Validations of Proofs Considered as Texts: Can Undergraduates Tell Whether an Argument Proves a Theorem?

Annie Selden, *Tennessee Technological University*
John Selden, *Mathematics Education Resources Company*

This article reports on an exploratory study of the way that eight mathematics and secondary education mathematics majors read and reflected on four student-generated arguments purported to be proofs of a single theorem. The results suggest that such undergraduates tend to focus on surface features of arguments and that their ability to determine whether arguments are proofs is very limited—perhaps more so than either they or their instructors recognize. The article begins by discussing arguments (purported proofs) regarded as texts and validations of those arguments, that is, reflections of individuals checking whether such arguments really are proofs of theorems. It relates the mathematics research community's views of proofs and their validations to ideas from reading comprehension and literary theory. Then, a detailed analysis of the four student-generated arguments is given and the eight students' validations of them are analyzed.

Key Words: Advanced mathematical thinking; College/university; Language and mathematics; Proof; Reasoning; Writing/communication.

Proofs, that is, arguments[1] that prove theorems, have been viewed from a wide variety of perspectives. Previous studies have considered the structure of proofs (Leron, 1983), their explanatory power (Hanna, 1989), errors and misconceptions in college students' proofs (Selden & Selden, 1987), the logic involved (Dubinsky & Yiparaki, 2000; Selden & Selden, 1995), university students' "proof schemes" (Harel & Sowder, 1998), and generic proofs (Rowland, 2002). In addition, one might describe the genre of proofs or discuss their rhetoric. Considerably more could also be done in examining the process of proof construction.

---

[1] We are using *argument* the way it is commonly used in the English-speaking mathematical community (Bagchi & Wells, 1998), where *proof* is often reserved for a correct argument consisting of a rather special kind of deductive reasoning that warrants the truth of a theorem (*démonstration,* in French). In that context, and in this article, argument refers to the same kind of deductive reasoning, except that it might not be correct. We are not referring to the more general kinds of arguments used outside of mathematics or to argumentation, for example, to discourse intended to persuade that might include alluding to experiences of the intended audience (see Perelman & Olbrechts-Tyteca, 1969 or Toulmin, 1958).

Here we focus on proofs as texts that establish the truth of theorems and on readings of, and reflections on, proofs to determine their correctness. We call such readings and the mental processes associated with them *validations of proofs* (Selden & Selden, 1995). 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. Toward 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. An annotated transcript of a sample (hypothetical) validation of a calculus proof can be found in Appendix I of Selden and Selden (1995, pp. 143–147).

In the sections that follow, we discuss a theoretical perspective on arguments, proofs, and validations and compare it with various 20th-century theories of texts and reading. We then give a detailed textual analysis of a single theorem and four brief student-generated arguments purporting to prove it. This analysis provides grounding for the previous theoretical discussion, as well as background for the subsequent description and analysis of eight undergraduate students' validations of the same four arguments. In the concluding sections, we offer some implications for teaching and pose some questions for future research.

VALIDATION AND OTHER KINDS OF READING

Our view of the nature of validation is reminiscent of late 20th-century ideas about reading found in works on reading comprehension and literary theory and criticism (as described in Bogdan & Straw, 1990 and Pearson, 2000). In these, the onus for the construction of meaning is on the reader rather than on the author or the written text. Each reader is seen as constructing his or her own story from personal background together with the written text. Different readers, or the same reader at different times, are seen as benefiting differently from a single written text. Validators of proofs, too, may benefit in differing ways according to their backgrounds. For example, some may develop an understanding of a theorem that goes well beyond just knowing it is true, and Hanna (1989) has pointed out that some proofs may be better for this than others. Also validation can sometimes involve emotionally intense reflection resulting in knowledge construction, for example, by forging new links between one's ideas.

At the beginning of the 20th century, theorists in reading comprehension and literary criticism regarded writing as mainly a matter of communication—the author was considered the ultimate arbiter of the meaning of a written text. To under-

stand a text, one studied the life and times of the author. By midcentury, the locus of meaning had shifted to the text itself,[2] and by the end of the century, the construction of meaning was viewed as residing with the reader. The idea that the meaning is in the text has been articulated by Olson (1977); it has been referred to as the "doctrine of autonomous texts" and was critiqued by Nystrand, Doyle, and Himley (1986), who included an example that may shed some light on the unique character of proofs. They pointed out that legal contracts are documents explicitly meant to be autonomous and independent of context, but nevertheless such documents cause disputes that courts are asked to settle by calling in experts to clarify the original terminology and context. This example supports the idea that independence of context is inherently unachievable. In contrast, such situations do not occur with proofs; one does not call in experts to find out what an author meant by, say, "compact" to establish the truth of a theorem.

Like midcentury structural critics, mathematicians[3] seem to treat a proof as being independent from its author. If something, perhaps the author's reputation, should suggest that a proof is probably correct, mathematicians may attempt to suspend their credulity similar to the way an audience in a classical theater was supposed to suspend its disbelief arising from unrealistic costumes and surroundings. That is, a mathematician may believe a theorem asserted by a well-known author to be true, but may read the arguments in its proof *as if* the statement of the theorem were in question. This predilection for separating authors from proofs appears to be reflected in the way proofs are written, for example, in the avoidance of autobiographical comment in published mathematical research. Indeed a general tendency to impersonalize mathematical texts, especially research papers, has been noted in analyses of their "natural language" aspects (Burton & Morgan, 2000).

By the early 1990s, reading and writing were beginning to be seen as different aspects of one unified process. Similarly, validation and proof construction might be seen as differing aspects of a single process. As such, they are probably best learned in a dialectical way. On the one hand, one constructs a proof with an eye toward ultimately validating it and may often validate parts of it during the construction process. In fact, the final portion of a proof construction is likely to be a validation of that proof. On the other hand, proofs are written for idealized readers, yet there is considerable variation in what actual mathematicians know. Thus, the validation of a proof is likely to require the construction of subproofs. That is, each process, validation and proof construction, entails the other.

Validation and proof construction also differ in important respects. Proof construction is much more like mathematical problem solving, in the sense of

---

[2] The idea that meaning can reside in the text was not new to the 20th century. It appears to go back at least to the time of Plato and became a major tenet for Martin Luther, who asserted that the meaning of Scripture was to be found in a deeper reading of the text (Olson, 1977, p. 263).

[3] In the context of this article, when mentioning mathematicians or the mathematics community, we are referring especially to those engaged in research in pure mathematics, that is, in finding and proving theorems.

Schoenfeld (1985), than is validation. Generally constructing a proof requires that more diverse ideas come to mind at the "right time" than validating it does. Also, like most reading, validation normally emphasizes proceeding linearly from the beginning to the end of a written proof—perhaps repeated several times. This linear order is unlikely to occur in proof construction. Given a theorem to prove, one must often attend not only to the beginning but also to the end of a proof before developing the middle. In addition, many proofs have a hierarchical structure based on subproofs and subconstructions that emerges during the process of proof construction. Even assuming an idealized prover (one making no mistakes or false starts, which is an inherently unrealistic assumption), such an emergent hierarchical structure may lead to generating subarguments in an order very different from that of the final written text. Indeed, for the working mathematician, a theorem and its proof sometimes emerge in a dialectical way over a long period of time, with adjustments in one alternating with adjustments in the other.

One distinction between current ideas of reading and that of validation is that written proofs are still seen as carrying considerable meaning. The proofs themselves (as opposed to their readers or authors) are regarded by mathematicians as the ultimate establishers of the truth of theorems.[4] One examines neither the life and times of a proof's author nor the sophistication of its readers in judging the truth of a theorem. Perhaps this is partly because any expanded or altered argument constructed by a validator is often seen as closely aligned with the original written proof, so that the original proof is deemed correct (in some general sense) provided the expanded, validator-generated text is seen by the validator as correct.

Another distinction between validation and reading in general appears to be the unusual degree of agreement about the correctness of arguments and the truth of theorems arising from the validation process. It is our experience that if two mathematicians disagree on the correctness of a proof, they will often attempt a joint validation of it (or a fragment of it). Typically, they will either expand the proof and agree on the expanded version or, failing that, they will find and agree on a mistake that cannot be fixed. Occasionally, they may find and correct small errors that they agree are insignificant and do not affect the overall correctness of the argument. This ability to reach agreement appears to depend to some extent on the way that mathematical definitions have come to be treated as more or less unchallengeable in the 20th century.

Today, definitions in mathematics are typically treated as analytic, as reducible to undefined terms except for foundational considerations, such as the meaning of set, element, or integer. Such considerations are not major issues for most mathematicians perhaps partly because "normal" foundations yield a mathematics that can be interpreted as paralleling an intuitive view of the world developed socially

---

[4] This view is supported by the ability of computers to check the correctness of many (more formally stated) proofs. For example, a large number of proofs have been machine verified after conversion to the Mizar language and are published online in the *Journal of Formalized Mathematics* (see mizar.org/JFM/).

from an early age. With both "normal" foundations and an intuitive view one can consider several objects together, count them, add, measure things, and so on.[5] In contrast with analytic definitions, synthetic definitions[6] are the everyday definitions that are commonly found in dictionaries—they are often ill-specified descriptions of things that already exist. It may be unclear when such everyday definitions (e.g., of democracy) are "complete" or whether attention to all aspects of them is essential for their proper use. Analytic definitions, however, bring concepts into existence—the concept (e.g., of *group* as used in mathematics) is whatever the definition says it is, nothing more and nothing less. One cannot safely ignore any aspects of such definitions. There is a sense in which synthetic definitions can be "wrong," that is, they can be inaccurate descriptions, but analytic definitions cannot, since nothing is being described.[7]

For example, our synthetic definition of *validation* could be wrong; we might have inaccurately described what we claimed is a recurring phenomenon. In contrast, few mathematicians would say the analytic definition of "group" was wrong—uninteresting, tasteless, lacking applications, or inaccurately remembered perhaps, but not wrong. Today's use of analytic definitions renders validations (and proofs and theorems) very reliable, whereas the earlier use of synthetic definitions sometimes left them problematic. This situation can be seen in *Proofs and Refutations* (Lakatos, 1976) where validations appear inherently unreliable, that is, unreliable even if there are no errors. There, theorems have proofs *and* refutations (counterexamples) apparently due to treating definitions as synthetic and hence challengeable. Given some literary license, *Proofs and Refutations* provides a fairly accurate description of the way that some mathematics developed historically. It illustrates how definitions and results can coevolve, but the compression of time involved in its fictional narration, together with its synthetic treatment of definitions, may suggest that validations, proofs, and theorems are far less reliable than they really are today.

Roles Played by Validation

Although we are focusing here on the validation abilities and practices of under-graduates who are in at least their 2nd year of mathematics study, validation

---

[5] However, nothing in principle restricts mathematicians to particular foundations. For example, in intuitionist/constructivist mathematics, one forgoes mathematical induction and the logical law of the excluded middle, resulting in a more limited mathematics without proofs by induction or contradiction.

[6] Our distinction between *analytic* and *synthetic* follows that made by Sierpinska, Defence, Khatcherian, and Saldanha (1997) in discussing linear algebra. Similar distinctions have been made by Edwards (1997) using *lexical* and *logical* and by Freudenthal (1973) using *constructive* and *descriptive*. In considering defining as a mathematical activity in geometry, Rasmussen and Zandieh (2000) and de Villiers (1998) adopted Freudenthal's terminology.

[7] Analytic definitions can, however, be *inspired* by regularities in the physical world, in which case a mathematical structure may be metaphorically represented in a physical situation. Vectors are useful in navigation because the motion of a sailboat (in part) "works like" vector addition.

appears to have a role to play throughout mathematics students' education, in mathematicians' practice, and in examinations of the nature of mathematics.

Validation in mathematics education. Holders of 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 and with understanding (Rodd, 2000). However, constructing or producing proofs is inextricably linked to the ability to validate them reliably, and a proof that could not be reliably validated would not provide much of a warrant. Solving moderately nonroutine problems also appears to call for abilities akin to validation, because one must ascertain whether a proposed solution is correct. In our ongoing investigation of mathematicians' views, we are finding that many mathematicians believe that checking the correctness of solutions ultimately depends on the same kind of deductive reasoning that is used in proofs and their validations.

Preservice secondary mathematics education majors and mathematics teachers also need to be able to validate proofs reliably because school mathematics curricula are likely to place increasing emphasis on proofs and problem solving (NCTM, 2000). In this regard, Cuoco has observed informally but on the basis of considerable experience that "The best high school teachers are those who have a research-like experience in mathematics" (2001, p. 171). And in an article in the *Mathematics Teacher*, Thompson (1996) suggests teaching indirect proof to high school students using number-theoretic statements that are similar to, although perhaps somewhat less complex than, the theorem used with undergraduates in our exploratory study.

Validation in the practice of mathematicians. In addition to being important for mathematics majors and mathematics teachers, validation appears to play a fundamental role in the production of new mathematics. A mathematician's belief in the reliability and unproblematic nature of validation provides the assurance needed to use a theorem in later work without repeatedly checking its correctness. That is, when a theorem is proved "it stays proved." This reliability seems to have been a major contributor to the rapid growth that has characterized 20th-century mathematics. Furthermore, many mathematicians seem to rely mainly on validation, sometimes carried out individually and sometimes in seminars, to learn new mathematics from the published research of others.

Validation within social constructivist (and other) perspectives. Our comments about the truth of theorems might suggest that we are taking a Platonist perspective, with its timeless abstract objects and absolute truth independent from human considerations. However, in this article, we are not favoring any single perspective on mathematics. Validation is a largely mental process with social origins and social consequences that bears examining relative to any of several views on the nature of mathematics. Furthermore, for many mathematicians, the truth of a theorem appears to be *functionally* just the existence of an argument that they

personally (or their presumed surrogates) have validated and deemed to be a proof of that theorem. In particular, a further investigation of validation might be helpful in viewing mathematics from a social constructivist perspective—where mathematics is seen as an entirely social, fallible[8] phenomenon with the idea of truth replaced by social agreement and where abstract Platonic objects have no role (Ernest, 1998). An accurate description of how new mathematics is created is crucial from this perspective because, in a sense, how it is done *is* mathematics. We suspect that, if analyzed together, proofs and validations thereof might in the future play a larger role within a social constructivist perspective. In such an analysis, foundations and the kind of logic used could be treated as hidden, that is, tacitly assumed, premises of theorems. By examining the extent of agreement and the way it arises from the validation process, one might find a way to see mathematics as locally[9] infallible and thereby reconcile this philosophical view with the more-or-less Platonic working views of most research mathematicians. Indeed, those working views may have emerged mainly from experience with the reliability of validation, rather than from any prior philosophical commitment.

Regardless of one's perspective on the nature of mathematics or its knowing, investigations of validation as a form of reading (and of the arguments being read), with emphasis on what validators do, could inform mathematics education. Our exploratory study is a start in this direction. We begin with an analysis of some specific arguments.

A TEXTUAL ANALYSIS OF A THEOREM AND FOUR "PROOFS"

In this section, we analyze a theorem and four undergraduate student-generated arguments purported to prove it. This analysis will provide concrete grounding for the above theoretical discussion of argument, proof, and validation, as well as for the subsequent description of our exploratory study of other undergraduate students' validations of these same four arguments. It may also expose aspects of these texts that might otherwise be missed. For example, we point out that certain portions of the first argument are extraneous and as such have no effect on its correctness as a proof. Nevertheless, errors in these extraneous portions were noted and thought to be important by some of our student validators.

The theorem and the four arguments are treated here primarily as texts, that is, as independent of specific authors and validators. We will, however, comment in general on how aspects of these texts may have come to be written or might affect readers, occasionally illustrating this point with comments from our student validators. The theorem is true and one of the arguments is a proof of it, whereas three

---

[8] Here, following Ernest (1998), we mean by *fallible* something more than just the (generally accepted) human tendency to make mistakes.

[9] By *locally* we mean the mathematics varies according to the foundations and logic used. For example, a mathematical constructivist would not accept the law of the excluded middle and, consequently, obtain different theorems than a mathematician who did accept the law of the excluded middle. We are not referring to geographic or social variations.

are not; that is, as proofs they are incorrect. These characterizations—true, correct, proof—are widely regarded by mathematicians as global features of texts alone that can be determined without knowing who the authors or readers are. Mathematicians say that an argument proves a *theorem*, not that it proves it for Smith and possibly not for Jones, who are readers of the argument. This view is held despite the fact that the texts are the products of human activities, that their physical representations could not exist without authors, and that validators are required for anyone to know the theorem is true. Perhaps such assignments of proof, truth, and correctness (or their opposites) to texts is due to mathematicians' remarkably uniform agreement on whether these properties are associated with particular texts—at least at the level of the mathematics discussed here and modulo modest concerns about style or completeness.

Texts (and, more generally, language and communication) have been analyzed from several perspectives: syntactic structure (e.g., Chomsky, 1957), semantic structure (e.g., Fillmore, 1968; Kintsch, 1974), and language in use (Halliday, 1977; Vachek, 1966). Relevant ideas emphasizing language in use that are especially applicable to mathematical texts have been summarized and illustrated by Morgan (1996). These ideas include such features as the ways reasoning is expressed (e.g., *because, so*), the symmetry or asymmetry of the relationship between authors and readers, and what texts might suggest about the nature of mathematics. Most of these features, for example the use of "we" in mathematical texts (Pimm, 1984), have no bearing on whether an argument proves a theorem; and in our experience, undergraduate students realize this and concentrate on the underlying mathematical concepts and inferences.

Thus, we provide a different kind of textual analysis, one that is a kind of line-by-line gloss, or elaboration, of the theorem and the four student-generated arguments emphasizing mathematical and logical points that a validator might, or might not, notice. For example, in considering the statement of the theorem, we discuss both its logical structure and other, more informal, ways that it could be written. For the four student-generated arguments, we discuss 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 is exhaustive), the correctness of inferences, computational errors, extraneous statements, and structural aspects of the argument. Global properties, such as whether an argument proves the theorem, as opposed to some other theorem, are also discussed.

In the following analyses, the names of the four arguments (e.g., *"Proof (a)": Errors Galore*) and the numbering of their parts are for the convenience of the reader; the names and numbering were not shown to the eight student validators in our study. The theorem to be proved was this:

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

Both the student authors of the four arguments and the student validators in our empirical study were given this statement of the theorem. It is a formal statement in that the variable *n*, the set to which it refers (the positive integers), and the universal quantification, *any*, are explicitly stated, and the conditional is expressed in its most common form, *if . . . , then*. However, other, less formal statements of this theorem convey the same logical information; their equivalence can be established using logic and little or no additional mathematical content. Often, theorems that undergraduate students at this level consider are stated in a less formal way. For example, one informal statement of this theorem is *A positive integer is a multiple of* 3 *whenever its square is.*

Undergraduate students similar to the ones in our empirical study appear to have considerable difficulty interpreting informal statements of theorems and converting them to formal ones (Selden & Selden, 1995). In our experience, formal statements appear to be the ones most easily interpreted in validations; they correspond to what we have called *proof frameworks.* These consist of the beginning and end of a proof—the part that can be written without knowing the meanings of all of the mathematical terms and from which the theorem can be reconstructed. It is an understanding (usually tacit) of the relationship between statements of theorems and their proof frameworks that allows validators to decide if an argument proves the theorem at hand as opposed to some other theorem. A proof framework[10] for this theorem is this: PROOF. Let *n* be a positive integer and suppose n^2 is a multiple of 3. . . . Then *n* is a multiple of 3. Often the introduction of a variable, for example, "Let *n* be a positive integer," is omitted but understood. This omission usually occurs when the variable appears in the statement of the theorem, as it does here. Our earlier work (Selden & Selden, 1995) suggests that, among the various possible ways to state a theorem, student validators should find the more formal statement used here the easiest to correctly connect with arguments purporting to prove it.

In what follows, we state each student-generated argument verbatim and then give our line-by-line analysis of it.

"Proof (a)": Errors Galore

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. ∎

Analysis of "Proof (a)"

[1] *Assume that n^2 is an odd positive integer that is divisible by* 3. This part of the argument is not wrong. In this context one can take *assume* to mean *suppose*,

_____

[10] Proof frameworks are often more complex than this. For examples from calculus and semigroup theory, see Selden and Selden (1995, pp. 129–130).

and an author of a proof can suppose any reasonable non-selfcontradictory statement which might aid the progress of the argument. But, as some of our student validators noted, this division into odd and even cases is a peculiar way to structure a proof of this particular theorem. It is peculiar because odd and even appear to be unrelated to multiples of 3.

That the phrase "n^2 is divisible by 3" is an equivalent way of saying "n^2 is a multiple of 3" is something a validator would need to assent to (by noting that the definitions of *divisibility* and *multiple* are mathematically equivalent). One student validator, whom we refer to in this article as SW, briefly noted (and assented to) the fact that "divisible by 3" was used here; subsequently, in the even case, "multiple of 3" was used.

[2] *That is $n^2 = (3n + 1)^2 = 9n^2 + 6n + 1 = 3n(n + 2) + 1$.* There are four errors here. First, the assumption of oddness was about n^2, not n, to which it is apparently being applied (i.e., $n = 3n + 1$). Second, n is being used in two places with two different meanings within one equation; in effect, the author is mistakenly asserting that $n = 3n + 1$. In the genre of mathematical proofs, it is not permissible to let the same symbol represent two different numbers except across independent subproofs. Doing so would be very likely to cause validators confusion. Indeed, student ST remarked that "all those n's trying to equal each other" misled her. Third, an odd integer n can be represented as $n = 2m + 1$, not as $3m + 1$, where m is a different integer from n. Fourth, the last equation is incorrect; $3n$ times n yields $3n^2$, not $9n^2$.

We would like to comment further on the author's apparent idea that $3n + 1$ can represent a general odd number. We doubt that this idea indicates a more-or-less fixed misconception in the author's knowledge base, but rather suspect it is an incorrect, partial recollection of the idea that $2n + 1$ is odd for integer values of n. There appears to be a more subtle lack in the author's knowledge, and we will offer a conjecture of its nature and how it occurs. Students of mathematics probably recognize (mostly tacitly) many kinds of problems and tasks, and each kind may be associated with many other ideas, forming a mental structure like a concept image (Tall & Vinner, 1981) for that problem or task—what we have called a problem-situation image. When a kind of problem or task is recognized, an associated problem-situation image may be activated (brought to mind) or partly activated (Baddeley, 1995). In particular, accomplished students might associate the task of representing a general odd number with a feeling of caution about getting the representation right. As they execute the task, this feeling of caution may engender checking. In the case at hand, that might mean substituting some numbers for n: substituting 1 for n (in $3n + 1$) yields 4 and indicates that $3n + 1$ need not be odd. We suspect that the author's problem-situation image(s) for this task did not include an appropriate feeling of caution. A mechanism for bringing something to mind similar to the one sketched here, concerning how to start solving a recognized kind of problem, has been more fully described in Selden, Selden, Hauk, & Mason (2000).

[3] *Therefore, n^2 is divisible by* 3. It was assumed in [1] that n^2 is divisible by 3. Thus, this statement cannot be wrong despite the intervening errors in [2]; however, it cannot be of use either. Furthermore, this statement completes one of two inde-

pendent subproofs, one for n^2 odd and one for n^2 even, and should have ended with n as a multiple of 3, not n^2.

The errors in [2] might be called extraneous errors. Because they do not affect the correctness of [3], they cannot affect whether or not the argument is a proof. Such errors are nevertheless undesirable because they can make arguments confusing and more difficult to validate. In our empirical study, most student validators spent considerable time trying to decipher [2], making comments like "$3n + 1$ doesn't make sense [for odd]," but did not seem to notice its irrelevance.

[4] *Assume that n^2 is even and a multiple of* 3. This assumption is an appropriate beginning of the second independent subproof (case). However, as already noted under [1], this division into cases, although not wrong, is unlikely to be useful.

[5] *That is $n^2 = (3n)^2 = 9n^2 = 3n(3n)$.* In the first equation $n^2 = (3n)^2$, it appears that the assumption that n^2 is a multiple of 3 is incorrectly applied to n instead of n^2. Also, in the implicit equation $n = 3n$, once again the right-hand n does not have the same meaning as the left-hand n. To indicate that n is a multiple of 3, one represents n as $3m$, where m is another positive integer. The second and third equations are correct.

[6] *Therefore, n^2 is a multiple of* 3. This conclusion is not incorrect; it does follow from [5] (if one replaces all but the first "n" with "m's"). But it also just repeats the assumption in [4], which is not likely to be useful.

[7] *If we factor $n^2 = 9n^2$, we get $3n(3n)$.* Here the author appears to be attempting to restate information from [5], which could not introduce any new errors. Since the word "factor" refers to expressions, not equations, we take it that [7] is a shortened form of "Now $n^2 = 9n^2$, and if we factor $9n^2$, we get $3n(3n)$."

[8] *which means that n is a multiple of* 3. ■ This inference from [7] is not correct even if we change $3n(3n)$ to $3m(3m)$. To conclude that n is a multiple of 3, one would expect to see 3 as a factor of n, not as a factor of n^2. The symbol "■" is one of those customarily used to indicate the end of a proof.

In summary, this argument is not a proof. It consists of two independent subarguments each of which should end with "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. It is left implicit that the two cases are exhaustive, that is, that every integer is either even or odd, but it is customary to leave such a well-known fact implicit.

We would like to mention two kinds of global errors that this argument does not contain. First, although the argument has a high density of errors, they are not what one might call mathematically syntactical errors (Selden & Selden, 1987). The mathematical terms and symbols fit together in a meaningful way (i.e., one can judge whether each individual statement is correct or incorrect). Second, the two subarguments really are independent. Occasionally, one finds arguments in which an inference in one subargument improperly depends on information from inside another supposedly independent subargument, and that error did not occur here.

Last, we have taken this argument at face value; that is, we have assumed that the author meant to write a proof and had some more-or-less clear (although incorrect) conception of how the argument would proceed. However, because the student author submitted this "proof" as part of a take-home test, it is possible that he or she was merely seeking some (partial) credit and was thus working in what Vinner (1997) has called a pseudoconceptual mode,[11] mimicking some previously encountered argument about multiples of 2. One possible similar argument might have been the following proof of the statement: For any integer n, n^2 is a multiple of 2 if and only if n is a multiple of 2.

PROOF. If n is odd, then $n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$, which is odd. Therefore, if n^2 is a multiple of 2, then n is a multiple of 2. Now, if n is a multiple of 2, that is, if n is even, then $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$, which is a multiple of 2. ∎

"Proof (b)": The Real Thing

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. ∎

Analysis of "Proof (b)"

[1] *Suppose to the contrary that n is not a multiple of* 3. The phrase "to the contrary" may indicate that this argument is meant to be a proof by contradiction. It might also indicate that the argument will prove the contrapositive of the theorem, that is, that the negation of the conclusion implies the negation of the hypothesis, which is logically equivalent to the theorem. The student author may not be clear about the distinction. If the proof is really by contradiction, it is implicit that n represents a positive integer and n^2 is assumed to be a multiple of 3. The argument can end with any contradiction, but that contradiction *could be* to the hypothesis, that is, n^2 is not a multiple of 3. If the contrapositive of the theorem is being proved, it is implicit only that n represents a positive integer, and the argument *must* end with n^2 is not a multiple of 3. For both methods of proof it is appropriate to suppose the negation of the conclusion, that is, n is not a multiple of 3, as is done here.

[2] *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. This representation and division into cases correctly follows from the fact that any positive integer can be represented as exactly one of $3k$, $3k + 1$, or $3k + 2$ for some positive integer k. It is implicit that

---

[11] Indeed, one student validator, ST, seems to have concluded that the author of "Proof (a)" was working pseudoconceptually or at least had no clear conception of a proof. After finding several errors and noticing that certain parts did not make sense, she stated, "Basically, it's a bunch of nothing."

the positive integer being represented is n and that, according to [1], n cannot equal $3k$; thus, n must equal $3k + 1$ or $3k + 2$. A validator would need to check that this statement is correct. To comprehend [2], it would not be essential to see the implicit representation of n as one of the three expressions, but it would be in order to comprehend [3].

[3] *Now* $n^2 = (3k + 1)^2 = 9k^2 + 6k + 1 = 3(3k^2 + 2k) + 1$. This statement begins the first of two independent, "parallel" subarguments. Assuming that $n = 3k + 1$, all of the equations are correct.

[4] *Since* $3(3k^2 + 2k)$ *is a multiple of* 3, $3(3k^2 + 2k) + 1$ *is not.* It follows from considerations similar to those under [2] that only every third positive integer is a multiple of 3. So if an integer is a multiple of 3, then one more than it is not. A validator would need to assent to this. It is then implicit in [4] that $3(3k^2 + 2k) + 1$ is the n^2 mentioned in [3]. A validator would not have to notice this to comprehend [4], but would need to see it to comprehend [5].

By a validator assenting to something, in this case that $3(3k^2 + 2k) + 1$ is not a multiple of 3, we mean that the validator is conscious of agreeing with it. This conscious agreement may be fleeting and not focused on, but it is distinct from mere passive reading in inner speech. It may be the only conscious phenomenon associated with a fairly complex inference. In this case, the validator must have some knowledge of the relationship between multiples and multiples plus one to see that $3(3k^2 + 2k) + 1$ fits that relationship and infer that $3(3k^2 + 2k) + 1$ is not a multiple of 3. While the assenting is conscious, the inference leading to it may occur very quickly and outside of consciousness. It would be a validation error for a validator to passively read, rather than assent to or reject, a portion of a proof, even if the proof were correct and the validator agreed that it was correct. Such a validator would not have tested that portion of the proof and could not really know that the proof was correct.

[5] *Now we will do the other possibility,* $3k + 2$. This assertion indicates that the first subargument is finished and that the second subargument is beginning with the correct, but implicit, assumption that $n = 3k + 2$.

[6] *So,* $n^2 = (3k + 2)^2 = 9k^2 + 12k + 4 = 3(3k^2 + 4k + 1) + 1$ *is not a multiple of* 3. All of the equations are correct. The final assertion depends on the same information used to justify [4] and should be assented to by a validator. This statement concludes the second subargument.

[7] *Because* n^2 *is not a multiple of* 3, *we have a contradiction.* ∎ This statement is correct, assuming that it is a proof by contradiction. The implicit reasoning is that, according to [2], one of the two subarguments (cases) must apply to n^2. Because both subarguments yield that n^2 is not a multiple of 3, there is a contradiction to the implicit assumption that n^2 is a multiple of 3. If this argument were treated as a proof of the contrapositive, it would be peculiar to mention "the contrary" in [1]. However, under that interpretation, this final step could be omitted entirely or replaced by "Thus, in either case, n^2 is not a multiple of 3." We regard this argument as a proof of the theorem, although it might be written more clearly. If the role of n^2 and the division into two independent subarguments (cases) were

made explicit rather than implicit in [2], [3], [4], [5], and [6], this proof would be less confusing for a validator, especially an inexperienced one.

"Proof (c)": The Gap

PROOF. Let n be an integer such that $n^2 = 3x$ where x is an 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. ∎

Analysis of "Proof (c)"

[1] *Let n be an integer such that $n^2 = 3x$ where x is an integer.* Except for the omission of the word "positive," this representation is an appropriate way to begin a proof of this theorem. However, it is very abbreviated, and for mathematics at this level, the argument would be more clearly linked to the theorem by supposing that n^2 is a multiple of 3 before claiming $n^2 = 3x$, where x is an integer.

[2] *Then $3|n^2$.* This inference is correct, but not likely to be useful. It is read "3 divides n^2" and is just another way of saying $n^2 = 3x$, where x is an integer, which was the supposition in [1].

[3] *Since $n^2 = 3x$, $nn = 3x$.* The inference is not wrong, but it is still just the information of [1] written in a slightly different form.

[4] *Thus, $3|n$.* Read "3 divides n," this statement means that n is a multiple of 3. Nothing in [1], [2], or [3] indicates why [4] should be true. Indeed, to fill the gap would amount to proving the theorem.

[5] *Therefore if n^2 is a multiple of* 3, *then n is a multiple of* 3. ∎ This conclusion is just a restatement of the theorem. The argument really ends with [4]. If the reasoning from [1] to [4] were correct (which it is not), [5] would be true. In the genre of mathematical proofs at and beyond the collegiate level, a theorem is not normally restated at the end of an argument purporting to prove it. However, this sort of summary sentence can be found in sample proofs in transition course textbooks, especially when an argument for the contrapositive has been given (Velleman, 1994).

"Proof (d)": The Converse

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. ∎

Analysis of "Proof (d)"

[1] *Let n be a positive integer such that n^2 is a multiple of* 3. This representation and supposition is an appropriate way to begin proving the theorem. Beginning this way, the theorem would be proved if the argument led to *n is a multiple of* 3.

[2] *Then $n = 3m$ where $m \in \mathbf{Z}^+$.* Here $\mathbf{Z}^+$ represents the set of positive integers. Statement [2] is incorrect. The meaning of multiple and [1] yields $n^2 = 3m$, where

$m \in \mathbf{Z}^+$, but [2] does not follow directly from [1] (except by an application of the theorem being proved).

[3] *So $n^2 = (3m)^2 = 9m^2 = 3(3m^2)$.* All three equations are correct, given [2].

[4] *This breaks down into $3m$ times $3m$* This statement follows from [3].

[5] *which shows that m is a multiple of* 3.■ This inference does not follow from [4] or anything else in the argument. The author probably "meant" that n^2, rather than m, is a multiple of 3. When changed, [3] yields that n^2, not n, is a multiple of 3, which means the converse has been proved (by [2], [3], [4], and the altered [5]).

All four of the undergraduate student-generated arguments above correctly use a logical and rhetorical device that is particularly common in the genre of mathematical proofs. That is, to prove a theorem claiming some property about all (each) of the elements in some set, for example, about all positive integers, it is sufficient to argue about a particular, but arbitrary, element of the set. In the language of secondary school algebra, one might say that the argument is made about an unknown constant rather than a variable. Typically, the element is named very early in the argument to avoid the possibility of inadvertently compromising its arbitrariness. At the conclusion of the argument, it is implicit that whichever element of the set might be of interest, the argument could have been about that element. Although this rhetorical device is simple, it is very powerful and simplifies the logic required of validators, thereby perhaps avoiding some errors. In the historical development of calculus, this device together with carefully formulated, analytic definitions (like the epsilon-delta definition of limit) allowed theorems about motion (for centuries a difficult concept) and, in particular, about rate of change, to have proofs in which nothing changes and motion is not mentioned.

The textual analysis above differs from a description of actual validations. It provides background information about the theorem and the four "proofs" useful for understanding validations of them. However, it does not include descriptions of actual validators' mental processes, abilities, missteps, and so on, in their attempts to determine which of the arguments are really proofs. We now describe how some undergraduate mathematics and mathematics education majors judged whether these same four arguments were proofs.

THE EXPLORATORY STUDY OF VALIDATION

Eight undergraduates (four secondary education mathematics majors and four mathematics majors) at a comprehensive state university in the United States were interviewed individually outside of class in the 4th and 5th weeks of a 15-week "bridge" or "transition to proofs" course. Although this course is normally studied in the 2nd or 3rd year at university, three of the eight students were in their 4th year. The eight formed the entire cohort of this learning-to-prove-theorems course required of all mathematics and secondary mathematics education majors at the institution. The goal of such courses, offered by many U.S. mathematics departments, is to help students make the transition from an earlier

traditional calculus sequence to later, more proof-based courses, such as abstract algebra and real analysis. They typically cover various aspects of logic; a little about sets, relations, and functions; proof by mathematical induction; and perhaps a few additional topics, such as elementary number theory or introductory graph theory, about which students are asked to prove theorems. The emphasis is on having students learn to make their own proofs as opposed to the amount of mathematics covered. Kurtz's *Foundations of Abstract Mathematics* (1992), a typical textbook, was used in the course. In addition, students were to spend much of their class time, in pairs, proving theorems at the chalkboard. All eight students had taken single-variable calculus; some had also taken multivariable calculus, differential equations, matrix algebra, or discrete structures. One student (referred to as BH) was concurrently enrolled in a dual-listed graduate/undergraduate number theory course.

The Structure of the Interviews

Each student was interviewed individually for about 1 hour in a semistructured manner using an interview script consisting of four phases.

Phase 1. During this phase, which could be termed the warm-up phase, each student was given a brief fact sheet (see Appendix 1) and this written statement: For any positive integer n, if n^2 is divisible by 3, then n is divisible by 3. They were asked to explain in their own words what the statement said, to give some examples of it, and to decide whether it was true and how they would know. Finally, they were asked to give a proof of the statement if they could. Two successfully did so. After some time had elapsed, those who could not complete a proof were advised that they need not continue, because proving the theorem was not the point of the interview; rather, they were to judge the correctness of other students' "proofs" of the statement. Students were told that since some of the "proofs" they would see might seem a bit unusual, having attempted a proof would help them appreciate the variety of approaches taken.

Phase 2. Each student was shown the four "proofs," one after the other, and asked to think out loud as they read each one and decided whether it was, or was not, a proof. If it was not a proof, they were to point out which part(s) were problematic. If they could, they were to say where, or in what ways, the purported proof had gone wrong. They were also informed that these "proofs" had been submitted as part of a take-home test in a previous year by transition-course students like themselves. The four "proofs" were presented to the students as they appear in Appendix 2.

They were told to take as long as they liked and seemed not to be hurried. For example, one student (JB) spent 15 minutes reading and rereading the first "proof" before deciding (incorrectly) that it was indeed a proof, except for a minor notational error. Several of the students did not want to commit themselves and were allowed during this phase to say that they were unsure whether a given argument was, or was not, a proof.

Phase 3. Having seen and thought about all four "proofs" one after the other, the students were then given an opportunity to reread them all together and rethink their earlier decisions with an opportunity to change their minds, and some did. At the end of this phase, they were not allowed to say that they were unsure, but were asked to make a decision and declare whether each argument was, or was not, a proof.

Phase 4. Finally, the students were asked eight general questions about proofs and how they read, understand, and validate proofs.

The data collected consisted of the audiotaped interviews, the interviewer's notes, and the students' work on the "proof" sheets provided during the interview. (See Appendix 2 for details.)

The Conduct of the Interviews

The interviews were treated as "teaching interviews." Not only was the interviewer the instructor for the transition course that the students were taking, it also seemed appropriate and opportune to the interviewer-instructor to answer their implicit and explicit questions. The interviewer would respond generally about the structure or style of written proofs but never indicate whether the four specific arguments that the students were examining were actually proofs of the theorem. This procedure kept the tone of the interviews as relaxed and conversational as possible.

In addition, as the interviews proceeded, the students sometimes seemed to need encouragement to keep reading and thinking about the four "proofs." Thus, to obtain as much information as possible, the interviewer intermittently made comments and answered or asked questions to keep students on task. All the students seemed to take the interviews seriously and seemed genuinely engaged with the task at hand.

Informing students of mathematical practices and norms. Keeping the students on task occasionally had the effect of suggesting to them practices and norms of the wider mathematics community. In the following excerpt, student MM has just begun to consider "Proof (a)" for the first time. She examines it silently for 1 minute, then asks whether another 3 should be in the expression $3n(n + 2) + 1$. With encouragement from the interviewer, she makes her suggested correction, writing $9n^2 + 6n + 1 = 3n(3n + 2) + 1$. She then continues reading silently for an additional 30 seconds. Then, the following interchange occurs.

MM.   I don't like this one 'cause it's confusing [laughs].

I.   OK, you don't like it. In addition to being confusing and not liking it–which is a very good reason not to like it and which is why I'm going to insist you all [meaning the students in the transition class] write nonconfusing proofs [both laugh]. OK?

MM.   There's an error.

I: Ahh, there's an error there. . . . Let's pretend the student meant to write that [indicating MM's correction $9n^2 + 6n + 1 = 3n(3n + 2) + 1$]. . . .

MM: OK.

I: . . . and see if that makes a proof.

The norm[12] being suggested here is that stylistic clarity (i.e., writing proofs in a nonconfusing way) is valued, but that lack of clarity in itself does not make an argument incorrect. In addition, it is implied through the overt act of correcting an algebraic error that small errors in written proofs (e.g., those similar to misprints) can be fixed during the validation process.

In the following example, student ST is in the midst of considering "Proof (a)."

ST: They also have $n^2 = (3n)^2$. They use the same variable for everything so it kind of throws (?) everything together.

[ST and the interviewer agree to fix this by writing $n^2 = (3m)^2$.]

ST: I think that's what, . . . that's what, . . . threw me off at the beginning because I was looking at all those n's trying to equal each other.

I: That's really hard, isn't it? There's a principle in there that you don't, uh, use the same letter to mean two different things in one proof. You can finish a proof—and then the idea is finished—and then use the letter to mean something new, but that's because there's a break in the ideas.

In this excerpt, ST comments on the confusion caused by using n in several different ways. The interviewer corroborates this and gives an additional bit of information regarding when it is permissible to use the same letter, such as n, again. This, and other similar comments made during the interviews, had the effect of introducing these students to some stylistic features of written proofs at the same time as the students saw (for themselves) a rationale for those bits of mathematical practice.

Further reflection encouraged by the interviewer's remarks. All students were kept on task by the interviewer's deliberate attempts to probe their thinking. We illustrate this procedure with comments from KC, a particularly expressive and responsive student. She told the interviewer that she ordinarily likes to read proofs aloud to herself, that is, she had no problem converting inner speech to regular speech during the interview. Below is a condensed version of her first attempt at validating "Proof (d)." It proceeded along a tortuous sequence of meandering stages, numbered 1–8 below, apparently assisted by the interviewer's probing questions. (For the entire transcript, see Appendix 3.)

1. KC reads the entire proof aloud slowly once and remarks that "Proof (d)" looks a lot like "Proof (a)" and says, "I think that's a proof."

2. On being questioned by the interviewer who is genuinely surprised by this analogy, she explains something about $3(3m^2)$ and tries to rewrite that portion of the proof, concentrating on the form she wants to see, "three times some integer."

3. After some time, the interviewer asks whether that is good enough for a proof

---

[12] This kind of norm (of the wider mathematical community) is reminiscent of sociomathematical norms, the negotiation of which has been investigated in classroom settings, for example, at the undergraduate level by Yackel, Rasmussen, and King (2000). It would be interesting to investigate how such norms are maintained and, in particular, adopted by novices in the wider mathematical community. The kinds of discussions associated with these interviews may not occur often enough in university courses to explain the maintenance of such norms.

now, and KC comments that her wording may not be right. She and the interviewer discuss this at some length and agree there are lots of equivalent ways to word things.

4. The interviewer then asks whether she has finished, and KC responds with an explanation, ". . . Let's see, they showed that *m* was a multiple of 3 but they didn't show *n* was a multiple of 3."

5. She tries to substitute back in and then suddenly blurts out, "OK, wait, wait . . . ," followed by 30 seconds of intense silent concentration. She then thumps her pencil twice emphatically on the desk and asserts, "I don't think that's a proof."

6. The interviewer asks why she changed her mind, and KC says, "They showed that *m* was a multiple of 3, but they didn't show *n* was a multiple of 3."

7. The interviewer then asks, "Where did that *m* come from anyhow?" to which KC responds, ". . . they let $n = 3m$ right up here . . ." [looking back to the beginning of the argument], and notes that they did not start in the right place, that is, $n^2 = 3m$.

8. She reiterates, "I don't think that's a proof."

In responding to the interviewer at stages (2), (3), (4), (6), and (7), KC is encouraged to reflect further on "Proof (d)" and notices features of the argument that went undetected earlier. However, it is not easy to conjecture what happened to KC during the 30 seconds of intense silent concentration at stage (5). Her insight cannot reasonably be attributed to anything specific that the interviewer did; she appears to have had something like a genuine "Aha!" experience (Barnes, 2000).

It would be very interesting to know what KC was conscious of during this experience. We suspect such an experience might be described immediately after its occurrence, but is often not well remembered later. The interviewer did question KC immediately about why she changed her mind, but not about the contents of her conscious experience. To have done so might have considerably altered the overall validation process.

Overall Results

On analyzing the data, it became clear that most students made a judgment regarding the correctness of each "proof" at four different times, namely, at the beginning and at the end of each of two readings (in Phases 2 and 3). These times were labeled Time 1 through Time 4. All students, except JB, maintained or increased the number of their correct judgments from Time 1 to Time 4. (See Figure 1, which displays the changing number of correct judgments by each student across time.) For example, KC gradually improved from one correct judgment at Time 1 to four correct judgments at Time 4. This improvement appears to be the result of further reflection.

Eight students were interviewed regarding the four "proofs," making a total of 32 person-proof judgments. Table 1 gives the percentage of correct judgments made at the various times.
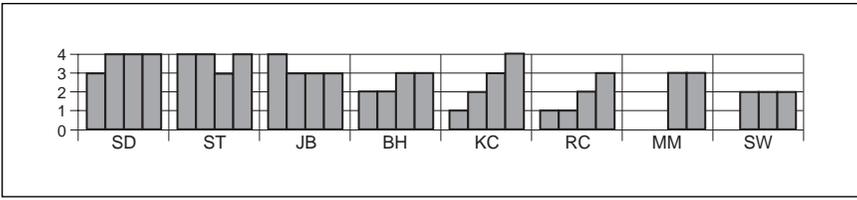
Figure 1. The number of each student's correct judgments across time (Time 1 to Time 4)

Table 1
Percentage and Number of Judgments Over Time

	Time 1	Time 2	Time 3	Time 4
Correct	46% (15)	56% (18)	72% (23)	81% (26)
Incorrect	25% (8)	28% (9)	22% (7)	18% (6)
Unsure	28% (9)	16% (5)	6% (2)	0% (0)

Note. The total number of judgments was 32.

What happened over time that might have caused students to change their minds? During the second phase of the interview, between Time 1 and Time 2, the interviewer asked and answered questions, encouraging the student validators to reflect further. By the beginning of the third phase of the interview, they had seen and pondered all four "proofs" and were more experienced. At Time 4, the interviewer would no longer accept "unsure," and the students made their final judgment for each "proof."

At Time 1, less than half (46%) of the students' judgments were correct. This remarkable fact suggests that given this task on a test, where no one would have been encouraging them to reflect further, the students would probably have done about as well by chance.[13] They initially made the most correct judgments on "Proof (b)," with five correctly stating that it was a proof and three being unsure. They also initially made the fewest correct judgments on "Proof (d)," with only two making the correct judgment and four incorrectly stating it was a proof. This difficulty with "Proof (d)" supports our observation below that the students were primarily checking local details instead of looking for global/structural difficulties. After the first sentence, "Proof (d)" can be regarded as an argument for the converse, modulo one notational error. The students also seemed to be relying on their feelings of understanding in a way that sometimes misled them. Student MM

---

[13] Although the students' ability to determine whether these arguments were proofs appeared to be very limited, we are not claiming this is unchangeable. Indeed, we saw some evidence of improved judgments even during the course of the interviews and conjecture that the students' initial lack of ability was largely due to a lack of appropriate experiences.

commented about "Proof (d)" that she "liked it" and could "see it better than the rest." Table 2 shows how the aggregated students' judgments on the four purported proofs changed over time. The entries represent numbers of students.

Table 2
Changes in the Pattern of Judgments

	"Proof (a)"	"Proof (b)"	"Proof (c)"	"Proof (d)"
Correct	4 3 5 5	5 6 7 8	4 4 6 6	2 5 5 7
Incorrect	3 4 3 3	0 1 0 0	1 1 1 2	4 3 3 1
Unsure	1 1 0 0	3 1 1 0	3 3 1 0	2 0 0 0

Note. Entries give the number of students responding "correct," "incorrect," or "unsure" at each of the four times. For example, under "Proof (a)" for Correct, there were 4, 3, 5, and 5 correct judgments at Time 1, Time 2, Time 3, and Time 4, respectively.

A clear trend can be seen for "Proof (b)" and "Proof (d)"; when comparing Time 1 with Time 4, all students' judgments improved or stayed the same. However, the data for "Proof (a)" and "Proof (c)" are more complex. For example, on "Proof (a)" the improvement from four correct judgments at Time 1 to five at Time 4 resulted from two students changing to correct judgments and one student changing to an incorrect judgment.

In general, the students considerably improved their judgments over time (even though this was neither the purpose nor the expectation of the interviewer). Furthermore, this improvement seems largely to have been due to interactions with the interviewer that somehow enhanced the way that these students used the experiences and knowledge that they brought to the task. This phenomenon appears similar to occurrences in the zone of proximal development that were described by Vygotsky (1978) in studying the sociocognitive development of children. It would be interesting to know more about such student-interviewer interactions when the interviewer's purpose *was* to enhance a student's mathematical practices. For example, small portions of these interviews might be seen as dialectical situations that allowed the interviewer to suggest certain aspects of the practices and norms of the wider mathematical community. In other portions of the interviews, a small comment or question apparently induced a student to refocus his or her attention on another part of the task (i.e., on another part of the argument being validated).

What the Students Said as They Carried Out Their Validations

Most of the errors detected were of a local/detailed nature rather than a global/structural nature. For example, in validating "Proof (a)," five students (JB, KC, SD, BH, ST) commented on notational errors like $n^2 = 9n^2$. Four (KC, SD, MM, ST) noted that odd/even was not correctly expressed in the symbolism, for example, $3n + 1$ for odd. And two (BH, ST) observed that divisibility by 3 could not be concluded in the odd case from the expression $3n(n + 2) + 1$.

Three students (KC, RC, MM) queried the interviewer regarding the use of odd and even cases in "Proof (a)," but they were willing to continue their validation efforts after mutually agreeing that, although this particular breakdown into cases might be peculiar, nothing was wrong with it logically. For example, student RC read "Proof (a)" silently for 1 minute and then the following interchange occurred.

RC: I didn't. . . . It looks like a proof. The only thing I'm questioning is how can they say it's odd or even. Can you do that? [It] would seem that. . . . It seems like it would take care of both cases. . . . The positive integers. . . .

I: We were just beginning to talk about that in class. . . . If you feel you want to do something by cases in a proof, as long as the cases cover everything, it's OK.

RC: OK.

I: Sometimes like this, they [meaning the odd and even cases] are nonoverlapping. But they can even be overlapping. It doesn't hurt to prove something twice as long as the union of the cases covers everything.

RC: It's kind of tricky. I would never have thought of that. Whatever's given me, I try to prove it.

[RC continues to consider the proof for some time.]

I: OK so, uh, if it's OK to do odd and even, do you agree with what they've done? Is that a proof?

RC: It looks like one. [He then examines "Proof (b)."]

Student BH, who had successfully proved the theorem earlier in the interview after a false start, thought that the conclusion in "Proof (a)" was being assumed in the even case. That is, she apparently ignored [4] and wrongly interpreted [5] to include the implicit supposition that $n = 3m$. On reading further, she concluded that the converse had been proved. She stated " . . . [they're] doing it the way I wanted to," suggesting that her initial failed attempt to prove the result influenced her to incorrectly see her own misstep in the argument. Only BH and ST had succeeded in proving the theorem earlier in the interview; they were also the only students who correctly stated that the converse had been shown in "Proof (d)." These sorts of structural comments were rare. Thus, as one might expect, the experience of proving a theorem can have positive effects on subsequent validations of other arguments for the same theorem. And, perhaps counterintuitively, such earlier experiences can sometimes have negative effects on validations, as described above when BH seems to have been more influenced by an early false start on a proof than by her later success.

What the Students Said They Did When Reading Proofs

Most students seemed content to attempt a careful line-by-line check to see whether each mathematical assertion followed from previous assertions. When queried (in Phase 4) about what they do when reading a proof, the students said they did many things: They made sure all steps were logical and looked to see that everything was supported. They checked the computations and whether anything was left out. They said, "[I] take it step by step; [I] like to read it **real** slow" and "I go through more than once. . . . See if from step to step, it follows." Several

students said that they would try to go through a proof with an example, but few gave evidence of doing so during their validations.

KC, whose validation of "Proof (d)" was paraphrased above, said, "[It's] not like reading a newspaper; more understanding is involved. I go through more than once. Something will come out at me that didn't the first time." When queried about how she tells that a proof is not correct, KC explained, "because they didn't show it and they didn't say it [meaning the conclusion]." She added, "Plus I expect to see the steps that say, that lead you up to n is . . . or whatever." She further explained, "And also if you find something that goes, . . . that, that, . . . that is definitely wrong in the steps above. Like the one up here [pointing to one of the four "proofs"]. So that's not . . . [a proof]." Given KC's limited experience with validating proofs, other than those in textbooks, this final remark about finding something definitely wrong in just one step seems quite reasonable. However, in the odd case of "Proof (a)," there are four mistakes that are extraneous to obtaining n^2 is divisible by 3 from itself. (See the textual analysis above.) In general, the students' self-descriptions sounded thorough and competent.

Making Sense as a Criterion for Determining Correctness of Proofs

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. For example, after initially reading through "Proof (a)" slowly, JB concentrated on the last statement, $n^2 = 9n^2$, reiterating several times that it did not make sense. Then he said, "I'm saying I don't think this is a proof because it [$n^2 = 9n^2$] doesn't make sense." After the interviewer suggested that perhaps the author had meant to write $n^2 = 9m^2$, JB concurred and continued his validation attempt for some time, finally deciding incorrectly it was a proof "after variable renaming." Although the judgment was incorrect, it is clear that JB's main focus was on making sense, perhaps residing in the form of some conscious, but unarticulated, feeling of understanding.

When KC first came to the gap in "Proof (c)," she reread the argument several times and tried to expand on it (on paper). When she could not, she explained her thinking to the interviewer, as follows.

KC: Maybe I just don't understand the concept of . . . I . . . , I understand 3 divides n^2. I understand all that stuff, but I don't . . . I don't . . . I understand getting from here . . . [from] there [$n^2 = 3x$] to there [$nn = 3x$], **but** from here [$nn = 3x$] to here [$3|n$] . . . .

I: Uh, huh. So when you tried to fill it out you couldn't basically?

      [KC agrees and says she could probably give an example, but that is not a proof.]

KC: I still don't think that, . . . that is . . . [a proof]. I don't know. I don't think that is . . . [a proof].

I: So that's not a proof?

KC: But I may decide later . . . [reserving the right to rethink the argument].

At this point, a sense of not understanding makes KC cautious about whether the argument is a proof. Another student, MM, in summing up her final judgments, wrote under "Proof (c)" that it was not a proof and "I don't understand the reasoning." Then, under "Proof (d)," she wrote an asterisk followed by "Proof—this one is the easiest for me to understand." However, this argument is not a proof of the theorem, but rather a proof of its converse. In her case, a feeling of understanding did not mean the argument was correctly linked (via a proof framework) to the theorem.

The students also indicated (in Phase 4) that they usually read proofs for understanding. Student BH, who was taking number theory concurrently, said "It's got to be true but got to make sense. Last night I was lost 20 minutes in number theory. I found somebody's thesis in the library. Just one step [was missing]. . . . " This comment suggests that the feeling of making sense, or of understanding, is a matter of considerable importance to BH—important enough to search through theses in the library. Student KC, commenting on her previous calculus textbook, said, "Some of the proofs in the calc book were 'off the wall.' I had to sit and write them up; [I had to] go home and see where the things were coming from."

For these students, understanding was required of them when reading proofs. When given a proof by an instructor or in a textbook, they quite reasonably assumed that it was correct. Indeed, they probably had very little previous practice in deciding whether "proofs" are correct, that is, in validation. Thus, they appear to have depended on their wider personal experiences of understanding, which did not always serve them well in validation. This unreliability may occur because a feeling of understanding itself and one's tendency to rely on it are influenced and sharpened by experiences in validation—experiences these students largely lacked. Also, although the students spoke of understanding quite freely, they apparently did not always mean more or less the same thing by it. Their responses to questions in Phase 4, together with their validation attempts, suggest that they had a range of meanings for understanding from verifying short computations in a text using their own procedural knowledge to a feeling of understanding associated with a gestalt view of an entire argument. Perhaps this variation is due to a general difficulty in negotiating meaning for concepts such as understanding, which do not refer to underlying physical objects (e.g., cars or basketballs).

CONCLUSIONS AND IMPLICATIONS FOR TEACHING

What students say about how they read proofs seems to be a poor indicator of whether they can actually validate proofs with reasonable reliability. They tend to "talk a good line." They say that they check proofs step by step, follow arguments logically, generate examples, and make sure the ideas in a proof make sense. However, their first reading judgments (see Table 1, Time 1) yield no better than chance results, suggesting that they cannot reliably implement their intentions.

On the other hand, even without explicit instruction, the reflection and reconsideration engendered by the interview process eventually yielded 81% correct

judgments (see Table 1, Time 4). This improvement suggests that instruction in validation could be effective. The interviews also suggest that students should be encouraged to attend more to possible global/structural errors, for example, proving the converse of the statement. In particular, students need to devote attention to what we have called proof frameworks (Selden & Selden, 1995), a topic that we suspect is relatively easy to master because it does not depend on a deep understanding of the mathematical concepts involved in the statement of a theorem.

Validation of proofs is part of the implicit curriculum, but it is a largely invisible mental process. Few university teachers try to teach it explicitly, although some may admonish students to "read with pencil and paper in hand" or the like. This advice is at best descriptive, but certainly not usefully prescriptive, so students tend to interpret such vague directions idiosyncratically, spotting mainly local notational and computational difficulties. They appear to be substituting their feeling of overall understanding, plus a focus on surface features, for validation. We suspect such a feeling of overall understanding is useful and reliable for mathematicians who will have included in it their notion of correctness, but is often unreliable and misleading for students.

"Bridge" courses in the United States tend to give students some idea of the logical structure of theorems, covering such topics as direct and indirect proofs and indicating that the converse is not logically equivalent to a theorem. These courses also cover some logic, set theory, functions, and assorted other topics. Often, the material on logic precedes student work with proofs, and perhaps for that reason, tends to be presented in an abstract, decontextualized way that may not be very effective. Textbooks for such courses have from none to just a few to a moderate number of exercises that involve the critiquing of "proofs." The directions vary; students may be asked to (a) find the fallacy in a "proof"; (b) tell whether a "proof" is correct; (c) grade a "proof," A for correct, C for partially correct, or F; or (d) evaluate both a "proof" and a "counterexample." Most of these "proofs" have been carefully constructed by textbook authors so there is just one error to detect. For example, in Smith, Eggen, and St. Andre (1990; p. 39) we find the following "proof to grade":

Suppose m is an integer.

Claim. If m^2 is odd, then m is odd.

"Proof." Assume m is odd. Then $m = 2k + 1$ for some integer k. Therefore, $m^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$, which is odd. Therefore, if m^2 is odd, then m is odd. ∎

The "proof" above does not present a student with anywhere near the complexity or kinds of difficulties seen in our student-generated arguments.[14] Consideration of such contrived "proofs" is no doubt valuable practice for those beginning to learn validation. However, we suggest that additional practice in validating such actual

---

[14] The Smith, Eggen, and St. Andre textbook (1990), and subsequent editions thereof, does contain more complex "proofs to grade," however, many transition course textbooks do not.

student "proofs" as those in this study, together with small-group discussion, could be beneficial, especially for preservice secondary teachers who one day may need to judge the correctness of their own students' proofs or novel solutions to problems.

FUTURE RESEARCH QUESTIONS

One might consider the students in our empirical study as novices at validation and ask what experts—advanced undergraduates, graduate students, and mathematicians—would do. The students in our study could not reliably validate relatively short purported proofs of a simple theorem, but mathematicians can. At some point, therefore, mathematics students learn to do so. When does this happen, and through what kinds of experiences is the process of validation learned?

Proofs of the theorem used in this study do not seem to have a very complex structure. What would happen if one considered a theorem whose proof was inherently more complex or where the potential for more subtle errors was greater? Students often have views of proof (Harel & Sowder, 1998) that differ markedly from mathematicians' views. How would such views be influenced by attempts to increase students' ability to validate proofs?

How does the ability to validate proofs relate to the ability to construct them? How does it relate to a knowledge of logic, especially when taught in the rather formal, symbolic, decontextualized way that is found in many transition courses? What are students' perceived and actual criteria for correctness, for example, that they agree the calculations are legitimate, that they can follow the argument line by line, that they can see why the argument begins where it does, and so on? The students in this study were sometimes misled by a feeling of overall understanding, whereas mathematicians can use such feelings beneficially. What accounts for the difference? How does one learn to have a feeling at the "right time" and not otherwise? What parts of validations consist of reading plus assent, without additional conscious experiences, such as inner speech or vision or feeling? What kind of teaching might promote a correct use of assent during validation?

Here, after introducing the idea of validation, we discussed its nature in a way that depended on information we take to be common knowledge among most mathematicians. However, without additional investigations, there is no way to be sure that we are right about what is common knowledge or that the common knowledge itself is correct. Taking much of this discussion as partly conjectural yields a number of additional research questions. For example, suppose one recorded what was being attended to during the construction of a proof and extracted, in order, those parts that appeared in its final form. Would this order, as we have suggested, differ greatly from the order of the final proof? In teaching, one of us has sometimes encouraged students to write a proof framework, that is, the beginning and ending few lines of a proof, before attempting the middle. Some students resist this advice, and it would be interesting to investigate whether such students have more difficulty constructing proofs than others.

We have also alluded to mathematicians' ability to agree on the correctness of proofs and the truth of theorems and that they can often decide this individually and believe "you do not vote on the truth of theorems." It would be interesting to investigate the degree to which these are the perceptions of mathematicians as well as the degree to which those perceptions are accurate. Information about such perceptions and abilities might enhance the role played by validation in social constructivist views of mathematics.

Last, what is a good way to teach validation? Is it a concept like proof or definition, which is perhaps best learned through experiences, such as finding and discussing errors in actual student "proofs"?

REFERENCES

Baddeley, A. (1995). Working memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 755–764). Cambridge, MA: MIT Press.

Bagchi, A., & Wells, C. (1998). Varieties of mathematical prose. *PRIMUS*, *8*, 116–136.

Barnes, M. (2000). 'Magical' moments in mathematics: Insights into the process of coming to know. *For the Learning of Mathematics*, *20*(1), 33–43.

Bogdan, D., & Straw, S. B., (Eds.). (1990). *Beyond communication: Reading comprehension and criticism*. Portsmouth, NH: Heinemann.

Burton, L., & Morgan, C. (2000). Mathematicians writing. *Journal for Research in Mathematics Education*, *31*, 429–453.

Chomsky, N. (1957). *Syntactic structures*. The Hague, Netherlands: Mouton.

Cuoco, A. (2001). Mathematics for teaching. *Notices of the American Mathematical Society*, *48*, 168–174.

De Villiers, M. (1998). To teach definitions in geometry or teach to define? In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 248–255). Stellenbosch, South Africa: University of Stellenbosch.

Dubinsky, E., & Yiparaki, O. (2000). On student understanding of AE and EA quantification. In E. Dubinsky, A. H. Schoenfeld, & J. Kaput (Eds.), *Issues in mathematics education: Vol. 8. Research in collegiate mathematics education. IV* (pp. 239–289). Providence, RI: American Mathematical Society.

Edwards, B. (1997). An undergraduate student's understanding and use of mathematical definitions in real analysis. In J. A. Dossey, J. O. Swafford, M. Parmantie, & A. E. Dossey (Eds.), *Proceedings of the nineteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 17–22). Columbus, OH: The ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Ernest, P. (1998). *Social constructivism as a philosophy of mathematics*. Albany, NY: State University of New York Press.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals of linguistic theory* (pp. 1–88). New York: Holt, Rinehart, and Winston.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, The Netherlands: D. Reidel.

Halliday, M. A. K. (1977). *Explorations in the functions of language*. New York: Elsevier.

Hanna, G. (1989). Proofs that prove and proofs that explain. In G. Vergnaud, J. Rogalski, & M. Artigue (Eds.), *Proceedings of the thirteenth conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 45–51). Paris: CNRS—Paris V.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Issues in mathematics education: Vol. 7. Research in collegiate mathematics education. III* (pp. 234–283). Providence, RI: American Mathematical Society.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kurtz, D.C. (1992). *Foundations of abstract mathematics.* New York: McGraw-Hill.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery.* Cambridge, UK: Cambridge University Press.

Leron, U. (1983). Structuring mathematical proofs. *The American Mathematical Monthly*, *90*, 174–184.

Morgan, C. (1996). The language of mathematics: Towards a critical analysis of mathematics texts. *For the Learning of Mathematics*, *16*(3), 2–10.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Nystrand, M., Doyle, A., & Himley, M. (1986). A critical examination of the doctrine of autonomous texts. In M. Nystrand with M. Himley & A. Doyle (Eds.), *The structure of written communication: Studies in the reciprocity between writers and readers* (pp. 81–109). Orlando, FL: Academic Press.

Olson, D. R. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review*, 47, 257–281.

Pearson, P. D. (2000). Reading in the twentieth century. In T. L. Good & M. Early (Eds.), *American education: Yesterday, today, and tomorrow* (Ninety-ninth yearbook of the National Society for the Study of Education, Part 2, pp. 152–208). Chicago: University of Chicago Press.

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation* (J. Wilkinson & P. Weaver, Trans.). Notre Dame, IN: University of Notre Dame Press. (Original work published 1958)

Pimm, D. (1984). Who is 'we'? *Mathematics Teaching*, *107*, 39–42.

Rasmussen, C., & Zandieh, M. (2000). Defining as a mathematical activity: A realistic mathematics analysis. In M. L. Fernández (Ed.), *Proceedings of the 22nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 301–305). Columbus, OH: The ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

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.

Rowland, T. (2002). Generic proofs in number theory. In S. R. Campbell & R. Zazkis (Eds.), *Learning and teaching number theory: Research in cognition and instruction* (pp. 157–183). Westport, CT: Ablex Publishing.

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

Selden, A., & Selden, J. (1987). Errors and misconceptions in college level theorem proving. In J. D. Novak (Ed.), *Proceedings of the second international seminar on misconceptions and educational strategies in science and mathematics* (Vol. III, pp. 457–470). Ithaca, NY: Cornell University.

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

Selden, A., Selden, J., Hauk, S., & Mason, A. (2000). Why can't calculus students access their knowledge to solve non-routine problems? In A. H. Schoenfeld., J. Kaput, & E. Dubinsky, (Eds.), *Issues in mathematics education: Vol. 8. Research in collegiate mathematics education. IV* (pp. 128–153). Providence, RI: American Mathematical Society.

Sierpinska, A., Defence, A., Khatcherian, T., & Saldanha, L. (1997). A propos de trois modes de raisonnement en algèbre linéaire [Three modes of reasoning in linear algebra]. In J.-L. Dorier (Ed.), *L'enseignement de l'algèbre linéaire en question* (pp. 259–268). Grenoble, France: La Pensée Sauvage.

Smith, D., Eggen, M., & St. Andre, R. (1990). *A transition to advanced mathematics* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Tall, D., & Vinner, S. (1981). Concept image and concept definition with particular reference to limits and continuity. *Educational Studies in Mathematics*, *12*, 151–169.

Thompson, D. R. (1996). Learning and teaching indirect proof. *Mathematics Teacher*, *89*, 474–482.

Toulmin, S. (1958). *The uses of argument.* Cambridge, UK: Cambridge University Press.

Vachek, J. (1966). *The linguistic school of Prague: An introduction to its theory and practice*. Bloomington, IN: Indiana University Press.

Velleman, D. J. (1994). *How to prove it: A structured approach*. Cambridge, UK: Cambridge University Press.

Vinner, S. (1997). From intuition to inhibition—Mathematics, education, and other endangered species. In E. Pehkonen (Ed.), *Proceedings of the 21st conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 63–78). Helsinki, Finland: University of Helsinki, Department of Teacher Education.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Yackel, E., Rasmussen, C., & King, K. (2000). Social and sociomathematical norms in an advanced undergraduate mathematics course. *Journal of Mathematical Behavior*, *19*, 275–287.

Authors

Annie Selden, Tennessee Technological University, Mathematics Department, Box 5054, Cookeville, TN 38505; selden@tntech.edu

John Selden, Mathematics Education Resources Company, 1015 Melrose Drive, Cookeville, TN 38501; js9484@usit.net

APPENDIX 1

Facts About the Positive Integers

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 $3n + 2$.

APPENDIX 2

Interview Data Collection Phases

PHASE 1: Warm-Up Exercises

For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.
    1. Explain, in your own words, what the above statement says.
    2. Give some examples of the above statement.

3. Does the above statement seem to be true? How do you tell?
4. Do you think you could give a proof of the above statement?

PHASE 2: Sequential consideration of 'Proofs' (a), (b), (c), (d).

PHASE 3: 'Recap' on the 'Proofs'
Below are several purported proofs of the following statement:
  For any positive integer n, if n^2 is a multiple of 3, *then n is a multiple of* 3.
For each one, decide whether or not it is a proof. Try to "think out loud" so you
can let me in on your decision process. If it is not a proof, point out which part(s)
are problematic. If you can, say where, or in what ways, the purported proof has
gone wrong.

(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 possi-
bility, $3k + 2$. 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. ∎

(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. ∎

PHASE 4. Final Questions
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?

APPENDIX 3

Transcript of KC Validating "Proof (d)"

[Note: Segmentation into stages is our own. Emphases in boldface indicate portions of the transcript discussed earlier in the condensed version.]
Initial validation attempt in Phase 2

Stage 1

I: OK, this is the last one. We have four of these.

KC: OK. [**Reads theorem aloud slowly once.** There's a pause at the end of the first sentence, after which she says "OK" and goes on reading.] Yah. That, that looks a lot like the one that we just did.

I: Ahh, which means it's good or it's not good?

KC: Yah. No, not just did. I mean this one over here [pointing to the paper with (a)].

I: Oh, it looks like. . . . [Spreads out the four papers with the 'proofs' on them.] We had (a), (b), (c), (d).

KC: Like that one. Like that one. [Pointing to (a).]

I: It looks like. . . . It looks a lot like the, ah, (a) one.

KC: Yah. Yah, **I think that's a proof.**

Stage 2

I: OK. OK. **You think this one's a proof.** Ookay.

KC: **'Cause, see, 'cause this is what I was talking about by 3 times $3m^2$.**

I: Ah, OK.

KC: Then you could let this $3m^2$ be some integer k.

I: OK. So you want to see this displayed like that?

KC: Yah.

I: OK.

KC: Then you could say that like…

I: Well, I don't want to have it upside down [turns paper around].

KC: You could also . . . say . . . this breaks down. Well, this one. . . . What I would do, is say, uhm, uhm 3 . . . then, $3m^2$ equals to k, some integer k such that $3k$, or something like, equals to $3k$, uhm. . . .

I: [Interrupting] Why don't you write that down? Your talking here.

KC: OK.

I: You need your pencil then.

KC: Like here, I probably would say [writes and talks more slowly]. Let $3m^2$ equal to k, for some positive integer k . . . [pause] . . . integer k. OK. Uhm. [long pause] Then 3 times $3m^2$ would equal to $3k$ . . . [pause] which . . . [pause] makes, uhm, or I don't know the wording may have to be wrong, uhm, which makes, uhm . . . [pause] which is . . . , I don't know how to say it, but that's. . . .

I: That's what you wanted to say.

KC: See that's what I want to say, and then, like, insert it in here. Then you could say that this, by proving that it's **three times some integer**, some. . . . Then you could say that it's now. . . .

I: It's now a multiple of 3. [Finishes sentence for KC.]

KC: Yah, it's now a multiple of 3, OK. [Says and writes] Now n^2 is a multiple of 3.

I: Uh, huh.

KC: I don't know if I'd word it in that way, but that's what I'm thinking.

I: That n^2 is a multiple of 3?

KC: I don't know if I'd word it in that way.

KC: 'Cause then you have it in the form **three times something else. . . .** Because this breaking down to $3m$ times $3m$, uh that, that kind of . . . because you have this already.

I: Uh, hmm.

KC: Do you understand what, what I'm saying?

Stage 3

I: Yah. Yah. I understand you want to get it displayed like that. Uhm, **and that's good enough for a proof now?**

KC: Now, my wording right here probably is wrong , I mean it could be. . . .

I: There's lots of different ways you can word things. . . . [Continues for some time.]

KC: I mean. I mean. I could. . . . A lot of this could be like… Even this could be reworded. You don't have to say "let" or just say, you know, k, uhm, for some integer k equals, or something, or just let this be something else. So you could have it in the form. . . . Or even that's fine. **But, this $3m$ times $3m$ stuff, that's gotta go.**

Stage 4

I: Then, **how do you know you're finished** with? I mean. . . .

KC: [Interrupts] Because, because you have it now, now you have it. . . . [Long pause] Let's see. **They showed that m was a multiple of 3, but they didn't show that n was a multiple of 3.** [Long pause] 'Cause you have to show that n is a multiple of 3.

Stage 5

I: Uh, huh.

KC: Uhmmm. [Long pause] Hmmm. [Long pause] So **you'd have to go back and you'd have to substitute back in here.** [Writes $(3m)(3m) = (3m)n = 3(mn)$.]

I: Uh, huh.

KC: Anyway that. . . . [Long pause] Well, it would. . . . OK, wait. OK, wait, wait. . . . Then you'd have. . . . Let's see . . . [long silence, lasting 30 seconds]. Well, [taps pencil quickly twice, and again quickly twice] . . . [pause] **I don't think that's a proof** [said definitely].

Stage 6

I: Oh! [surprised] Oh, OK. [KC laughs.] **Now why did you change your mind?**

KC: Well, because they didn't, they didn't say . . . that n was . . . they didn't say n was a multiple of 3. They didn't figure that out. **They showed that m was a multiple of 3, but they didn't show that n was . . .** and. . . .

Stage 7

I: [Interrupting] **Where'd that m come from anyhow?**

KC: Well, the m came from… **they let $n = 3m$ right up here** [pointing to top of proof].

I: Yah.

KC: Just to, just to to. . . .

I: That's just like your using k, as far as I can tell.

KC: See, you're, and . . . they're supposed to have let n^2 be a multiple of 3, so it should have been "let $n^2 = 3m$."

Stage 8

I.   Uh, hmm.

KC: So. . . . So, at least I showed you how I was, . . . I would have worded. . . . [Both laugh.] **I don't think that's a proof.**

I: OK this one's not a proof. OK, that's one's not a proof. OK, we can come back and change our minds.

Subsequent "recap" validation attempt in Phase 3

KC: [Reads "proof" aloud, followed by a pause of 18 seconds.] No.

I: Not a proof?

KC: No, because they still, I mean, they didn't define . . . they didn't prove what the n was, which is what they were supposed to prove.

I: So they didn't prove what they were supposed to prove. Which was what?

KC: That, that n is a multiple of 3.

I: Uh, hmm. . . . OK.