' arguments (p. 346). Leading students to develop accurate standards for what constitutes a good argument requires that the teacher have a solid understanding of these standards. At a more practical level, teachers need to be able to determine if the justifications and proofs that students submit are acceptable and to provide feedback when they are not.

Research studies indicate that mathematics majors and teachers of mathematics lack the ability to determine whether an argument constitutes a valid proof. Selden and Selden (2003) presented eight undergraduates with arguments purporting to prove the assertion, "If 3 divides n^2, then 3 divides n" and asked them to determine if each argument would constitute a valid proof. They found that students initially performed at chance level while completing this task, but did show improvement due to the reflective questions engendered in the interview process. Similarly, Alcock and Weber (2005) asked thirteen undergraduate mathematics majors to assess the validity of a flawed proof in real analysis; only six rejected the proof as invalid and only two did so for legitimate mathematical reasons. In his interviews with 16 highly qualified in-service high school teachers, Knuth (2002) reported that these teachers had serious difficulties with the task of proof validation, and many were prepared to accept flawed arguments as valid mathematical proofs. Martin and Harel (1989) also found that pre-service elementary teachers' acceptance of a proof often was based more on the form of the argument presented to them than on the content. In particular, these prospective teachers would accept flawed arguments in a traditional two-column format, but would find valid proofs written in paragraph form to be unacceptable. The results of the studies involving pre-service and in-service teachers have led researchers to question whether these teachers could effectively meet the goals laid out in the *Principles and Standards* (e.g., Knuth, 2000, 2002).

Each of the previously mentioned studies either demonstrated that or illustrated why students and teachers *cannot* validate 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 stress 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 illustrates that most students do not infer and check warrants when validating proofs (Alcock & Weber, 2005). Both Selden and Selden's and Weber and Alcock's analysis is theoretical and not based on systematic observations of those who can validate proofs effectively.

More research is needed on the practice of proof validation. One way to deepen our understanding of this mathematical practice is to observe the behavior of mathematicians who are engaged in the mathematical activity of proof validation (e.g., Stylianou, 2003; RAND Mathematics Study Panel, 2003). In this paper, I will investigate the practice of proof validation by describing the processes that eight mathematicians used to determine if a collection of mathematical arguments constituted valid proofs.

2. Perspectives on proving

Individuals' epistemological beliefs about proof necessarily influence their conceptions of how one should determine whether a given argument constitutes a proof. In the mathematics education communities, there is not a consensus about what is meant by a mathematical proof (e.g., Balacheff, 2002). In this section, I discuss three perspectives on proving and what implication each perspective might have regarding the processes involved in proof validation.

A traditional, formalist perspective on proof holds that "a mathematical proof is a formal and logical line of reasoning that begins with a set of axioms and moves through logical steps to a conclusion" (Griffiths, 2000, p. 2). A literal interpretation of this perspective is that a proof can be viewed as a formal structure whose validity can be assessed by determining if it obeys well-defined and explicitly stated mathematical conventions and logical rules. Lakatos (1978) argues that this position implies that although proof construction may be a creative endeavor that can involve the use of intuitive representations, conceptual knowledge, and heuristics, proof validation is largely a formal activity. When one encounters an assertion in a proof, one needs to judge whether that assertion follows logically from earlier statements in the proof- either by identifying a theorem or logical rule from which the assertion could be derived or perhaps constructing a basic sub-proof formally demonstrating how the new assertion could be deduced from previous ones. Informal reasoning processes, such as drawing diagrams or inspecting particular examples, should play only a minimal role in this process (primarily limited to the construction of sub-proofs) and are not, by themselves, sufficient to determine whether any aspect of a proof is correct. In the last several decades, many philosophers and mathematics educators, as well as some mathematicians, have challenged this formal conception of proof (e.g., Davis & Hersh, 1981; Ernest, 1991; Hanna, 1991, 1995; Hersh, 1997; Thurston, 1994; Tymoczko, 1986), in large part because they believe this is not consistent with the practice of mathematicians (e.g., Lakatos, 1976; Tymoczko, 1986; Hersh, 1997). An alternative interpretation of the traditional view of proof is that an argument constitutes a proof if it is possible to write the argument as a formal proof in axiomatic set theory without changing the substance of the argument (Tymoczko, 1979; Devlin, 1992). How mathematicians judge whether an argument can be rewritten in this form is an open question.

A second perspective is that a proof is an argument that is convincing to a mathematician. This viewpoint can be found in the writings of Davis and Hersh (1981), who argue that a proof is an argument that convinces a mathematician who knows the subject, Volminik (1990) who claims a proof is an argument that would convince a reasonable skeptic, and Mason, Burton, and Stacey (1982), who describe the act of proving as constructing an argument that would convince an enemy. Some philosophers and mathematicians further argue that the rigor and formalism of proof play only a secondary role in evaluating the correctness of the proof. Manin (1977), for instance, contends that acceptance of a proof is aimed more at weighing the plausibility of the argument being presented than at verifying the specific steps of the deductive process. Hanna (1991) argues that when considering the plausibility of an argument, non-mathematical factors, such as the reputation of the proof's author, may influence mathematicians' acceptance of a proof.

A third conception of proof is the view that whether an argument constitutes a proof is a matter of social negotiation and agreement. Balacheff (1987) asserts "we call proof an explanation accepted by a given community at a given time". Similarly, Manin (1977) claims "an argument becomes a proof after the social act of accepting it as a proof". In the following passage, the mathematician William Thurston describes how he developed his understanding of mathematical proof:

> When I started graduate school at Berkeley, I had trouble understanding how I could "prove" a new and interesting theorem. I didn't really understand what a "proof" was. By going to seminars, reading papers, and talking to other graduate students, I gradually began to catch on. Within any field, there are certain theorems and certain techniques that are generally known and generally accepted. When you write a paper, you refer to these without proof […] Many of the things that are generally known are things for which there may be no known written source. As long as people in the field are comfortable that the idea works, it doesn't need to have a formal written source. At first, I was suspicious of this process. I would doubt whether a certain idea was really established. But I found that I could ask people, and they could produce explanations and proofs, or at least refer me to other people or written sources that would give explanations and proofs […] Mathematical knowledge and understanding were embedded in the minds and the social fabric of the community of people thinking about a particular topic. This knowledge was supported by written documents, but the written documents were not really primary (Thurston, 1994).

Thurston's passage implies that when evaluating a proof, one must be aware of the social norms of the mathematical community in which the proof was situated; determining whether a particular technique is permissible without justification in a proof does not involve locating a written document formally establishing the legitimacy of this technique, but determining whether this technique is generally known and accepted by one's community.

Although these three perspectives were presented separately, this is not to imply they are independent of one another or in opposition to each other. On the contrary, one could argue that these perspectives are complementary. For instance, Devlin (1992) argues that, in practice, mathematicians' determining whether an argument can be written as a formal proof in axiomatic set theory and judging if an argument is convincing are synonymous. Other philosophers have claimed an argument needs to satisfy criteria from more than one perspective to be considered a proof. For instance, Tymoczko (1979) claims a proof is an argument that is convincing to a mathematician, can be written in formal, axiomatic set theory, and is surveyable (i.e., understandable by a mathematician).

In this paper, I will present the results of a study in which I asked eight mathematicians to determine whether eight number-theory arguments were valid proofs. These findings will contribute to the mathematics education and philosophical literature on proof in two ways. First, for each of the three perspectives listed in this section, the processes used in proof validation are underspecified. If one contends a proof is an argument that convinces mathematicians, it is then important to understand mathematicians' standards of conviction and the processes that they use to determine if a proof satisfies those standards. Similarly, if one believes that a proof is an argument that can be written in the language of axiomatic set theory, it is natural to ask how mathematicians judge whether an argument can be written in this format. The data reported in this paper will shed light on these important issues. Second, the mathematics education community's philosophy of mathematics has a strong influence on how and what mathematics is taught in our classrooms (Thom, 1971; Ernest, 1991). For instance, the New Math movement that emphasized sets and logic in elementary mathematics in the United States in the 1970's and the current "definition-theorem-proof" format of proof-oriented collegiate mathematics (cf., Weber, 2004) courses are strongly influenced by formalist philosophies of mathematics (Thom, 1971; Tymozko, 1986). There is an increasing recognition that the philosophy of mathematics should be compatible with and influenced by research on the practices of mathematicians (Tymozko, 1986). Thus the results of this study should have implications for philosophies of proof and how the validation of proof should be introduced in mathematics classrooms.

3. Research Methods

Participants. Eight mathematics professors at a regional university in the southern United States participated in this study. All participants were actively involved in mathematical research. Their fields of study included topology, real analysis, algebra, combinatorics, and applied mathematics. The participants all had Ph.D's in mathematics from well-known research universities, published multiple papers in respected journals in their areas of expertise, had experience refereeing papers for mathematics journals, and had three to ten years of experience as faculty members in a mathematics department. It is worth noting that the eight participants in this study shared traits that are not universal among mathematicians. In particular, they all were active in research, working at a regional university, and relatively young. It is possible that the findings from this study may not generalize to mathematicians who lack these characteristics, such as mathematicians who are no longer involved in research or mathematicians who are employed at elite research institutions. On the other hand, the participants in this study conducted their research in a wide variety of mathematical disciplines and had diverse backgrounds in their teaching experiences. Hence, they appeared to be representative in these respects of research-active mathematicians working at regional universities.

The arguments that the mathematicians read were in the domain of number theory. No participant studied number theory as their field of research. However, the participants did include one mathematician who studies combinatorics and two algebraists; number theory played a significant role in their research. Four of the arguments that the mathematicians inspected were purportedly written by students in a transition-to-proof course. Only one of the participants (labeled Mathematician A) had experience teaching a transition-to-proof course; however all participants had experience teaching proof-oriented mathematics courses to advanced undergraduates.

Materials. Two sets of purported proofs were used in this study. The first set of proofs, from hereon labeled "the elementary arguments", consisted of four arguments identical to those used in Selden and Selden's (2003) study on proof validation. These arguments are referred to as arguments a, b, c, and d, and are presented in the appendix of this paper. These four arguments all purported to prove the statement "If n^2 is divisible by 3, then n is divisible by 3". Selden and Selden judged proof b to be a valid proof, while explaining how the other three arguments contained serious logical errors. The Selden and Selden (2003) study documented undergraduates' inability to validate these elementary arguments; this study will illustrate how mathematicians can successfully determine if these elementary arguments are valid proofs.

The second set of proofs, from hereon labeled "the advanced arguments", were four more sophisticated arguments that were of the level of complexity that one might expect to find in an advanced number theory course for undergraduates. These arguments are referred to as arguments e, f, g, and h and are also presented in the appendix. Proofs e and h are valid proofs that were published in the *American Mathematical Monthly* (Holdener, 2002; Watkins, 2000). Proof g is a valid proof that was generated by a colleague of mine[1]. Proof f is an invalid argument that I constructed to use as a foil. These advanced arguments were included to provide the mathematicians with the opportunity to validate proofs that were more sophisticated and to see how mathematicians would behave when it was less obvious whether a particular inference was valid. These were also included to investigate the way that the mathematicians would validate proofs in two different contexts (grading a student-generated proof vs. refereeing a proof submitted by a colleague). Comparisons between mathematicians' behaviors when validating the elementary and advanced arguments are given in section 5.5.

Procedure. Each participant met individually with the author of this paper for a three-stage interview. In the first stage of the interview, the participant was observed as he or she determined if the elementary proofs were valid. The participant was first told that he or she would be presented with the proofs of four students from a transition-to-proof course. The participant was handed each proof, one at a time, and was asked to "think aloud" (using the verbal protocol methodology of Ericsson and Simon, 1993) while determining whether the argument was a valid proof. After making a judgment on the argument, a new argument was given to the participant. This process continued until the participant made a judgment on all four elementary arguments. Participants took between five and twenty minutes in evaluating these first four arguments, with six of the participants completing this stage of the interview in less than ten minutes.

---

[1] I am grateful to Kelly Pearson for this proof.

In the second stage of the interview, each participant was told that they would read arguments that a mathematician might submit to an expository mathematics journal such as the *American Mathematical Monthly* and again were asked to determine if these arguments constituted valid proofs. Each participant was given the advanced arguments one at a time and asked to think aloud while making a judgment on the arguments. This process continued either until arguments e, f, g, and h were presented to the participant or until 45 minutes had elapsed. Five participants only had time to look at arguments e and f and only two participants examined all four advanced arguments.

In the third stage of the interview, each participant was asked a set of general questions about how they would determine if an argument was a valid proof. Some of these questions were adapted from the Selden and Selden (2003) study and some were formed based on interesting observations from pilot studies that I conducted prior to this study. These questions included:

- What do you do when you read a proof?
- When you are trying to determine if a proof is valid, do you check every step of the proof?
- Do you read the proof more than once?
- Do you make small subproofs or expand steps?
- How do you tell when a proof is correct or incorrect?
- Why do we have proofs?
- Would you check a proof by a mathematician in the same way that you would check a proof by a student?
- Do you ever use your intuition to help you verify a proof?

- Do you ever look at examples to help you verify a proof?
- When you read a proof, are you only looking to see if the proof is correct or are you trying to get something more out of it?

During the reflective stage of the interview, the interviewer would also describe aspects of the participant's behavior that he found interesting and ask the participant to reflect upon it.

Analysis. Each interview with the mathematicians was transcribed. Each transcription was carefully read and summarized to give a general description of the participants' behavior during proof validation. The results of this preliminary analysis are presented in section 4.1 and 4.2.

The purpose of the next stage of the analysis was to investigate how mathematicians determined whether a new assertion in the proof was a valid deduction from previous assertions. There were collectively 225 instances in which a participant read an assertion whose validity could reasonably be judged. (Explicit assumptions and statements that introduced variables are not included in this list, since in a proof, the prover can assume or define as he or she pleases). In 122 of these 225 instances, the participant immediately accepted the assertion as valid without explicit comment. In 20 instances, the assertion was rejected as invalid. There were also six cases where the participant was unable to determine if the assertion, and hence the proof itself, was valid. The analysis in this paper concentrates on the reasoning used by the participants to determine whether an assertion was valid in the remaining 77 instances. Of these 77 assertions, 71 appeared in the advanced arguments. Hence, the analysis of the participants' behavior on these assertions is primarily relevant to how they performed on the second stage of the interview—determining whether the advanced arguments submitted by mathematicians for publication in an expository journal are valid.

The participants' comments when examining these 77 assertions were grouped using the constant-comparative method (Strauss & Corbin, 1990). The comments that a participant made to conclude a particular assertion was valid constituted an episode. An initial description was given to summarize each episode. Similar episodes 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. Descriptions and illustrations of each category will be provided in section 4.3. Throughout the analysis, the protocol of the participants while completing the validation tasks was the primary source of data. Participants' comments while answering the interview questions were used as secondary sources of data to corroborate, challenge, or deepen the findings from the analysis described above.

4. Results

4. 1. General results

The participants' judgments on each of the proofs are presented in Table 1.

***** Insert Table 1 About Here *****

For proofs a, d, e, f, g, and h, there was unanimous agreement among the participants who read the proof and felt that they were able to make a judgment. One participant reported that proof b was invalid because of the awkward presentation of its first three lines, while the remaining seven participants judged proof to be valid. Four participants judged proof e to be valid, while the remaining four were unable to determine if a particular assertion within proof e was justified. Three participants chose not to make a judgment on the proof, while the fourth judged the proof as "probably valid", but said that he would need more time to confirm this judgment. The only argument in which there was significant disagreement was argument c. The participants' conflicting judgments on this purported proof will be discussed in section 5.1.

4. 2. Stages of validation

Seven of the eight participants' validations could be divided into two distinct phases. For both the elementary and advanced arguments, when given an argument to validate, these participants would first determine the structure of the argument- that is, the proof technique being employed—primarily by explicitly checking which assumptions were being used in the argument. If the participant found the structure of the proof to be acceptable, he or she would then proceed to checking each line of the argument. (The remaining participant, labeled Mathematician A in this paper, proceeded immediately to a line-by-line check of the argument in all cases). For most of the arguments, the participant would state the proof technique being used aloud (e.g. proof by contradiction, direct proof) before verifying individual statements. To illustrate, when Mathematician E was asked to verify proof b, he stated:

> Math E: [reading the first line of the proof] Suppose to the contrary that n is not a multiple of 3. [Commenting] So I like, so proof by contradiction. [reads next line to himself and comments] So negating the conclusion. It looks good so far.

In this excerpt, the participant noted that the first line of the proof started by negating the desired conclusion and inferred that a proof by contradiction was being employed. He then interpreted the next statement, which appears to define n as $3k + 1$ or $3k + 2$, as a valid negation of the conclusion, and determined that the assumptions in the proof were valid for proofs by contradiction.

Similarly, when Mathematician C read proof a, he stated:

> Math C: Prove, seeing that n^2 is an odd positive integer divisible by 3. I'm looking down to see, are they gonna come back? Yeah, they come back to do the even case, so let's split this up into cases.

The first line of proof a asserts that n^2 is an odd positive integer that is divisible by 3. The statement to be proven is that "For any positive integer n, if n^2 is divisible by 3, then n is divisible by 3". Assuming that n^2 is divisible by 3 implies a direct proof is being used. The assumption that n^2 is odd does not follow directly from the problem statement. When the participant reads this assumption, he infers a proof by cases is being used. Before proceeding with the proof, the participant first makes certain that the other case, the one in which the assumption that n^2 is an even integer divisible by 3 is used, is also present in the proof. Finally, I present one illustration in which a mathematician determines a proof structure is invalid and immediately rejects the proof. The transcript below is Mathematician G reading proof d. This proof is also of the statement, "If n^2 is divisible by 3, then n is divisible by 3". The second line of the proof defines n as $3m$ where m is a positive integer. This would directly imply that n is a multiple of 3, and so would conclude the antecedent of the conditional statement to be proven. When the participant reads this proof, he asserts:

> Math G: Let n be a positive integer such that n^2 is divisible by three… Oh boy! Again, ha, you're assuming, there is something wrong there because you're assuming the conclusion. Yeah, for D you are assuming the conclusion, so obviously you are going to get the correct answer. So this is invalid.

The assumption that the antecedent of the conditional statement is true is not part of any legitimate proof framework for proving conditional statements. Immediately after Mathematician G read a statement based on this assumption, he rejected the proof as invalid. Selden and Selden (2003) note that technically, within a formalist conception of proof, a misplaced assumption does not automatically render a proof invalid. One can make whatever assumptions that he or she chooses in a proof. As long as the proof in d did not rely on this assumption in its final argumentation, the assumption alone would not invalidate the proof. Nonetheless, this participant immediately rejected the proof when he read this assumption. All other participants behaved similarly throughout their interviews.

In discussing their validation processes, the participants described how they began their validation process by first checking the structure of the proof before proceeding to line-by-line verification. Four representative responses to the question, "How do you determine whether or not an argument is a valid proof?" are given below:

> Math B: I will first try to understand the structure of the proof, to get an overview of the argument that's being used. After that, if that's reasonable, I'll check the individual steps to make sure each of them are valid.

> Math E: I first look for a thread of logic. I ask myself, what is this person trying to accomplish? What techniques are they doing? What is the thread of logic running

throughout the proof? After that I look at the statements in the proof. If I don't see how a statement follows, I try to verify the statement is clear and obvious.

Math C: I just read through the proof to just to get the big picture, just to get the feel, to get the flow of it and then go back and get the details […] I would try and glance over it to see the general flow. And then go line by line.

Math F: I ask myself, what is the person trying to show and how are they trying to show it? Then I look at each step in the proof, see if the step follows from the previous ones, and see if the steps put together actually does what the author is trying to do.

4. 3. Line-by-line validation

The remaining analysis focuses on the 77 instances in which a participant was initially unsure as to whether an assertion in a proof was a legitimate consequence of previous assertions, but then provided an explanation for why the assertion did indeed follow. Sixty-seven explanations were first sorted into two categories: construction of a property-based argument and example-based argument. A summary of the frequency of each type of explanation is provided in Table 2 below. Each of these categories is defined and illustrated below. The remaining ten explanations were idiosyncratic and could not be grouped with other explanations to form a category.

***** Insert Table 2 About Here *****

4. 3. 1. Property-based arguments

There were 48 instances in which a participant was initially unsure if an assertion followed validly from previous statements, but resolved this difficulty by reasoning based on properties of the objects, structures, and concepts to which the proof pertained. In 15 cases, the participants *constructed a sub-proof* that logically demonstrated how the new assertion from the proof was a necessary consequence of previous assertions. An argument was considered a sub-proof if it used an acceptable proof technique, carried out the logical details of this technique, and with minor alterations concerning the representation but not the content of the argument, would be considered a valid proof. An example is provided below. The excerpt below illustrates one participant assenting to the second line of proof e by constructing a sub-proof.

Prior assertion: $n \equiv 3 \pmod 4$.

Assertion: Note that *n* is not a perfect square.

Math A: So if you take an odd number and square it, $2k + 1$, and I assume that when you square it out… [the participant writes: $(2k+1)^2=4k^2+4k+1=4(k^2+k)+1$]

that would be, yeah, that would be 1 mod 4. OK, note that *n* is not a perfect square. OK, I think I'm OK with that.

In this instance, the participant chose an appropriate proof technique (showing that the square of odd numbers is always congruent to 1 (mod 4), not 3 (mod 4), by algebraic manipulations), carried out the logical details of this technique, and produced an argument that, with minor refinement, could be used as a formal proof that if $n \equiv 3$ (mod 4), then *n* was not a perfect square. However, the participants did not always show this degree of rigor in their explanations. There were 33 instances in which the participant resolved his or her uncertainty by *constructing an informal justification*, rather than a formal proof, for why the new assertion followed from previous ones. More specifically, a justification was considered to be an informal justification if the following two conditions held: a) The justification was based on properties of the objects, structures, and concepts to which the proof pertained. b) The justification given by the participant would not, by itself, constitute a valid proof, either in the sense that a specific mode of argumentation (e.g., proof by exhaustion, algebraic manipulation) was not specified or substantial details of that argumentation were not carried out. An example from one participants' reading of the fifth line of proof e is provided below.

Prior assertions: $ab = n$. $n \equiv 3 (\text{mod } 4)$.

Assertion: Either $a \equiv 3 (\text{mod } 4)$ and $b \equiv 1 (\text{mod } 4)$ or $a \equiv 1 (\text{mod } 4)$ and $b \equiv 3 (\text{mod } 4)$.

Math A: I guess that's a fact. I mean, how do you take two numbers in $\boldsymbol{Z}_4$ and multiply them together to get three?

In this excerpt, the participant's reasoning expresses the essence of why the assertion is true. His comments also implicitly suggest ways in which a formal proof could be constructed—for example, a proof by exhaustion in which all products of pairs of numbers in $\mathbf{Z}_4$ were computed. However, there is no indication that the participant carried out the details of this argumentation—that is, he did not compute all possible product computations. A second excerpt illustrating an informal justification from proof c is provided below.

Prior assertion: $n \cdot n = 3x$.

Assertion: Thus, $3|n$.

Math B: Hmm, let's see. I suppose they're using something like, if there is a 3 on one side, there's a three on the other. There's a 3 in $3x$ so there must be a 3 in nn. And because 3 is prime, it must be in one of the two. So yeah, there's a 3 in n. 3 divides n. Yeah, I think I agree with that […] My qualm is that they should say

> that, say that they could make this step because 3 is prime or by some theorem, but the logic of the proof, I think that's correct.

Again, in this excerpt, the participant presents an intuitive argument for why this assertion is true, based on the fact that 3 is prime and properties of the prime numbers. His comments also suggest ways that such a statement could be proven, perhaps by looking at the prime decompositions of n^2 and $3x$. However, substantial details of this argument were not carried out. It is noteworthy that, although this participant accepted the argument as valid, he expressed some reservations on whether the argument was sufficiently rigorous. Other participants expressed similar reservations, and this issue is discussed further in section 5.1.

4. 3. 2. Example-based arguments

There were 19 instances in which a participant accepted a general assertion in a proof solely on the examination of examples. These instances did not include cases in which a property-based argument was later produced. The ways in which participants used examples can be parsed into four categories, each of which is described below.

Identification of a systematic pattern. There were four instances in which mathematicians accepted a statement by examining systematically chosen examples, noticing a pattern in these examples, and then conjecturing a general statement that would hold for a wider class of objects based on this pattern. The following example occurred when a participant examined the second line of proof e:

> Previous assertions: $n \equiv 3 \pmod 4$.
>
> Assertion: Note that n is not a perfect square.

> Math H: I'm using examples to see what, where the proof is coming from. So 5^2 is 25 and that's 1 mod 4. 36 is 0 mod 4. 49 is 1 mod 4. 64 is 0 mod 4. I'm thinking that, ah! So it is… 24 times 24, that's 0 mod 4. So a perfect square has to be 1 mod 4, doesn't it? n^2 equals 1 mod 4 or 0 mod 4. Alright.

From his inspection of the integers 5, 6, 7, and 8, this participant conjectured that perfect squares were only congruent to 0 or 1 mod 4, and then tested his conjecture by seeing if this held true for 24. He then concluded that n could not be a perfect square since it was congruent to 3 mod 4. In the interview following the validations, the participant indicated that he would not ordinarily used examples in his validations, but found such reasoning to be appropriate in this particular number-theoretic context.

> I: I noticed that at times you used examples to help you validate the proofs.
>
> Math H: Yes… I think with the proofs with number theory, they [examples] are a little easier to use. You can show it's true for some and then use… induction arguments to show that it's true for all of them. Topology [Math H's area of research] you don't quite have that.

Using a specific example to construct a generic proof. There were seven instances in which the participant first verified that the general assertion held for a specific example. He or she then used this verification to form a generic proof (in the sense of Mason and Pimm (1984)) to establish why the general assertion held. That is, the participant noted that his or her justification for the particular example that was inspected did not depend on any properties of that particular example not shared by all objects to which the statement pertained. This is illustrated with the example below in which a participant examines the third line of proof e:

Previous assertions: $n \equiv 3 \pmod 4$. Note that n is not a perfect square.

Assertion: $\sum_{d|n} n = \sum_{d<\sqrt{n},\, d|n} \left(d + \frac{n}{d}\right)$

Math D: I don't understand this statement so let me look at an example. Let me look at 8. Eight has four factors and they add up to… 15. Only two are less than $\sqrt{8}$. OK so we add 1 and 8 and 2 and 4 and that's 9 plus 6 and, well what do you know? It worked! It worked for this example… Oh I see, each of these numbers multiplied to get 8, one will be less than the square root of 8 and one will be greater… yeah, OK, I see.

This participant initially examined the specific instance where n was equal to eight to try to gain an understanding of the statement. He then saw that the statement was true for eight and formed a generic proof that led him to accept the general assertion as true. It is interesting to note that this participant examined this general assertion with an instance that did not satisfy assumptions that appeared earlier in the proof. It was assumed that n was congruent to 3 (mod 4), yet the professor set n equal to 8.

Failure to find a counter-example. There were six instances in which participants sought to contradict a general assertion by constructing a counter-example. When the search was unsuccessful, they accepted the assertion as valid. It should be noted that the search to find a counter-example never appeared to be random, and seemed to rely on their conceptual understanding of number theory. Consider the example below in which a participant evaluates the fifth line of proof g.

Previous assertions: n is a natural number.

> Assertion: There exists an odd integer m and a non-negative integer l such that $n=2^l m$.
>
> Math E: Hmm… can we express every integer in that way? Well, 1 is $l = 0$, $m = 1$. 2, 4, and 8 are powers, but can we express every integer in that way? What about 3? Um, let $m = 3$ and $l = 0$. And 5, let $m = 5$ and $l= 0$. What about 6? 6 is 2 cubed times 3 [sic]. OK I guess this sounds reasonable.

In his interview after his validations, the participant indicated that his search for counterexamples was partially dictated by his understanding of the integers.

> Math E: In the case that we were looking at, with $n = 2^l m$… yes, since it concerned numbers being written in a way involving powers of 2, the first thing I did was check powers of 2. Then I checked numbers other than the powers of 2, but realized that they were all odd numbers. So I tried even numbers that were not powers of 2 and saw that they worked too.

Validation from a single example. There were two cases in which participants accepted a general assertion as valid by verifying that it held for a single example. In these cases, it is possible that the participant may have initially been more or less sure that the assertion was true and checked a single instance to confirm their initial judgment. In the excerpt below, the participant examines the ninth line of proof g.

> Assertion: If $\frac{n}{2^k}$ is even, $(3^{\frac{n}{2^k}}+1) \equiv 2 (\mathrm{mod}\, 8)$ and contains exactly one factor of two.
>
> Math E: Is that true? 3^2 plus 1 is 10, yeah, that seems fine.

In this excerpt, this participant appeared to accept the claim, if a is even, 3^a+1 is congruent to 2 (mod 8), as valid after examining the single case where a=2.

5. Further considerations

5. 1. The role of context in proof validation

The most interesting disagreements that the mathematicians had occurred when validating proof c, a student-generated five-line proof presented in its entirety below:

> Proof: Let n be an integer such that $n^2=3x$, where x is any integer.
> Then $3|n^2$.
> Since $n^2=3x$, $n \cdot n=3x$.
> Thus, $3|n$.
> Therefore if n^2 is a multiple of 3, then n is a multiple of 3.

Selden and Selden (2003) judged this proof to be invalid because there is a "gap" between the third and fourth lines of the proof. The proof seems to offer no justification for how $3|n$ is a consequence of $n \cdot n=3x$. However, all eight participants noticed that there is an implicit justification that could be used to bridge this gap, namely the theorem that, "if p is prime and $p|ab$, then $p|a$ or $p|b$". All participants struggled in determining whether the proof was valid. My general interpretation of the mathematicians' confusion was that they believed their task was underspecified; they did not know what theorems could be regarded as established in a transition-to-proof course at the time the student wrote the proof. Five participants ultimately decided the proof was not valid but each of these participants expressed some reservation in making this judgment. The transcript of one participant's protocol is presented below:

> Math A; Since $n^2 = 3x$, $nn= 3x$. Thus $3|n$… Well, you know, I don't know about that. I guess it would be silly to give this problem if you've already seen the fact that if p is a prime and divides ab, then p divides a or p divides b, then this problem would be silly anyway. So I'm assuming they haven't seen that fact at

> this point in the course. In which case, the uh… you know, I would, I would actually, I would want, I think this could be salvaged but I would want some justification here. I would ask for some justification along the lines of, you know express *n* as a product of primes or something, or use some fact like if a prime divides a product of integers, then *p* divides one of those integers … I wouldn't say this was the worst, the first proof you gave me was the worst one.
>
> I: So would you say valid, invalid, or you could not draw such a distinction?
>
> Math A: Um, OK, valid invalid, uh. Well, I guess I would say invalid, since there's a leap, an unjustified leap of logic there. I so, I wouldn't consider this to be a proof that would get full credit.

Here, Mathematician A infers that the theorem, "if p is prime and $p|ab$, then $p|a$ or $p|b$" must not have been established by the community of students in the transition-to-proof course since it would be unreasonable to ask a student to prove the statement in question if it was. Therefore, he judged the proof to be invalid. Two other participants argued that the proof would be valid, but again had reservations because they did not feel this proof would be appropriate from students in a transition-to-proof course.

> Math D: Well, I'm thinking of what level do we want to know these problems at? Because certainly it's true that if *a* times *b* is divisible by some prime number, then *a* is divisible by it or *b* is divisible by it. But, I'm kinda, sorta confused this is, these are things in [the transition-to-proof course]? So what level of sophistication are they [the students] at and what do they know to prove it? [Pause]

> I: So you would say you wouldn't be able to determine if the proof was valid if you didn't know…
>
> Math D: Yeah because this is certainly… if *a* times *b* is divisible by three, then either *a* or *b* is divisible by three, so I would, this would be a correct and valid proof, I'm just saying.
>
> I: But it might not be for an introductory [transition-to-proof] student?
>
> Math D: No, because if you haven't proved that theorem yet, then they can't use it. You see what I'm saying? That's why I'm kind of… [long pause]

In this excerpt, Mathematician D recognizes that this proof would be acceptable to the broader mathematical community, but claims it would not be valid for the community of students if they have not yet established the pivotal theorem used implicitly in the proof. Finally, one mathematician claimed he could not make a judgment on the proof because he was not given sufficient information.

> Math F: So 3 divides *n*. Um, yeah… I think I remember something from way back that if we have a ring and a prime divides *ab*, then the prime divides *a* or *b*. Is that right? [pause] Yes, I think it is. I think this step actually follows. OK.
>
> I: So would you say that this is a valid proof then?
>
> Math F: Well, yes, I mean, it could be, I mean, this was handed in by a student in my [transition-to-proof] course, right? So I would need to know, had he seen this theorem before. If he had, then yes, this is a valid proof. If not, I don't know, I guess not, I guess this is not a proof. I don't know, if you gave me this proof, I

> would say it was fine, but if a student gave it to me… I don't know, I guess it would depend on what the student knew at that point. I guess it depends.

Again, in this excerpt, Mathematician F withholds judgment on the proof since he did not have sufficient information on the context in which the proof was written. These excerpts support Thurston's (1994) assertion that the validity of some proofs cannot be determined independently from the context and community in which they were produced.

5. 2. The role of participants' background in proof validation

Other comments during their reflective portion of the interviews suggest that some participants' validation standards and strategies were dependent upon their familiarity of the domain that they were investigating. One representative response is given below:

> I: Would you check a proof in analysis [Mathematician H's field of research] in the same way that you would check a proof in another field?
>
> Math H: No, I know a lot about analysis so I have a good sense about what types of things in a proof are going to work and what are not. I have a big bag of examples and counterexamples. I can read a proof and understand the flow of the argument and have a good intuition about whether that flow is right. If I'm working in a field that I do not understand, like number theory [laughs], I have to go through the argument and check every step logically. Because, since… I don't trust my intuition in these cases.

Mathematician A's response to this question was somewhat different. He indicated that his background would influence whether he would try to determine if the proof was valid:

I: Would you check a proof in algebraic combinatorics [Professor A's field of research] in the same way that you would in a field that you were unfamiliar with?

Math A: You know, no. In algebraic combinatorics, I would probably be capable of following the proof, I would probably work through it a lot more carefully. In a proof in a field that I just wasn't familiar with at all, or that I haven't looked at since graduate school, my inclination would probably be just to believe it. If I got to a point where I didn't understand it, I'd just say, OK, probably, I guess it's true. If it was in print or something like this.

5. 3. Comparisons between mathematicians' validations of elementary and advanced arguments

Participants in this study validated proofs in two different contexts: determining if elementary arguments were purportedly generated by students in a transition-to-proof course and if more advanced arguments that were purportedly written by a mathematician for publication in an academic journal. The processes used by the participants for each of the validation tasks were similar. As described in section 4.2, seven participants engaged in the same two stages of validation for both the elementary and advanced arguments; they first checked to see if the proof had a legitimate structure and then proceeded to a line-by-line verification. For most of the assertions in the elementary proofs, the participants immediately accepted the assertions as valid without justification or rejected them as invalid. This is likely because the assertions in these proofs were basic and the arguments presented to the participants were not complicated. However, in the six cases in which the participants did accept a statement as valid, but did not do so immediately without justification, their reasoning patterns were consistent with their behavior while validating the more advanced assertions. For some assertions, the participants produced a formal argument demonstrating that the assertion in question was a logical consequence of previous statements in the proof. In other cases, the participants produced an informal argument (see section 4.3.1 for an excerpt of how one mathematician justified that "$3|n$" followed from the fact that "$3x = nn$, for some integer x"). In another instance, one participant accepted a statement (line 3 of proof c) because he was unable to find a counterexample to the statement used to support it.

When asked if they would check a proof by a mathematician the same way they would check a proof by a student, six of the eight participants claimed they would not. In general, the participants said they would expect more justification from the student. Three participants claimed students should be given less leeway than mathematicians in their proof-writing. One representative response is provided below:

> Math G: I would say that I would check a proof by a student more rigorously than a proof by a mathematician. Umm, I guess in a certain sense a regular mathematician can assume more than a student does. 'Cause when you have students in a class, you're very narrow in defining what you can use to prove the next problem.

Four participants also indicated that when they read a dubious assertion within a proof, they were more inclined to assume the assertion was wrong if the assertion was written by a student than if the assertion was written by a mathematician. Two representative responses are given below:

> Math A: To be honest, probably not. *There is something about trusting the source.* My assumption would be, when I was reading a proof in a journal or a proof that was handed to me by a mathematician friend and they were pretty sure it was true, my assumption would be that the steps are probably correct and that I need to work hard to make sure that I can understand and justify every step. In a student proof … I'm probably going to be less inclined to work through things that strike me as odd. (Italics are my emphasis)

> Math F: No, I don't think I would. When I see something strange in a student proof, I usually just assume its nonsense, where with a mathematician's proof, I will try to look deeper to understand what he's saying. Maybe that's unfair and maybe it's not, but if I gave every student the same benefit of the doubt that I gave my colleagues, with all the bad proofs they hand in, I would never get done grading.

These excerpts illustrate Hanna's (1991) claim that the authority of the author is a factor when mathematicians validate proofs.

5. 4. Perceptions of the value of proof validation in the reading of proofs

At the end of the interviews, participants were asked, "When you read a proof, are you only looking to see if the proof is correct or are you trying to get something more out of it?". All eight participants emphatically claimed that they were trying to get something more out of a proof beyond ascertaining the proof's validity. In fact, three participants went further, asserting that when they read a proof, they are often less concerned about whether the proof is valid and more interested in the insights behind the proof. Two transcripts illustrating this are presented below:

> I: When you check a proof, are you just trying to see if the proof is correct, or are you hoping to get something more out of it?
>
> Math E: I tend to hope the proof will give me some insight into the problem it was solving. It's not always about correctness. Checking for validity is subordinate really, I'm really looking more to gain some insight.

> Math A: One other bias, I'll just throw this out there. To be honest, when I read papers, I don't read the proofs. In the journal papers, and the papers that I read in my research, maybe that's bad or maybe that's not, if I'm convinced that the result is true, I don't necessarily need to read it, I can just believe it.
>
> I: When you check a proof, are you just trying to see if the proof is correct, or are you trying to get something more out of it?
>
> Math A: I think I hope to get something more out of it. Checking just for correctness, I guess what I said was it's not that important to me. If it offers me some insight, that's often what I'm looking for if I take the time to read the proof. Yeah, any of these proofs here that I was reading, they were, I was interested in them, even just being reminded of some things I may have forgotten.

These results corroborate the arguments of Hanna (1995), Thurston (1994), and others who emphasize that the primary purpose of sharing proofs is not only to formally establish that a theorem is correct, but also to communicate mathematical ideas.

6. Discussion

6. 1. Limitations of this study

The purpose of this study was to investigate the processes and reasoning used by mathematicians in proof validation. It is certainly possible that there are important processes used for validation that were not observed. For example, some of the reasoning used by the mathematicians may have been sub-conscious and not observable by verbal protocol analysis. Such reasoning is necessarily beyond the scope of the methodology used in this study. Further, the participants in this study spent no more than twenty minutes in validating a single proof. In the actual practice of mathematicians, the act of validating a proof may span over several weeks or longer. If mathematicians were given a longer period of time to inspect these proofs, different modes of reasoning may have been exhibited. Further still, several mathematicians indicated that they would use different strategies for judging the validity of number theoretic proofs than they would to validate proofs from other mathematical domains. Finally, the mathematicians' proof validations were investigated in two settings—grading a proof that a student submitted and validating a proof submitted by a mathematician for publication in an expository journal. There are other situations in which a mathematician might read a proof, such as refereeing an article for a research journal or studying a published proof of an important theorem, that might involve different validation processes than those exhibited by the participants in this study. In summary, there may be many aspects of proof validation that were not observed in this study. Nonetheless, it still seems reasonable to say that this study can be used to identify *some* of the processes that mathematicians use to validate proofs.

6. 2. Summary of results

When validating proofs, most participants' validations could be divided into two distinct phases. First, the participants checked to see if the argument had a legitimate structure, primarily by explicitly checking whether the assumptions used in the argument were part of a permissible proof framework. Next, the participants would proceed to checking individual lines of the argument to determine whether they followed validly from the proof's assumptions or other previously deduced statements in the proof.

Participants used a wide repertoire of reasoning strategies to determine whether one line in proof followed legitimately from previous statements. In some case, formal logic was used to construct a sub-proof that bridged previous statements and the new assertion, but other strategies were also used frequently by the participants. There were cases in which the participants judged an assertion to follow validly based on informal arguments or the careful inspection of examples. It should be noted that the informal or inductive reasoning that mathematicians used were not applied haphazardly. From their comments in the interview following their validations, the participants indicated that their conceptual understanding of number theory helped them decide when their inductive strategies were appropriate and how they should be applied. These findings illustrate that proof validation is not a purely formal process, but is dependent upon the validator's informal understanding of the domain in which the proof is couched.

6. 3. Why did the participants justify assertions using inductive reasoning?

One particularly interesting finding from these data was that the justifications that participants used to convince themselves of the legitimacy of a particular assertion in a proof employed modes of argumentation that would not be permissible in the presentation of a formal proof. For example, mathematicians would not judge an open theorem to be true simply because they were unable to find a counterexample to this theorem, yet the participants in this study would sometimes accept particular assertions within a proof to be true for this reason. An interesting question is why, and in what sense, did the participants gain conviction from their inspection of examples. Worded differently, why were participants willing to accept that some assertions were true without a deductive argument? This question cannot be addressed by the data in this paper, but would be an interesting topic for future research with implications for the philosophy of mathematics. I suggest two hypotheses below.

First, the participants may have seen how their inductive reasoning could be mapped or "translated" into a deductive argument. Based on her interviews with mathematicians, Raman (2002; 2003) argues that mathematicians see strong links between their informal reasoning and their rigorous proofs. Further, she claims that some mathematicians will be fully convinced of an assertion if they construct an informal argument in support of that assertion and see a method for mapping this argument into a formal proof. These mathematicians do not need to actually carry out this translation to obtain complete conviction (Raman, 2002). This hypothesis is consistent with the position that mathematicians will accept arguments as valid if they see how the argument can be expressed as a formal, axiomatic proof even if the argument is not explicitly expressed in this way. It seems plausible that the participants in this study were convinced by their generic proofs because they could see how these example-based arguments could be mapped to general, deductive proofs. However, it is more difficult to describe other inductive reasoning patterns, such as accepting a statement based on a pattern obtained from the systematic inspection of examples, in such a framework. Alternatively, it is possible that the participants were not trying to obtain certainty that particular deductions were valid, but only believe these deductions were valid with a high degree of confidence. Inglis, Meija-Ramos, and Simpson (in press) observed mathematicians as they attempted to determine whether particular statements from number theory were true. The mathematicians in these studies did not use examples to conclusively determine that a general statement was true, but would use examples to obtain a high degree of confidence in the statement's veracity. It is plausible that in some episodes, the participants were doing the same, particularly in the two episodes in which the participants accepted a statement as true based on the inspection of a single example. In these cases, believing that the assertion in the proof was valid with a high probability was sufficient for them to accept the statement and continue reading. Such actions would be consistent with Manin's (1977) contention that mathematicians' acceptance of a proof is sometimes more dependent upon how plausible the argument seems to be rather than the details of the deductive process in that argument. In this study, I did not ask the participants if or why they were certain that the assertions in the arguments they read were valid deductions, so these issues cannot be addressed by the data presented in this paper.

6. 4. Educational implications

Investigations into the practices of professional mathematicians should have a strong influence on what is taught in mathematics classrooms (e.g., RAND, 2003); however the link between the behaviors of mathematicians and the teaching of mathematics is not straightforward. One example of how the practice of mathematicians influenced the teaching of mathematics is with the case of problem-solving heuristics. In Polya's writings on problem solving practice (e.g., Polya, 1957), he emphasized the importance of heuristics in mathematical problem solving. These insights led mathematics educators to realize that one must not only teach facts, procedures, and concepts to students, but must also help students develop heuristics and strategies for using this knowledge effectively when they are doing mathematics. The teaching of heuristics was recognized as a central goal of mathematics education (NCTM, 1980). However, simply making students aware of the heuristics that mathematicians used or giving students practice applying these heuristics proved to be ineffective instruction (e.g., Begle, 1979; Schoenfeld, 1985); it was only after a deep mathematical analysis of the heuristics being taught and the careful design of instruction that the teaching of heuristics led to substantial learning gains (Schoenfeld, 1985).

The educational implications of this study concern the finding that the processes involved in proof validation are highly dependent upon contextual factors, including the mathematical domain in which the proof is situated, the community that is evaluating the proof, and the author of the proof. Regarding the first point, the concept of proof is often introduced to students in introductory proof courses that strive to be "content-free" (Marty, 1986). In these courses, the coverage of mathematical content is deliberately shallow, primarily designed to be only deep enough to give students a setting in which they can practice basic proof techniques. The data presented in this paper illustrate that many of the validation strategies used by mathematicians are based upon their conceptual understanding of the domain being studied. While there may be basic validation strategies that can be taught in introductory proof courses, such as identifying if an argument is using a valid proof framework (Selden & Selden, 1995), some validation strategies are content-dependent and, it would seem, must be learned in the context of studying new mathematical domains.

As illustrated in section 5.1, the participants in this study viewed the decision as to whether a particular argument is a proof as dependent upon the community that produced and is evaluating the proof. In particular, the participants made a distinction between the *content* of a theorem used in a proof (i.e., was the theorem a true mathematical assertion?) and the *status* of the theorem (i.e., did the community in question have a legitimate reason to view the theorem as established?) (cf., Harel, 2006). In deciding whether an argument constitutes a proof, a student or teacher should not only assess whether the proof techniques employed were acceptable to the mathematical community at large, but also whether they were understood by the community in which the proof was situated (e.g., Stylianides, 2007). If students were aware of the need for making the latter distinction, they may come to appreciate the social functions of proof in helping mathematical communities understand why certain theorem are true. Of course, evaluating whether or not a community has accepted a particular theorem or proof technique for legitimate mathematical reasons can be a difficult decision; more research is needed on what types of learning environments may foster students' appreciation for and ability to make such judgments.

One final note concerns how the authority of an arguments' author should affect students' judgment of whether that argument constitutes a proof. Many in mathematics education believe it should not. For instance, Selden and Selden (2003) claim, "like midcentury structural critics, mathematicians seem to treat a proof as being independent from its author" (p.6) and Harel and Sowder (1998) argue that basing one's decision about a proof solely on the authority of its author is an undesirable yet common way of reasoning. However, most of the participants in this study claimed that they would evaluate proofs differently depending upon the authority of the author. In particular, some said they would spend more times working through the details of a seemingly dubious argument if they trusted the author of the proof. We might want students to do the same. If a student was handed a proof from a respectable source, such as a teacher or a textbook, and the student thought the proof was invalid, it might be beneficial for the student to act on the assumption the proof was valid and to work hard at trying to resolve his or her difficulties with the proof. However, if the proof was generated by the student himself or herself, we might want the student to simply correct the argument. In short, while we would obviously want students to accept or reject arguments upon their own merits, authority still may play a role in their decisions.

Acknowledgements

Thanks to Matthew Inglis, who patiently read several earlier drafts of this manuscript and offered many insightful suggestions, and to Brian Mitchell, who assisted with early stages of this project. Thanks also to Lara Alcock, Carolyn Maher, Eva Thanheiser, Steven Williams, and the anonymous reviewers for helpful comments on earlier drafts of this manuscript. Finally, I am indebted to the mathematicians who generously offered their time to participate in my study.

References

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

Alibert, D. and Thomas, M. (1991). Research on mathematical proof. In D. Tall (Ed.) *Advanced Mathematical Thinking*. Dordrecht: Kluwer.

Balacheff, N. (1987). Processus de preuves et situations de validation. *Educational Studies in Mathematics*, 18(2), 147-176.

Balacheff, N. (2002). The researcher epistemology: A deadlock from educational research on proof. In F.L. Lin (ed.) *2002 International Conference on Mathematics Education- Understanding proving and proving to understand*. Taipei: NCS and NUST.

Begle, E.G. (1979). *Critical variables in mathematics education: Findings from a survey of the empirical literature*. Washington, D.C.: MAA and NCTM.

Blanton, M. L., Stylianou, D. A. & David, M. M. (2003). The nature of scaffolding in undergraduate students' transition to mathematical proof. In N. Pateman, B. Dougherty, & J. Zilliox (Eds.) *Proceedings of the 27th Annual Meeting for the International Group for the Psychology of Mathematics Education,* 2, 113-120, Honolulu, Hawaii.

Davis, P.J. and Hersh, R. (1981). *The mathematical experience*. New York: Viking Penguin Inc.

Devlin, K. (1992). The nature of proof. *Notices of the American Mathematical Society*, 39(9), 1065.

de Villiers, M. D. (1990). The role and function of proof in mathematics. *Pythagoras*, 24, 17-24.

Downs, M. L. N. & Mamona-Downs, J. (2005). The proof language as a regulator of rigor in proof, and its effect on student behavior. In *Proceedings of the 4th Conference on European Research in Mathematics Education*. Sant Feliu de Guixols, Spain.

Ericsson, K.A. & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Cambridge, MA: Bradford Books/MIT Press.

Ernest, P. (1991) *The Philosophy of Mathematics Education*, London: Falmer Press.

Griffiths, P. A. (2000). Mathematics at the turn of the millennium. *American Mathematical Monthly*, 107(1), 1-14.

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange*, 21(1), 6-13.

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

Hanna, G. (1995). Challenges to the importance of proof. *For the Learning of Mathematics*, 15(3), 42-49.

Harel, G. (2006). Mathematics education research, its nature, and its purpose: A reply to Lester's paper. *Zentralblatt fuer Didaktik der Mathematik*, 38, 58-62.

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

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.

Herbst, P. (2002). Engaging students in proving: A double bind on the teacher. *Journal for Research in Mathematics Education*, 33(3), 176-203.

Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics*, 24(4), 389-399.

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

Holdener, J. (2002). A theorem of Touchard on the form of odd perfect numbers. *American Mathematical Monthly*, 109, 156.

Inglis, M., Meija-Ramos, J.P., & Simpson, A. (in press). Modeling mathematical argumentation: The importance of qualification. To appear in *Educational Studies in Mathematics.*

Knuth, E. (2000). The rebirth of proof in school mathematics in the United States. *International Newsletter on the Teaching and Learning of Mathematical Proof*. (Spring 2000). Available from the following website: http://www.lettredelapreuve.it/Newsletter/000506Theme/000506ThemeUK.html. Last downloaded: July 12, 2005.

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

Lakatos, I. (1978). *Mathematics, science, and epistemology*. Cambridge: Cambridge University Press.

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

Manin, Y. (1977). *A course in mathematical logic*. New York: Springer Verlag.

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

Marty, R.H. (1986). Teaching proof techniques. *Mathematics in college* (Spring-summer), 46-53. The City University of New York.

Mason, J. & Pimm, D. (1984). Generic examples: Seeing the generic in the particular. *Educational Studies in Mathematics*, 15, 277-289.

Mason, J., Burton, L., & Stacey, K. (1982). *Thinking Mathematically*. London: Addison Wesley.

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

National Council of Teachers of Mathematics (1980). *An agenda for action: Recommendation for school mathematics in the 1980's*. Washington, D.C.: NCTM.

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

Polya, G. (1957). *How to solve it (2nd Ed.)*. New York: Doubleday Anchor Press.

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.

Raman, M. (2003). Key ideas: What are they and how can they help us understand how people view proof? *Educational Studies in Mathematics*, 52, 319-325.

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

Rowland, T. (2002). Generic proofs in number theory. In S. Campell & R. Zazkis (Eds.). *Learning and teaching number theory: Research in cognition and instructions* (pp. 157-184). Westport, CT: Ablex Publishing.

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

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.J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, 38, 289-321.

Stylianou, D. A. (2002). Interaction of visualization and analysis - The negotiation of a visual representation in problem solving. *Journal of Mathematical Behavior*, 21(3), 303-317.

Thom, R. (1971). "Modern" mathematics: An educational and philosophical error? *American Scientist*, 59, 695-699.

Thurston, W.P. (1994). On proof and progress in mathematics, *Bulletin of the American Mathematical Society,* 30, 161-177.

Tymoczko, T. (1979). The four color problem and its philosophical significance. *Journal of Philosophy*, 76, 57-83.

Tymoczko, T. (Ed.) (1986). *New directions in the philosophy of mathematics*. Secaucus, NJ: Birkhauser.

Volminik, J.D. (1990). The nature and role of proof in mathematics education. *Pythagoras,* 23, 7-10.

Watkins, W. (2000). The quadratic residues -1 and -3. *American Mathematical Monthly*. 107, 934.

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 courses: 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. & 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.

Williams, E. (1980). An investigation of senior high school students' understanding of the nature of mathematical proof. *Journal for Research in Mathematics Education,* 11(3), 165-166.

<u>Captions for tables.</u>

Table 1. Participants' judgments on each of the proofs.

Table 2. Types of arguments for accepting assertions.

Table 1. Participants' judgments on each of the proofs

Proof Attempt	Valid Proof	Invalid Proof	Could not Decide	Did not Attempt	Other
a.	0	8	0	0	0
b.	7	1	0	0	0
c.	2	5	0	0	1
d.	0	8	0	0	0
e.	4	0	3	0	1
f.	0	8	0	0	0
g.	2	0	1	5	0
h.	2	0	0	6	0

Table 2. Types of arguments for accepting assertions

Type of justification	Number (Total: N = 77)
Property-based argument (N=48)	
Sub-proof	15
Informal argument	33
Example-based argument (N = 19)	
Identification of systematic pattern from examining several examples	4
Understanding of why an individual example worked	7
Failure to find counter-examples	6
Validation from a single example	2
Other	10

Appendix- Arguments used in this study
(Note: the lines of the proof are numbered in this appendix to improve the readability of the text of the manuscript. No line numbers appeared on the arguments handed to the participants).

Theorem: For any positive integer n, if n^2 is divisible by 3, then n is divisible by 3. (This theorem was used for proofs a-d).

(a).

1. Proof: Assume that n^2 is an odd positive integer that is divisible by 3.
2. That is, $n^2=(3n+1)^2=9n^2+6n+1=3n(n+2)+1$.
3. Therefore, n^2 is divisible by 3.
4. Assume that n^2 is even and a multiple of 3.
5. That is $n^2 = (3n)^2=9n^2=(3n)(3n)$.
6. Therefore, n^2 is a multiple of 3.
7. If we factor $n^2=9n^2$, we get $3n(3n)$; which means that n is a multiple of 3.

(b).

1. Suppose to the contrary that n is not a multiple of 3.
2. We will let $3k$ be a positive integer that is a multiple of 3,
3. So that $3k+1$ and $3k+2$ are integers that are not multiples of 3.
4. Now $n^2=(3k+1)^2=9k^2+6k+1=3(3k^2+2k)+1$.
5. Since $3(3k^2+2k)$ is a multiple of 3, $3(3k^2+2k)+1$ is not.
6. Now we do the other possibility, $3k+2$.
7. $n^2=(3k+2)^2=9k^2+12k+4=3(3k^2+4k+1)+1$ is not a multiple of 3.
8. Because n^2 is not a multiple of 3, we have a contradiction.

(c).

1. Proof: Let n be an integer such that $n^2=3x$, where x is any integer.
2. Then $3|n^2$.
3. Since $n^2=3x$, $n\cdot n=3x$.
4. Thus, $3|n$.
5. Therefore if n^2 is a multiple of 3, then n is a multiple of 3.

(d).

1. Proof: Let n be a positive integer such that n^2 is a multiple of 3.
2. Then $n=3m$ where $m \in \mathbf{Z}^+$.
3. So $n^2=(3m)^2=9m^2=3(3m^2)$.
4. This breaks down into $3m$ times $3m$ which shows that n is a multiple of 3.

(e).
Theorem: If n is an odd perfect number, then $n \equiv 1 \pmod{12}$ or $n \equiv 9 \pmod{12}$.
Note: n is perfect if and only if $\sigma(n) = \sum_{d|n} d = 2n$.

1. Proof: We first show that if n is a perfect number, n is not congruent to 3(mod 4).
2. Suppose $n \equiv 3 \pmod 4$. Note that n is not a perfect square.
3. $\sigma(n) = \sum_{d|n} d = \sum_{d|n, d<\sqrt{n}} (d + \frac{n}{d})$
4. Note that since d and $\frac{n}{d}$ are divisors of n, $d(\frac{n}{d}) = n$, and $n \equiv 3 \pmod 4$,
5. then either $d \equiv 1 \pmod 4$ and $\frac{n}{d} \equiv 3 \pmod 4$ or $d \equiv 3 \pmod 4$ and $\frac{n}{d} \equiv 1 \pmod 4$.
6. In either case, $4|(d+\frac{n}{d})$ for all d dividing n.
7. Hence, $4|\sigma(n)$ but 4 does not divide $2n$ which is congruent to 2(mod 4).
8. So n is not a perfect number.
9. We now show that if n is a perfect number, n is not congruent to 5(mod 6).
10. Suppose $n \equiv 5 \pmod 6$. Note that n is not a perfect square.
11. $\sigma(n) = \sum_{d|n} d = \sum_{d|n, d<\sqrt{n}} (d + \frac{n}{d})$
12. Note that since d and $\frac{n}{d}$ are divisors of n, $d(\frac{n}{d}) = n$, and $n \equiv 5 \pmod 6$,
13. then either $d \equiv 1 \pmod 6$ and $\frac{n}{d} \equiv 5 \pmod 6$ or $d \equiv 5 \pmod 6$ and $\frac{n}{d} \equiv 1 \pmod 6$.
14. In either case, $6|(d+\frac{n}{d})$ for all d dividing n.
15. Hence, $6|\sigma(n)$ but 6 does not divide $2n$ which is congruent to 4(mod 6).
16. So n is not a perfect number.
17. Thus, if n is a perfect number, n is not congruent to 3(mod 4) or 5(mod 6).
18. If n is odd, but n not congruent to 3(mod 4) or 5(mod 6), then either
19. $n \equiv 1 \pmod{12}$ or $n \equiv 9 \pmod{12}$.
20. So if n is an odd perfect number, then $n \equiv 1 \pmod{12}$ or $n \equiv 9 \pmod{12}$.

Proof adapted from Holdener (2002).

(f).
Theorem: There are infinitely many primes that can be written as $4k+1$ (where $k\in\mathbf{Z}$).

1. Proof. Suppose there are finitely many primes of the form $4k+1$.
2. Then these primes can be enumerated $p_1, p_2, p_3, \ldots, p_n$.
3. Define a number a as follows. Let $a=\left(\prod_{i=1}^{i=n} p_i\right)+4$.
4. Note that a is congruent to 1(mod 4).
5. Note that every number congruent to 1(mod 4) is divisible by a prime congruent to 1(mod 4).
6. However, for all i such that $1\leq i\leq n$, $p_i \mid \prod_{i=1}^{i=n} p_i$ and p_i does not divide 4.
7. Thus, p_i does not divide a.
8. So a is congruent to 1(mod 4) and no prime congruent to 1 mod 4 divides a.
9. This is a contradiction.

(g).
Theorem. $2^n|(3^n\text{-}1)$ if and only if n=1,2, or 4.

1. Proof. The "if" part is obvious. To prove the "only if" part, we note that
2. $3^n - 1 = 2\sum_{i=0}^{n-1} 3^i$ (since $\sum_{i=0}^{n-1} p^i = \frac{p^n - 1}{p-1}$).
3. Suppose $2^n|(3^n\text{-}1)$.
4. Let n be given.
5. We can express n as $n = 2^l m$ (where l and m are integers, and m is positive and odd).
6. Then, 3^n-1
   $= (3^{n/2} + 1)(3^{n/2} - 1)$
7. $= (3^{n/2} + 1)(3^{n/4} + 1)(3^{n/4} - 1)$
   ...
   $= (3^{n/2} + 1)(3^{n/4} + 1)...(3^{n/2^l} + 1)(3^{n/2^l} - 1)$
8. $= (3^{n/2} + 1)(3^{n/4} + 1)...(3^{n/2^l} + 1)(3^m - 1)$
   $= (3^{n/2} + 1)(3^{n/4} + 1)...(3^{n/2^l} + 1)(2\sum_{i=0}^{m-1} 3^i)$
9. If $\frac{n}{2^k}$ is even, $\left(3^{n/2^k} + 1\right) \equiv 2 \pmod 8$ and thus contains exactly one factor of two.
10. If $\frac{n}{2^k}$ is odd, $\left(3^{n/2^k} + 1\right) \equiv 4 \pmod 8$ and thus contains exactly two factors of two.
11. $\sum_{i=0}^{m-1} 3^i$ is the sum of an odd number of odd addends and thus is odd.
12. Counting the factors of two in the expression above yields that 3^n-1 contains at most ($2l$+1) factors of 2.
13. Since $2^n|(3^n - 1)$, 3^n-1 must contain at least $n=2^l m$ factors of 2.
14. So $2^l m \leq 2l+1$.
15. If l=0, $m\leq 1$ so m=1 and thus n=1.
16. If l=1, $2m\leq 3$ so m=1 and thus n=2.
17. If l=2, $4m\leq 5$ so m=1 and thus n=4.
18. If l=3, $2^l m \leq 2l+1$ will be impossible since $2^l > 2l+1$.
19. So $2^n|(3^n\text{-}1)$ only if n=1, 2, or 4.

Proof provided by Kelly Pearson.

(h). Theorem: If p is an odd prime and $p \equiv 1 \pmod 6$, then –3 is a quadratic residue(mod p).
Note: -3 is a quadratic residue(mod p) if $\exists k \in \boldsymbol{Z}_p (k^2 = -3)$.

1. Proof: Suppose $p \equiv 1 \pmod 6$.
2. Assume that –3 is not a square (quadratic residue) in the field $\boldsymbol{Z}_p$.
3. Extend $\boldsymbol{Z}_p$ to $\boldsymbol{Z}_p(\sqrt{-3})$.
4. $\boldsymbol{Z}_p(\sqrt{-3})$ has p^2 elements (since $\sqrt{-3}$ is not a member of $\boldsymbol{Z}_p$).
5. $\boldsymbol{Z}_p(\sqrt{-3})^*$ (that is, the multiplicative group formed with the non-zero elements of $\boldsymbol{Z}_p(\sqrt{3})$) has $p^2 - 1$ elements.
6. Now consider the factor group $\boldsymbol{Z}_p(\sqrt{-3})^*/\boldsymbol{Z}_p^*$. This group has order $\frac{p^2-1}{p-1} = p+1$.
7. Consider the coset represented by $3+\sqrt{-3}$ in the factor group $\boldsymbol{Z}_p(\sqrt{-3})^*/\boldsymbol{Z}_p^*$.
8. Direct computation (given below) shows $3+\sqrt{-3}$ has order 6 in this factor group.
9. Hence, by Lagrange's Theorem, $6|(p+1)$, so $p \equiv -1 \pmod 6$, contradicting our assumption.

10. Computations (you may assume these computations are correct):

$$\left(3+\sqrt{-3}\right)^2 = 6+6\sqrt{-3}$$

$$\left(3+\sqrt{-3}\right)^3 = 24\sqrt{-3}$$

$$\left(3+\sqrt{-3}\right)^4 = -72+72\sqrt{-3}$$

$$\left(3+\sqrt{-3}\right)^5 = -432+144\sqrt{-3}$$

$$\left(3+\sqrt{-3}\right)^6 = -1728.$$

Note that no prime $p \equiv 1 \pmod 6$ divides 6, 24, 72, 144. Hence $(3+\sqrt{-3})$ has order 6.

Proof adapted from Watkins (2000).