DEPARTMENT OF MATHEMATICS

TECHNICAL REPORT

VALIDATION OF PROOFS AS A TYPE OF READING AND SENSE-MAKING: ILLUSTRATED BY AN EMPIRICAL STUDY[1]

DR. ANNIE SELDEN[2] and DR. JOHN SELDEN

SPRING 2015

No. 2015-4

TENNESSEE TECHNOLOGICAL UNIVERSITY
Cookeville, TN 38505

VALIDATION OF PROOFS AS A TYPE OF READING AND SENSE-MAKING: ILLUSTRATED BY AN EMPIRICAL STUDY[1]

Annie Selden[2] and John Selden

In this paper, we consider proof validation as a type of reading and sense-making of texts within the genre of proof. To illustrate these ideas, we present the results of a study of the proof validation abilities and behaviors of sixteen U.S. undergraduates after taking an inquiry-based transition-to-proof course that emphasized proof construction and demonstrated proof validation. Participants were interviewed individually towards the end of the course using the same protocol that was used by Selden and Selden (2003) at the beginning of a transition-to-proof course. Results include a description of the participants' observed validation behaviors, a description of their proffered evaluative comments, a description of their sense-making attempts, and the, perhaps counterintuitive, suggestion that taking an inquiry-based transition-to-proof course emphasizing proof construction may not enhance students' abilities to judge the correctness of other students' proof attempts. We also discuss distinctions between proof validation, proof comprehension, proof construction and proof evaluation and the need for research on their interrelations.

Key words: Validation, proof, reading, sense-making, transition-to-proof

Introduction

In this paper, we consider proof validation as a type of reading and sense-making of texts within the genre of proof (Authors, 2013). Further, in line with reading comprehension researchers, we view reading as an active process of meaning-making in which readers use their knowledge of language and of the world, including the mathematical world, to construct situation models of texts in light of their backgrounds and experience (e.g., Kintsch, 2004; Pressley & Afflerbach, 1995; Zwaan & Radvansky, 1998).

To illustrate the above ideas for proof validation, which we view here as a kind of reading, we present the results of a study of the observed proof validation behaviors of 16 U.S. undergraduates after taking an inquiry-based transition-to-proof course emphasizing proof construction. Students were interviewed individually towards the end of the course employing the same protocol as that used in the Selden & Selden (2003) study in which proofs were regarded as texts that establish the truth of theorems. They described *proof validation* as the reading of, and reflection on, proofs to determine their correctness. More explicitly, they said

> A validation is often much longer and more complex than the written proof and may be difficult to observe because not all of it is conscious. Moreover, even its conscious part may be conducted silently using inner speech and vision. Validation can include asking and answering questions, assenting to claims, constructing subproofs, remembering or finding and interpreting other theorems and definitions, complying with instructions (e.g., to consider or name something), and conscious (but probably nonverbal) feelings of rightness or wrongness. Proof validation can also include the production of a new text—a validator-constructed modification of the written argument—that might include additional calculations, expansions of definitions, or constructions of subproofs. Towards the end of a validation, in an effort to capture the essence of the argument in a single train-of-thought, contractions of the argument might be undertaken. (Selden & Selden, p. 5).

Below we provide detailed descriptions of the observed validation behaviors that the 16 U.S. undergraduates took – something either not done, or only partially done, in prior validation studies and perhaps not at all for this particular level of student. Past validation studies include: first-year Irish undergraduates' validations and evaluations (Pfeiffer, 2011); U.S. undergraduates' validations at the beginning of a transition-to-proof course (Selden & Selden, 2003); U.S. mathematics majors' validation practices across several content domains (Ko & Knuth, 2013); U.S. mathematicians' validations (Weber,

[1] This is an extension of a paper presented at the 17th Annual Conference on Research in Undergraduate Mathematics Education, Denver, Colorado, February 2014 and published in its *Proceedings*, available at http://sigmaa.maa.org/rume/RUME17.pdf.

[2] Annie Selden is Professor Emerita of Mathematics from Tennessee Technological University. John Selden is retired from and also formerly taught mathematics at Tennessee Technological University. The Seldens have retired to Las Cruces, New Mexico and are now Adjunct Professors of Mathematics at New Mexico State University.

2008); and U.K. novices' and experts' reading of proofs, using eye-tracking, to compare their validation behaviors (Inglis & Alcock, 2012).

Our ultimate goal is to understand both the proof construction process and the proof validation process. Our specific empirical research questions were: Would taking an inquiry-based transition-to-proof course that emphasized proof construction significantly enhance students' proof validation abilities? And, what behaviors do students exhibit when reading other students' proof attempts, when asked to judge their correctness? In answering these questions, we came to see the students' validation attempts as a kind of reading and sense-making – a practice that goes beyond determining correctness.

Theoretical Perspective

In a way similar to Selden, McKee, & Selden (2010), we view the proof construction process as a sequence of mental or physical actions in response to situations in the partly completed proof. This process, even when accomplished with few errors or redundancies, contains many more actions, or steps, than appear in the final written proof and cannot be fully reconstructed from a final written proof. For example, actions, such as "unpacking" the conclusion to see what one is being asked to prove, and drawing a diagram, may not appear in the final written proof, and hence, are often unavailable to students for consideration and reflection.

Many proving actions appear to be the result of the enactment of small, automated situation-action pairs that have been termed *behavioral schemas* (Selden, McKee, & Selden, 2010). A common beneficial behavioral schema consists of a situation where one has to prove a universally quantified statement like, "For all real numbers x, $P(x)$" and the action is writing into the proof something like, "Let x be a real number," meaning x is arbitrary but fixed. Focusing on such behavioral schemas, that is, on small habits of mind for proving, has two advantages. First, the uses and interactions of behavioral schemas are relatively easy to examine. Second, this perspective is not only explanatory but also suggests concrete teaching actions, such as the use of practice to encourage the formation of beneficial schemas and the elimination of detrimental ones. (See the case of Sofia and her "unreflective guess" behavioral schema in Selden, McKee, and Selden, 2010, pp. 211-212).

While a number of proof construction actions (e.g., Selden, McKee, & Selden, 2010) have been investigated, our thinking about proof validation actions is still in its infancy. However, it seems reasonable to conjecture, based on the extant proof validation literature (e.g., Inglis & Alcock, 2012; Selden & Selden, 1995, 2003, Weber, 2008) that examination of the overall structure of a proof is crucial in order to determine whether the given attempted proof, if correct, actually proves the statement (theorem) that it sets out to prove. In addition, it also seems that a careful line-by-line reading of an attempted proof is useful for determining whether individual assertions are warranted, either explicitly or implicitly (e.g., Weber & Alcock, 2005). However, one often also wants to "get a sense of" a proof—What makes it work? What are the key ideas? These questions refer to the explanatory function of proof (deVilliers, 1990). One can consider these questions as part of the epistemic aims of the reader of a proof (Weber, Inglis, Mejia-Ramos, 2014).

Importance of Validation

Although we are focusing here on the validation abilities and practices of U.S. undergraduates who are at least in their second year of mathematics study, validation appears to have a role to play throughout mathematics students' education and in mathematicians' practice.

Holders of U.S. bachelor's degrees in mathematics are normally expected, not only to know considerable mathematics content, but also to be able to construct moderately complex proofs and to solve moderately nonroutine problems. Indeed, one major way that an individual's mathematical knowledge of a theorem is sometimes taken to be warranted is by the ability to "produce" a proof, not in a rote way, but in the way a mathematician would produce it, namely, with understanding (Rodd, 2000). However constructing or producing proofs appears to be inextricably linked to the ability to validate them reliably, and a "proof" that could not be validated would not provide much of a warrant. Similarly, solving moderately nonroutine problems appears to call for abilities akin to validation because one should ascertain whether the proposed solution is correct.

Pre-service secondary mathematics education majors and in-service secondary mathematics teachers also need to be able to validate proofs reliably because school mathematics curricula are likely to place increasing emphasis on justification and proof (NCTM, 2000). In particular, the CCSS-M Mathematical Practice Standard MP3 states that students should be able to "Construct viable arguments and critique the reasoning of others." (CCSS-M, 2014). In this regard Cuoco has observed informally, but based on considerable experience, that "The best high school teachers are those who have [had] a research-like experience in mathematics." (2001, p. 171).

In addition to being important for mathematics majors and mathematics teachers, validation appears to play a fundamental role in mathematicians' practice. While some mathematicians can sometimes obtain conviction in other ways (Weber, Inglis, Mejia-Ramos, 2014), mathematicians' belief in the general reliability and unproblematic nature of validation supports the assurance needed to use a theorem in later work. That is, when a theorem is proved one can expect it to "stay proved." This reliability seems to have been a major contributor to the rapid growth that has characterized twentieth century mathematics.

Setting of the Research: The Course and the Students

The course has been taught by the authors for several years at a U.S. Ph.D.-granting university. It is meant as a second-year university transition-to-proof course for mathematics and secondary education mathematics majors, but is often taken by a variety of majors and by more advanced undergraduate students.[3] The course was taught in a very modified Moore Method way (Coppin, Mahavier, May, & Parker, 2009; Mahavier,1999) and was modified somewhat each year. The students are given course notes with definitions, questions, requests for examples, and statements of theorems to prove. In addition, the course notes contain one detailed sample set theory proof construction, some explanations of types of proof frameworks (Selden & Selden, 1995) and a number of operable interpretations of definitions.[4] That is, we include information about how to use a definition in a proof and about how to prove that a definition is satisfied. We have found this level of detail useful, sometimes even necessary, for our second-year university students. Our aim is to gradually withhold such help, so students become autonomous by the end of the course.

To illustrate this, we have found that a formal definition, such as, $f(A) = \{ y \mid \text{there is an } a \in A \text{ so that } f(a) = y \}$ is often difficult for students to use. So we have added the following operable interpretation: To *show* "$y \in f(A)$" you show "there is an $a \in A$ such that $f(a) = y$." To *use* "$y \in f(A)$" you may say "there is an $a \in A$ such that $f(a) = y$." Yet, despite having included such operable interpretations, we have sometimes overheard students, during group work in a similar subsequent transition-to-proof course, say of our operable interpretations that they don't know what these mean. It is now our conjecture that students may, in addition, need specific examples of how to use a definition in proof construction and of how to show that a definition is satisfied.

The students in this study proved the theorems in the course notes outside of class and presented their proofs in class on the blackboard and received extensive critiques. These critiques consisted of careful line-by-line readings and validations of the students' proof attempts, often with corrections and insertions of missing warrants. In a sense, the second author modeled proof validation for the students. This was followed by a second reading of the students' proof attempts, indicating how these might be written in "better style" to conform to the genre of proofs (Author, 2013). Once these corrections and suggestions had been made, the student, who had made the proof attempt, was asked to write it up carefully, including the indicated corrections and suggestions, for duplication for the entire class. In this way, by the end of the semester, the students had obtained one correct, well-written proof for each theorem in the course notes. Given these careful critiques of student work—consisting of the line-by-line checking of students' proof attempts (i.e., modeling proof validation), followed by a second reading to help with the "style" in which the proofs are written, and finally, a carefully rewritten final proof by a student—we expected that these students might "adopt" some of the second author's techniques of validation and be able to implement them in their own proving attempts.

In addition, about once a week, the class worked in groups to co-construct[5] proofs of upcoming theorems in the notes. Sometimes, if the students seemed to need it, there were mini-lectures on topics such as logic or proof by contradiction. However, these mini-lectures were not preplanned; rather they occurred spontaneously, as the need arose. The homework, assigned each class period, consisted of requests for proofs of the next two or three theorems in the course notes. These proof attempts were handed in at the beginning of the next class to the first author, who determined "on the spot", based on the students' written work, which students would be asked to present their proof attempts on the blackboard that day. The students were aware that being asked to present their proof attempts did not necessarily mean that these were correct, but rather that their proof attempts would probably provide interesting points for the second author to discuss. In addition to presenting their attempted proofs in class, the students had both mid-term and final examinations, which consisted of theorems, new to them, to prove. The mathematical topics considered in the course included sets, functions, continuity, and beginning abstract algebra in the form of a few theorems about semigroups and homomorphisms. However, the teaching aim was to have students experience constructing as many different kinds of proofs as possible, especially in abstract algebra and real analysis, and *not* to have them learn a particular mathematical content. The course notes were self-contained, that is, all relevant definitions were provided.

---

[3] We have found that students are often afraid of a transition-to-proof course, and that sometimes instead of taking it in their second year of university, before courses like abstract algebra and real analysis (with which it is supposed to help), they take it later.

[4] Bills & Tall (1998, p. 104) considered a definition to be *formally operable* for a student if that student "is able to [actually] use it in creating or meaningfully reproducing a formal argument [such as a proof]". An "operable interpretation" is meant to help a student do that.

[5] A detailed description of the co-construction of real analysis proofs is given (Authors, 2010)

Methodology of the Study: The Conduct of the Interviews

Sixteen of the 17 students enrolled in the course opted to participate in the study for extra credit. Of these, 81% (13 of 16) were either mathematics majors, secondary education mathematics majors, or were in mathematics-related fields (e.g., electrical engineering, civil engineering, computer science).

Interviews were conducted outside of class during the final two weeks of the course. The students received extra credit for participating and signed up for convenient one-hour time slots. They were told that they need not study for this extra credit session. The protocol was the same as that of Selden and Selden (2003).

Upon arrival, participants[6] were first informed that they were going to validate four student-constructed "proofs" of a single number theory theorem, indeed, that the proof attempts that they were about to read were submitted for credit by students, who like themselves, had been in a transition-to-proof course. The participants were asked to think aloud and to decide whether the student-constructed proof attempts were indeed proofs. Participants were encouraged to ask clarification questions and were informed that the interviewer would decide whether a question could be answered. They were given the same information, in the form of a Fact Sheet, about multiples of 3 that was provided to the participants of the Selden and Selden (2003) study. (See Figure 1).

FACT 1. The positive integers, $\mathbf{Z}^+$, can be divided up into three kinds of integers -- those of the form $3n$ for some integer n, those of the form $3n + 1$ for some integer n, and those of the form $3n + 2$ for some integer n.

For example,

1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11…
		$3n$	$3n+1$	$3n+2$	$3n$	$3n+1$	$3n+2$			
			where $n = 1$			where $n = 2$				

FACT 2. Integers of the form $3n$ (that is, 3, 6, 9, 12, . . .) are called *multiples of* 3.

FACT 3. No integer can be of two of these kinds simultaneously. So m is not a multiple of 3 means the same as m is of the form $3n+1$ or $3n+2$.

Figure 1 The Fact Sheet given to the study participants.

There were four phases to the interview: A warm-up phase during which the participants gave examples of the theorem: *For any positive integer n, if n^2 is a multiple of* 3, *then n is a multiple of* 3 and then tried to prove it; a second phase during which they validated, one-by-one, the four student-constructed proof attempts of the theorem (Fig. 2); a third phase during which they were able to reconsider the student-constructed proof attemps (presented altogether on one sheet of paper), and a fourth debrief phase during which they answered questions about how they normally read proofs (Fig 3.).

For any positive integer n, if n^2 is a multiple of 3, *then n is a multiple of* 3.

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

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

(c). Proof: Let n be an integer such that $n^2 = 3x$ where x is any integer. Then $3|n^2$. Since $n^2 = 3x$, $nn = 3x$. Thus $3|n$. Therefore if n^2 is a multiple of 3, then n is a multiple of 3. ■

[6] Because the proof attempts were constructed by undergraduate students and because the participants in this study were also undergraduate students, we will henceforth refer to the undergraduates in this study as "participants" to avoid confusion.

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

Figure 2 The student-constructed proof attempts that the participants saw during Interview Phases 2 and 3

The interviews were audio recorded. The participants wrote as much or as little as they wanted on the sheets containing the student-constructed proof attempts. Participants took as much time as they wanted to validate each proof, with one participant initially taking 25 minutes to validate "Proof (a)".

1. When you read a proof is there anything different you do, say, than in reading a newspaper?
2. Specifically, what do you do when you read a proof?
3. Do you check every step?
4. Do you read it more than once? How many times?
5. Do you make small subproofs or expand steps?
6. How do you tell when a proof is correct or incorrect?
7. How do you know a proof proves this theorem instead of some other theorem?
8. Why do we have proofs?

Figure 3 The final debrief questions asked of participants in interview Phase 4

The interviewer, who is the first author, answered an occasional clarification question, such as the meaning of the vertical bar in $3|n^2$, but otherwise only took notes, and handed the participants the next printed page when they were ready for it.

The data collected included: the sheets on which the participants wrote, the interviewer's notes, and the recordings of the interviews. These data were analyzed multiple times to note anything that might be of interest. Tallies were made of such things as: the number of correct judgments made by each participant individually; the percentage of correct judgments made by the participants (as a group) at the end of Phase 2 and again at the end of Phase 3; the validation behaviors that the participants were observed by the interviewer to have taken; the validation comments that the participants proffered; the amount of time taken by each participant to validate each of the student-constructed proof attempts; the number of times each participant reread each purported proof; the number of participants who underlined or circled parts of the ; the number of times the participants substituted numbers for n; and the number of times the participants consulted the Fact Sheet. Many of these are discussed below.

Commentary on the Four Student-Constructed Proof Attempts

First, we make some general remarks on interesting or unusual aspects of the four student-constructed proof attempts. The theorem is true and one of the student-constructed proof attempts is actually a proof of it, while three are not, that is, as proofs they are incorrect. Selden & Selden (2003, pp. 10-18) provided a textual analysis -- a kind of line-by-line gloss or elaboration -- of the theorem and the four student-constructed proofs that emphasized mathematical and logical points that a validator might, or might not, notice.

For example, they discussed such matters as the use of alternative terms (e.g., *assume* for *suppose*, *divides* instead of *multiple*), the role of individual sentences in furthering the argument, proper and improper uses of symbols, implicit assumptions (e.g., that a division of the argument into cases had been exhaustive), the correctness of inferences, computational errors, extraneous statements, and structural aspects of the four student-constructed proof attempts. Global properties such as whether an argument proves the theorem, as opposed to some other theorem, were also noted.

Here we consider the four student-constructed proof attempts, one-by-one, only very briefly. "Proof (a)" (Fig. 2) is not a proof. It consists of two independent subarguments each of which should have ended with the conclusion "*n* is a multiple of 3" or its equivalent, "*n* is divisible by 3." However, the odd case did not end this way, and the even case made this claim but did not properly justify it. In addition, while taking odd and even cases, when n^2 is a multiple of three, would not be wrong, these two cases, if proved correctly, would be essentially the same. That is, even and odd have nothing to do with the multiples of three.

If "Proof (b)" is treated as a proof of the contrapositive, which one of the participants (CY) in the current study did, it would be peculiar to mention "the contrary" in the beginning. Under a contrapositive interpretation, the final step could have been omitted entirely or replaced by "Thus in either case n^2 is not a multiple of 3." We regard this as a proof of the theorem, although one that might have been written more clearly. Had the role of n^2

and the division into two independent subarguments (cases) been made explicit rather than implicit, the proof would have been less confusing for validators, especially inexperienced ones.

"Proof (c)" has a gap in the reasoning, although some mathematicians have pointed out that the result is an immediate consequence of knowing that if p is prime and $p|ab$ then p|a or p|b. This observation points to the importance of the context of known results for proof and validation. For a mathematician, this is the whole of the relevant literature, but for a student, it is the material in the relevant part of the course notes or book. These students knew essentially nothing about number theory beyond our Fact Sheet (Fig. 1). The participants in both studies had been told, at the beginning of the interviews, that the student-constructed proof attempts were written by students like themselves, who were in a transition-to-proof course. This gave them crucial contextual information. Indeed, several participants in the current study wondered what the students who wrote the student-constructed proof attempts had been allowed to assume, but nothing that would "fill the gap" had been discussed with the participants during the course.

"Proof (d)" begins with the conclusion and arrives at the hypothesis, although not in a straightforward way. Hence, it can be considered a proof of the converse of the stated theorem.

Results: Participants' Observed Validation Behaviors

Given that validation can be difficult to observe, it is remarkable how verbal and forthcoming the participants in this study were. This enabled us to gather a variety of data, much of which is presented and discussed below.

All participants appeared to take the task very seriously and some participants spent a great deal of time validating at least one of the student-constructed proof attempts. For example, LH[7] initially took 25 minutes to validate "Proof (a)"[8] before going on, and VL initially took 20 minutes to validate "Proof (b)". The minimum, maximum, and mean times for validating each purported proof are given in Table 1.

Table 1: Time (in minutes) taken initially to validate the student-constructed proof attempts (during Phase 2)

	"Proof (a)"	"Proof (b)"	"Proof (c)"	"Proof (d)"
Maximum time	25	20	16	9
Minimum time	5	2	3	2
Mean time	8.8	8.5	6.3	4.5

The following validation behaviors[9] were observed as having been enacted by the participants; the percentages and absolute numbers are given in parentheses:

1. Underlined, or circled, parts of the student-constructed proof attempts (100%, 16);
2. Pointed with their pencils or fingers to words or phrases, as they read along linearly (50%, 8);
3. Checked the algebra, for example, by "foiling" $(3n+1)^2$ (62.5%, 10);
4. Substituted numbers for n to check the purported equalities (37.5%, 6);
5. Reread all, or parts of, the student-constructed proof attempts (87.5%, 14);
6. Consulted the Fact Sheet to check something about multiples of 3 (56.25%, 9).

Summarizing the above, participants used focus/reflection aids (1. & 2.); checked computations or tested examples (3. & 4.); revisited important points – perhaps as a protection against "mind wandering" (5.); and checked their own knowledge (6.). These actions all seem to be beneficial validation behaviors.

Results: Participants' Proffered Evaluative Comments

The participants sometimes voiced what they didn't like about the student-constructed proof attempts. For example, CY objected to "Proof (b)" being referred to as a proof by contradiction. He insisted it was a contrapositive proof and twice crossed out the final words "we have a proof by contradiction". Fourteen (87.5%) mentioned the lack of a proof framework,[10] or an equivalent, even though they had been informed at

[7] Initials, like LH, designate individual participants.

[8] As is the custom in some transition-to-proof course textbooks, we have placed quotations marks around the names of the proof attempts, e.g., "Proof (a)", to remind readers that these attempts may not actually be proofs.

[9] Ko and Knuth (2013, p. 27) referred to validation behaviors, such as checking line-by-line or example-based reasoning as "strategies" for validating proofs. We prefer the term "behaviors" as the act of underlining or circling parts of proofs is evidence of focus, not strategy, which usually entails a plan.

[10] A *proof framework* is a "representation of the 'top level' logical structure of a proof, which does not depend on a detailed knowledge of the mathematical concepts, but is rich enough to allow the reconstruction of the statement being proved or one equivalent to it." (Selden & Selden, 1995, p. 129). In practice, in this transition-to-proof course, this meant writing the hypotheses at the top of the nascent proof, leaving a blank space for the

the outset that the students who wrote the proof attempts had not been taught to construct proof frameworks.

Below are some additional features that seemed to bother some participants:

1. Lack of clarity in the way the student-constructed proof attempts were written. Some referred to parts of the proof attempts as “confusing”, “convoluted”, “a mess”, or not “making sense” (68.75%);
2. The notation, which one participant called “wacky”;
3. The fact that “Proof (d)” started with *n*, then introduced *m*, and did not go back to *n*;
4. Not knowing what the students who had constructed the proof attempts knew or were allowed to assume:
5. Having too much, or too little, information in a purported proof. For example, one participant said there was “not enough evidence for a contradiction” in “Proof (b)”;
6. The “gap” in “Proof (c)” which was remarked on by six participants.

Results: Individual Participants’ Voiced Local and Overall Comments

Some participants made comments that indicated local concerns, but some comments were of an overall evaluative nature. That is, the overall comments often seemed to have more to do with making sense, having enough information, or being a “strong” proof, rather than with the structure or validity of the student-constructed proof attempts.

Indeed, no participant even commented on the strange division of “Proof (a)” into odd and even cases. This general lack of global, or overall, structural comments is similar to prior findings in the literature (e.g., Inglis & Alcock, 2012; Selden & Selden, 2003).

Local Comments

Some of the local comments on “Proof (a)” were:

MO: [I] don’t like the string of = s.
KW: $3n+1$, if $n=1$, is not odd, [rather it] would be even.
AF: This [pointing to $n^2 = 9n^2$] isn’t equal.

Two of local comments on “Proof (b)” were:

FR: [I am] not seeing the closing statement.
KK: [This is] not a proof because we don’t introduce *n*, but we use *n*.

A sample local comment on the use of the universal quantifier in “Proof (c)” was:

CJ: [The bit about] where *x* is *any* integer worries me.

A local comment on the notation used in “Proof (d)” was:

LH: Why would you use *m*? ... [It’s] kind of confusing with that *m*.

Overall Comments

Some overall comments on “Proof (a)” were:

CL: [It] needs more explanation -- I can’t see where they are going.
CY: [The] first case doesn’t seem right.
KW: [They are] Not going where they need to go.
FR: Not a proper proof.
MO: [This is a] partial proof.

Three overall comments on “Proof (b)” were:

CL: [This makes] a lot more sense to me [than “Proof (a)”].

---

details, and writing the conclusion at the bottom of the proof, and perhaps, also unpacking the conclusion and writing as much as possible of the structure of the proof.

SS: [It's] not written well.
AF: [I] feel like it's a proof because [they're] showing that the two integers in between [referring to $3k+1$ and $3k+2$] are not multiples of 3.

Four overall comments on Proof (c)" were:

CY: [I] just can't get my head around [it].
CJ: [I] need more information. [I] don't buy it.
KK: [This one is] closer [to a proof] than the others.
MO: [This one's a] sound proof.

Two sample overall comments on "Proof (d)" were:

MO: [He is] putting [in] more information than needs to be [there].
[This does] not help his proof.
LH: [This one's] not a strong proof.

Interpretation of Results

Participants' comments, illustrated in the above two sections, did not seem to focus primarily on whether the theorem had been proved. They included evaluative comments about whether they liked the student-constructed proof attempts, found them confusing or unclear in some way, or were lacking in some details or information.

According to the reading comprehension literature (e.g., Kintsch, 2004; Zwaan & Radvansky, 1998), unless reading is done totally superficially, the reader makes a *situation model* of the text.

> The situation model represents the information provided by the text, independent of the particular manner in which it was expressed in the text, and integrated with background information from the reader's prior knowledge. What sort of situation model readers construct depends very much on their goals in reading the text, as well as the amount of relevant prior knowledge that they have. … Situation models are not necessarily verbal. Texts are verbal, and textbases are propositional structures, but to model the situation described by a text, people often resort to imagery."
> (Kintsch, 2004, p. 1274-5).

We conjecture that the participants in our study may have been attempting to make a "situation model" of each of the proofs, that is, they were trying to understand, and make sense of, where the authors of the proof attempts were "coming from". Perhaps that is why they made comments about the student-constructed proof attempts not "making sense", having "wacky" notation, or being "confusing", "convoluted", or "a mess". Indeed, in their study, Selden & Selden (2003) observed similar attempts at sense-making. They said of their participants:

> In general, for these students, a feeling of understanding or not—that is, of making sense or not--seemed to be an important criterion when making a judgment about the correctness of these four "proofs." For them it seemed a question of whether the written text, together with their efforts at comprehension, engendered a personal feeling of understanding. (p. 26).

Students are not unique in their interest in understanding proofs. As Rav (1999) has stated, one important reason that mathematicians read proofs is to expand their understanding.

Results: What Participants Said They Do When Reading Proofs

In answer to the final debrief questions (Fig. 3), all participants said that they check every step in a proof or read a proof line-by-line. All said they reread a proof several times or as many times as needed. All, but one, said that they expand proofs by making calculations or making subproofs. In addition, some volunteered that they work through proofs with an example, write on scratch paper, read aloud, or look for the framework. All of these actions can be beneficial. Indeed, it is quite reasonable to suspect something might be wrong with a proof, if in an initial line-by-line reading, one or more logical implications cannot be warranted by the reader (Weber

& Alcock, 2005). Such a situation calls for a rereading, or a rethinking, of the proffered argument.

In addition, ten (62.5%) said they tell if a proof is correct by whether it "makes sense" or they "understand it". These are cognitive feelings that, with experience, can be useful. Four (25%) said a proof is incorrect if it has a [single] mistake, and four (25%) said a proof is correct "if they prove what they set out to prove." These last two views of proof call for some caution during implementation.

It is possible for a proof to have a minor mistake, perhaps a calculation error, that can be easily fixed, and hence, not "make sense" locally, but otherwise be correct. Indeed, it has been claimed that a past editor of the *Mathematical Reviews* once said that "approximately one half of the proofs published in it [that is, published in regular mathematics journals and abstracted in the *Mathematical Reviews*[11]] were incomplete and/or contained errors, although the theorems they were purported to prove were essentially true." (de Villiers, 1990, p. 19). Consequently, it appears that, for most mathematicians, a mistake that is easily fixable does not mean the entire proof should be judged invalid. Additionally, it is possible, especially for student proof attempts, to "end in the right place", but still have significant errors. Thus, cognitive feelings such as those expressed in the previous paragraph need to be informed by appropriate proof construction and validation experiences.

In addition to interpreting their task as first making sense of what they were reading, probably due to their prior experiences with reading and making situation models, we conjecture that the participants in this study might have felt it important, perhaps even necessary, to gain a top-level view of each proffered argument, that is, to be able to comprehend it holistically (Mejia-Ramos, Fuller, Weber, Rhoads & Samkoff, 2012, pp. 10-11) before making a judgment on its validity. Indeed, Selden & Selden (2003, p. 5) said of an ideal validation that, "Towards the end of a validation, in an effort to capture the essence of the argument in a single train-of-thought, contractions of the argument might be undertaken." Thus, perhaps the participants implicitly felt that making sense of the other students' proof attempts, that is, of where the student authors' were "coming from", was a prerequisite to being able to judge whether they were indeed proofs.

Discussion and Teaching Implications

In answer to the first research question, the participants in this study took their task very seriously, but made fewer final correct judgments (73% vs. 81%) than the undergraduates studied by Selden and Selden (2003) despite, as a group, being somewhat further along academically. In this study, 56% (9 of 16) of the participants were in their fourth year of university, whereas just 37.5% (3 of 8) of the undergraduates in the Selden and Selden (2003) study were in their fourth year.

Because the participants in this study were completing an inquiry-based transition-to-proof course emphasizing proof construction, in which validation had been modeled extensively by the second author, we conjectured they would be better at proof validation than those at the beginning of a transition-to-proof course (i.e., those studied by Selden and Selden, 2003), but they weren't. We have tentatively concluded that if one wants undergraduates to learn to validate "messy" student-constructed proof attempts, in a reliable way, one needs to teach validation explicitly, perhaps through validation exercises or activities.

We stress this because it may seem counterintuitive. We note that, as students most mathematicians have received considerable implicit proof construction instruction through feedback on assessments and on their dissertations. However, most have received no explicit validation instruction, but are apparently very skilled at it; for otherwise, they would submit some invalid proofs for publication.

There is at least one caveat regarding this study. It could be that we have been comparing "apples to oranges" as the participants in this study were not given a pre-test on proof validation, but instead were being compared to different students at the beginning of a transition-to-proof course at a different university (Selden & Selden, 2003). This could be remedied with another study that administered both a pre-test and a post-test consisting of validation items. However, it is difficult to imagine a course that gave more attention to helping students with proof construction or that demonstrated validation of proofs more explicitly than this one, with such limited effect on students' validation behaviors.

As to how one might possibly teach students to validate "messy" student-constructed proofs, Boyle and Byrne (2014, Table 1) have suggested a rubric, which they refer to as a "proof assessment tool" meant to help university teachers give formative feedback to students on their proofs. Byrne has used this rubric to have her transition-to-proof course students comment on each other's proof attempts (personal communication, March 8, 2014). It will be interesting to see whether Byrne needs to give her students explicit instruction on how to use this rubric or whether her students can use it "off the shelf" without explicit instruction, and how proficient her students became at giving helpful feedback to each other on proofs and on validation.

[11] *Mathematical Reviews* is a journal and online database published by the American Mathematical Society (AMS) that contains brief synopses, written and signed by mathematicians with appropriate expertise, of many published articles in mathematics, statistics, or theoretical computer science.

Finally, we note that, at least in the U.S., many future teachers of secondary or tertiary mathematics take a transition-to-proof course. Thus, for future pedagogical purposes, it would be useful for today's mathematics and mathematics education majors to be able to distinguish between a proof being valid (i.e., guaranteeing the truth of the claimed theorem) and having additional positive or negative features. Our own teaching suggests that making this distinction would be helpful, and the results of this study suggest more work on making this distinction explictly is needed.

Future Research

In addition to proof validation, there are three related concepts in the literature: proof comprehension, proof construction, and proof evaluation. There has been little research on how these four concepts are related. In this study, we investigated one of these relationships -- whether improving undergraduates' proof construction abilities would enhance their proof validation abilities and have obtained some negative evidence.

Proof comprehension means understanding a (textbook or lecture) proof. Mejia-Ramos, Fuller, Weber, Rhoads, and Samkoff (2012) have given an assessment model for proof comprehension, and thereby described proof comprehension in practical terms. Examples of their assessment items include: Write the given statement in your own words. Identify the type of proof framework. Make explicit an implicit warrant in the proof. Provide a summary of the proof.

In this regard, Weber (personal communication, March 10, 2014) noted that "Our model [of proof comprehension], to some extent, was to go beyond validation which we felt was clearly an important activity in proof-reading, … but not the one that undergraduates and mathematicians typically engage in during their proof reading."

Proof construction means constructing correct proofs at the level expected of university mathematics students (depending which year they are in their program of study).

Proof evaluation was described by Pfeiffer (2011) as "determining whether a proof is correct and establishes the truth of a statement (validation) and also how good it is regarding a wider range of features such as clarity, context, sufficiency without excess, insight, convincingness or enhancement of understanding." (p. 5).

While it is still an open question as to how the above four concepts are related, in addition to our study, Pfeiffer (2011) conjectured that practice in proof evaluation could help undergraduates appreciate the role of proofs and also help them in constructing proofs for themselves. She obtained some positive evidence, but her conjecture needs further investigation. As for proof comprehension, it is an open question as to whether practice in proof comprehension would help any of proof evaluation, proof validation, or proof construction. In this regard, Weber recently wrote:

> It would seem that Pfeiffer's notion of evaluation is also different than ours, asking the reader to rate the proof along some non-validity dimensions … I would argue that understanding the proof in terms of our [proof comprehension] model would be an important precursor to making the evaluations that Pfeiffer described. If one wanted to judge the methods and significance of a proof, one would seem to want to be able to have a high-level summary of the proof and identify the main methods used in the proof. (personal communication, March 10, 2014).

Additionally, there is anecdotal evidence, obtained from several mathematics department chairpersons, that some of today's U.S. transition-to-proof courses/textbooks are thought to be inadequate for the task of actually transitioning students from lower-level undergraduate mathematics courses to upper-level undergraduate proof-based mathematics courses, such as abstract algebra and real analysis. Whether this is the case, and to what degree, should be investigated.

Finally, we feel that there is a need to develop characteristics of a reasonable learning progression for tertiary proof construction, going from *novice* (lower-division mathematics students) to *competent* (upper-division mathematics students), on to *proficient* (mathematics graduate students), and eventually to *expert*[12] (mathematicians).

References

[12] The terms *novice*, *proficient*, *competent*, and *expert* have been adapted from the Dreyfus and Dreyfus (1986) novice-to-expert scale of skill acquisition.

Authors (2010, 2013).
Bills, L., & Tall, D. (1998). Operable definitions in advanced mathematics: The case of the least upper bound. In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education*, Vol. 2 (pp. 104-111). Stellenbosch, South Africa: University of Stellenbosch.
Boyle, J., & Byrne, M. (2014). Supporting students to construct proofs: An argument assessment tool. *Proceedings of the 17th Annual Conference on Research in Undergraduate Mathematics Education*.
Common Core Standards Initiative (2014). *Common Core State Standards for Mathematics* [*CCSS-M*]. Downloaded March 11, 2014 from http://www.corestandards.org/Math/.
Coppin, C. A., Mahavier, W. T., May, E. L., & Parker, G. E. (2009). *The Moore Method: A pathway to learner-centered instruction* (MAA Notes No. 75). Washington, DC: MAA.
Cuoco, A. (2001). Mathematics for teaching. *Notices of the American Mathematical Society*, *48*, 168-174.
de Villiers, M. (November, 1990). The role and function of proof in mathematics. *Pythagoras*, *24*, 17-24.
Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the age of the computer*. Oxford: Blackwell.
Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, *43*(4), 358-390.
Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.) *Theroretical models and processes of reading*. 5th Edition, International Reading Association.
Ko, Y.-Y., & Knuth, E. J. (2013). Validating proofs and counterexamples across content domains: Practice of importance for mathematics majors. *Journal of Mathematical Behavior*, *32*, 20-35.
Mahavier, W. S. (1999). What is the Moore Method? *PRIMUS*, *9*, 339-354.
Mejia-Ramos, J. P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics*, *79*(1), 3-18.
Pfeiffer, K.(2011). Features and purposes of mathematical proofs in the view of novice students: Observations from proof validation and evaluation performances. Doctoral dissertation, National University of Ireland, Galway.
Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum Assoc.
Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, *7*, 5-41.
Rodd, M. (2000). On mathematical warrants: Proof does not always warrant, and a warrant may be other than a proof. *Mathematical Thinking and Learning*, *2*, 221-244.
Selden, A., McKee, K., & Selden, J. (2010). Affect, behavioural schemas and the proving process. *International Journal of Mathematical Education in Science and Technology*, *41*(2), 199-215.
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*, *34*(1), 4-36.
Selden, J., & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics*, *29*, 123-151.
Weber, K., & Alcock, L. (2005). Warranted implications to understand and validate proofs. *For the Learning of Mathematics*, *25*(1), 34-38 & 51.
Weber, K., Inglis, M., & Mejia-Ramos, J. P. (2014). How mathematicians obtain conviction: Implications for mathematics instruction and research on epistemic cognition. *Educational Psychologist*, *49*(1), 36-58.
Weber, K. (2008). How do mathematicians determine if an argument is correct? *Journal for Research in Mathematics Education*, *39*(4), 431-439.
Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 126*(2), 162-185.