Teaching Mathematical Proofs

Robert W. McCreery, student #13879059

Introduction

Proof occupies a unique status within the secondary mathematics curriculum. Where most topics require an ability to digest concepts and algorithms, then to apply them correctly in a given situation, proving assertions requires a secondary level of understanding: a thorough understanding of the concepts and the ability to combine relevant ones into a logical argument (Selden & Selden, 1987).

Proofs, therefore, involve not only solving concrete problems, but also covers abstract problems, and demonstrating conclusively that the solution or argument is correct. Student seldom have an introduction to these skills and techniques, and the conventions for them, before being instructed on the process (Andrew, 2009)

The unique nature of proof therefore demands a broad set of skills and knowledge, and these skills are reflected in every component of the Working Mathematically skill set (NESA, N.D.). The syllabus topics, outcomes, and relevant Working Mathematically categories, are shown in the following table.

The issues, errors, and misconceptions are therefore also unique, and so are the challenges facing both students and teachers.

Table 1: Syllabus Topics and Outcomes, and applicable Working Mathematically categories

NESA Syllabus Topic and Outcome		NESA Working Mathematically categories					
Outcome Code	**Description**	**Communicating**	**Fluency**	**Understanding**	**Justification**	**ProblemSolving**	**Reasoning**
ME-P1: Proof by Mathematical Induction							
ME12-1	applies ***techniques involving proof*** or calculus to model and solve problems			√			√
ME12-6	chooses and uses appropriate technology to solve problems in a range of contexts		√			√	
ME12-7	***evaluates and justifies conclusions***, communicating a position clearly in ***appropriate mathematical forms***	√			√		√
MEX-P1 The Nature of Proof and **MEX-P2 Further Proof by Mathematical Induction**							

MEX12-1	understands and uses different ***representations of numbers and functions*** to model, ***prove results*** and find solutions to problems in a variety of contexts	√					
MEX12-2	chooses appropriate strategies to ***construct arguments and proofs*** in both ***practical and abstract*** settings		√	√			√
MEX12-7	applies various mathematical techniques and concepts to model and ***solve structured, unstructured and multi-step*** problems			√		√	√
MEX12-8	communicates and justifies abstract ideas and relationships using appropriate l***anguage, notation and logical argument***	√			√		

Issues

Literature on the teaching of proof, across various mathematical disciplines, identifies many issues and distinct common themes. These challenges confront both students and teachers; some of the challenges are shared between these groups. Principal issues include the following:

Inadequate prior knowledge

For most students, mathematics is the notion of proofs is relatively alien to them when they first encounter them in class; the ability to write a proof, or to recognise a valid proof, will likely not have been introduced to them yet (Andrew, 2009; Selden & Selden, 1987).

A study by Rice et al. (2016) comparing errors in proofs written by Computer Science majors and Mathematics majors found that, even at an undergraduate level, students with stronger mathematics backgrounds are far less likely than their counterparts to make major errors in their proofs. This study supports the importance of background knowledge, and the mode of thinking required, in composing and writing proofs.

Lack of clarity on characteristics of a good proof

A commonly highlighted issue is the lack of consensus on what constitutes a successful, well-constructed proof (Selden & Selden, 1987) Moreover, teachers and students frequently have different views of what a correct proof is (Andrew, 2009; Selden & Selden, 1987) . Moore (2016) and Andrew (Andrew, 2009) asserted that, while the proof-writing process and conventional forms are typically taught, the expectations (for example, level of detail and notation standards are frequently left unclear.

In two of the studies examined, the authors identified a single attribute to describe the quality that a good proof, within most areas of mathematics, would exhibit. (The other studies framed their evaluation criteria in terms of errors made.)

For Moore (2016), the defining characteristic is *demonstrated understanding*, shown through *correctness*, *clarity*, and *fluency*, in descending order of importance.

Brown and Michel (2010) described the goal as expressing mathematical maturity, through readability, validity, and fluency; the method they propose, RVF, takes its name from the resulting acronym. *Readability* for them comprises coherence, flow, correct syntax/grammar, comprehensibility, logical continuity, and relevance (i.e. every assertion must have a purpose). *Validity* is evinced through appropriate method appropriate, correct calculations, and correct use of logical rules. *Fluency* is shown through the apt use of mathematical language and notation.

Grading and feedback

With the success characteristics being relatively abstract, the evaluation and grading processes also were naturally also seen as difficult; it is this process that prompted most of the studies (Andrew, 2009; Moore, 2016; Selden & Selden, 1987; Strickland & Rand, 2016)

Moore's study (2016) found that some teachers addressed errors as black-and-white and their severity context-independent, while others were interested in the cognitive process that led to the error: did it indicate a slight miscue, or was it a sign of a significant lack of grasp of the concept, careless or deliberate? The consensus in his group was that grading prohibitively time-consuming and detail is difficult to convey in feedback to students; therefore, their feedback often lacked detail beyond marking where an error occurred, or even simply the marks deducted.

Correspondingly, several of the studies noted that students found feedback difficult to action, because it did not contain the necessary information (Andrew, 2009; Selden & Selden, 1987; Strickland & Rand, 2016)

Errors and Misconceptions

According to Moore (2016), proofs are often taught by requiring students to write them, without knowing details of the rubric or evaluation criteria. Students are often unaware of grading criteria, which Moore's study found to be variable; rubric or criteria unclear or not conveyed to students before writing. Therefore, the errors encountered can range widely.

The studies considered here all include classifications of errors (see Figure 4: Strickland/Rand coding matrix and Appendices). Some of them consider errors in relation to the aspects of a quality proof (see Lack of clarity on characteristics of a good proof), which gives a sense of the range. Proof errors may be related, for example to:

- logic (application of rules, or logical flow)
- mathematical usage (e.g. computation)
- fluency (language notation)
- grammar/syntax
- form (the conventional presentation and required components of a proof)

Selden and Selden (1987) addressed Reasoning errors specifically, dividing them into Misconception and Other errors. Their list is in large part specific to set theory, though, and is not discussed further.

However, they identified two distinct pedagogical approaches relating to mathematics, which they asserted is a major cause of misconceptions about proofs.

The first approach, the static view, is easily recognised and prevalent. It involves learning concepts, then algorithms using those concepts. Finally, it leads to implementing the algorithms, usually in specifically defined scenarios. (The second, the Socratic-Moore method, involves sharing materials to study, discussion, and questions, but no lecturing or answers. By its nature, it avoids this misconception, but the resources discussed here may still be useful learning materials in that context.

Students may – even if they do not realise it – think of mathematics in this static framework – learn, apply, execute. Proof being outside that construct, and therefore not what they think of as a mathematical skill, may present significant difficulties. Synthesizing results and solutions from many bits of knowledge may be new to students (Selden & Selden, 1987).

This idea, that proof is not a mathematical exercise, may be the most significant misconception about it. If students are to create algorithms, not just follow them, they will need to validate their creations, they need reasoning, justification, and communication skills (see Table 1).

Resources that could be used to teach

Proofs, having a general structure but otherwise being thought exercises rather than algorithms, lend themselves less to technical applications than to logical processes. To address the prevailing issues associated with teaching and grading proofs, these authors proposed rubrics of their own design. They intended to establish common understanding of expectations, a platform for consistent grading, and a framework for useful feedback. Andrew (2009) proposed two reasons that a consistent framework for evaluation would alleviate misconceptions:

- meaningful, consistent feedback;
- clarity on what affects marks, so students focus attention more productively.

All of the rubrics include feedback that is targeted, but not prescriptive; two of the authors (Andrew, 2009; Moore, 2016) include this trait as a benefit or design aspect of their solution, in that they guide students but leave it to them to identify and correct the exact error.

The rubrics are discussed below, in descending order of simplicity.

RVF

The RVF rubric, Brown and Michel's abstraction of an original objectives-based rubric (2010), reduces and generalises the criteria to the three primary attributes they identified in their study. The authors chose a 0- scale where the criterion is non-Boolean (and 0/1 otherwise), because it is still close to a Boolean in efficiency, but it allows some latitude for considering the three desired attributes and overall considerations.

An example appears below; descriptions of the criteria are in the Appendices.

Figure 1: RVF rubric (Brown & Michel, 2010)

• Author's solution is readable	0	1	2	3
• Author's arguments and calculations are valid	0	1	2	3
• Author evinces fluency		0	1	2

The following version might be more self-explanatory:

Figure 2: RVF rubric update

Criterion	**Met?**
Readability: steps relevant and connected, body is coherent, grammar & syntax	0 1 2 3
Validity: method appropriate, calculations correct, deductions logically correct	0 1 2 3
Fluency: Use and convey understanding of language and notation of maths	0 1 2

Proof Error Evaluation Tool (PEET)

Andrew (2009) developed the Proof Error Evaluation Tool (PEET) to enable quick marking and to enable quality feedback that students can easily interpret and act upon. In contrast to the RVF rubric, it is not intended to generate marks or a grade, but to facilitate discussion. As with the RVF, the output could be used for a secondary purpose (i.e. as input for marking).

The PEET identifies twenty-two errors, classified as Structure (with eight sub-categories) or Understanding (six sub-categories), listed in the Appendices.

When grading the proof, the teacher applies the listed codes in line, giving the student both the location and nature of the error, and can add a description if desired (see example below).

Figure 3: PEET application example (Andrew, 2009)

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

"Proof (d)". PROOF. Let n be a positive integer such that n^2 is a multiple of 3. Then $n = 3m$ where m is a positive integer. 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.

Let n be a positive integer such that n^2 is a multiple of 3.	S4. This statement is not used anywhere in the proof.
Then $n = 3m$ where m is a positive integer.	S7. Is attempting to proof $q \geq p$.
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.	U2. This statement is confusing because it was already shown that $n^2 = 3(3m^2)$, clearly a multiple of 3, but then it is stated that n^2 is a multiple of 3 because it "breaks down into $3m$ times $3m$."

Strickland/Rand

Strickland and Rand (2016), like Andrew, looked to classify errors, rather than enabling marking. This was by design, with the idea that, for example, a teacher would have higher expectations at the end of a term than they would at the start, so relative grades for the same errors might differ. Divorcing the classification from the grade avoided inaccurate or unfair results.

They also recognised that judging an error's severity was subjective, so they devised a qualitative severity classification: logic-related, maths-related, rhetorical (language/notation), or ambiguous. Their rubric, then, categorises errors in two dimensions: the type of error made, and the source of the misunderstanding.

Strickland and Rand's model also benefited from related work; they adapted categories from other models, and they cited Selden & Selden and Andrew, both also included in this paper, acknowledging they adapted much of their work was founded on these same references (Strickland & Rand, 2016)

The matrix below lists the errors and their categories on the left, the severity categories across the top, and examples for various error/severity combinations, is shown below. Explanations for the numbered references are in the Appendices.

Figure 4: Strickland/Rand coding matrix

Error Type		Fundamental	Content	Rhetorical	Ambiguous
Basic Problems - Acategorical	Wrong Problem				
	Wrong Method				
	Scratch Work				
	Other				
Omissions	Error-caused Omission	An error caused large parts of the proof to be omitted. Must be used in conjunction with another code specifying the triggering error. See (1) below.			
	Assertions	Entire result is asserted	Local assertion. See (2) below.	Context requires more details. See (3) below.	
	Omitted Sections	Did not address one or more sections of the proof. (Trailed off halfway, skipped cases, etc.)			
	Local Omission	Omitted a single isolated step			
Importing known material	Misusing Theorem	Ex - applying the converse of a theorem	Misunderstanding known results or theorems. See (4) below.	Misphrasing a known theorem, while using it correctly	
	Vocabulary & Grammar	Rare or nonexistent. See (5) below.	Misuse of definition or vocabulary. See (6) below.	Prose is poorly written or missing, or student fails to follow mathematical conventions	
Flow of proof	Logical Order	Didn't proceed through the proof in a linear fashion, or ideas were not in logical order			
	Circular Argument	Uses conclusion in the course of the proof			
	False Implication	P does not imply Q. See (7) below.			
	Extraneous Detail	Attempted or proved irrelevant results. See (8) below.			
Miscellaneous	Locally Unintelligible	A single statement within the proof is completely unintelligible			
	Proof by Example	Usually this error is Fundamental.	Ex - Student confuses a universal proof for an existential proof.	Ex - Student fails to include "Without loss of generality..."	
	Notation	Notation is awkward/confusing/misleading. See (9) below.			
	Weakened Result	Proving a weaker result (eg - a result that only applies to a subset of the scope of original statement)			
	False Statement	Made a false statement or incorrect computation.		Apparent typo	

Comparison of tools

The RVF approach provides marking criteria to be met, resulting in scores. It is primarily a guide for grading, though it provides specific values to analyse the three abstract factors that inform the grade. It encapsulates the detail that error classification requires – see the other two methods for comparison – into these scores. The method implies a level of sophistication in the works and independence in their writers. The RVF model appears suited to higher levels of work, with students who have some background with proofs.

The PEET and Strickland/Rand approaches are methods of classifying and communicating errors. They do not result in a specific grade, nor are they intended to. However, they provide more direction in terms of the nature of errors – and certainly these systems could provide structure and rationale for applying marks or grades. The preference may depend on the error categories of each method. Of course, the user can always exchange or adapt the error sets to suit their needs.

Considerations for Use

In each case, the first advantage of a rubric is being able to communicate standards and expectations to students before they begin work. While a rubric provides a consistent platform, though, teachers may always interpret or apply the framework slightly differently, or take other factors (such as benefit of the doubt, use of abbreviations, or assumptions that affect relative importance of different attributes of the proof). Discussion of the rubric and assumptions will be informative and provide assurance to students about how to direct their efforts.

Walking through an example, such as the one shown in Figure 3, would be helpful for discussion in class, so that students would see what to expect and understand how they might respond to it. That knowledge, in turn, could save teachers' time in answering questions later.

Rubrics, too, can be refined with use; when difficult questions or issues arise, or new, similar errors appear that do not fall within an existing category, the error rubrics can be adjusted.

Because the PEET and Strickland/Rand models are both for identifying errors, and the RVF model is for grading, teachers may consider using two together, say the PEET for diagnosing, with the results informing an RVF scoring.

All of the approaches give direction, without being prescriptive: a common thread with all of the approaches is the belief that the students will benefit from having to identify and correct the errors themselves – which implies the value of revising the exercises. Strickland and Rand (2016) recommend including a revision/resubmission task at least once in a class, to ensure students benefit from the feedback – but not for every exercise.

Summary/conclusion

Implicit here, and perhaps the most important conclusion, is that the skill of proof composition needs to be taught in a structured way, and it needs to incorporate a meaningful feedback cycle. The syllabus already includes three separate units on proof, so focusing on proof techniques within these units can fit within the existing framework. Some of the various attributes useful in proof can be developed in earlier units, for example in early geometry or work with triangles, junior students can learn to provide explanations or rules for their answers, and the marking rubric can reflect the importance of stating the reasons.

The approaches for employing these rubrics need not involve technology, but they can be implemented using existing technology and productivity tools; platforms such as Google Forms, Microsoft Excel, and the classroom sharing/distribution platforms may provide the best technological context for these improvements.

Appendices

RVF

(Brown & Michel, 2010)

Criteria

Your written responses will be assessed via attention to the following categories, which will be explained below: *validity*, *readability*, and *fluency*. Assume that an equal weight will be applied to each category that is applicable, noting that readability and fluency are always applicable whereas validity may not be.

Readability. Your written work will be examined for readability. Obviously, if your written work is not readable it cannot be assessed, but since the ability to communicate Mathematics is a focal point for this class, special attention will be paid to this quality. Strings of equations with no explanation or motivation do not constitute a readable response, and although a carefully chosen, motivated, and explained figure may be worth 1000 words, 1000 figures in and of themselves will be worth nothing.

Fluency. Mathematics is a concise and precise language and I wish to enhance your fluency with this language. Therefore, part of every assessment will focus on your ability to incorporate correct, established notation and terminology into your written work. An example of a lack of fluency is the use of "top" and "bottom" for the numerator and denominator of a fraction. Another example which is evidence of fluency is using "Let f be a real valued function of a real variable" interchangeably with "Let $f : \mathbb{R} \to \mathbb{R}$". Other things which are definable perhaps only by an exhaustive list of examples will be assessed as well.

Validity. If a solution requires calculations or entails a string of deductions, the extent to which your solution is valid will be analyzed. Validity corresponds to the discernable extent to which your method used is appropriate, your calculations are correct, and your deductions follow the rules of logic.

PEET

(Andrew, 2009)

Error	Category
S1. Introduced variables without defining them or performed operations that were undefined.	Proof setup
S1. The approach taken in proving a statement will not work.	
S1. The proof was to be completed using a specific method, but this method was not used.	
S2. Made a false assumption somewhere in the proof.	Correct assumptions
S3. Didn't proceed through the proof in a linear fashion, and ideas were not in logical order.	Linear/sequential order
S4. The proof contained extraneous details or steps that did not really contribute to the proof.	Stray details/conciseness
S4. The length of the proof was unnecessarily long and thus extremely difficult to follow.	
S5. The write-up was illegible at times, making it difficult to read and/or understand.	Neat presentation

S6. Relied too much on calculator or computer-generated information in one step of the proof.	Technology's place
S7. Needed to show p ≥ q but did not show directly, or by (q) ≥ (p), or by contradiction.	Proof type
S7. Only gave an example to establish the truth of a mathematical statement.	
S8. Used nonstandard or confusing notation.	Correct use of symbols/notation
U1. Wrote a statement that was not justified, explained, or verified.	Sufficient details
U2. Wrote a statement or paragraph that was ambiguous, confusing, and/or unnecessarily complex.	Clarity
U3. Failed to include an illustrative picture that would make the proof easier to understand.	Pictures in the proof
U3. Relied too much on a picture to prove something was true.	
U4. Did not sufficiently justify a crucial step in the proof.	Crucial step/main idea
U4. An error caused important parts of the proof to be left unaddressed.	
U5. Made a false statement or incorrect computation in the proof.	Correct implications and statements
U5. Incorrectly claimed that one statement implied or equaled another statement.	
U6. Included some cases but not others (which were not trivial).	All cases present
U6. Did not address one aspect of the problem.	

Strickland/Rand

(Strickland & Rand, 2016)

Severity Categories

Fundamental Errors	errors that arise from a basic misunderstanding of how logical rules operate or what it means to construct a logical argument.
Content Errors	errors that show a misunderstanding of the mathematical content in a proof.
Rhetorical Errors	errors that reflect unfamiliarity with professional norms and conventions of proof-writing as related to presentation.
Ambiguous Errors	without interviewing the student, the rater feels that they have insufficient information to hypothesize the cause of the error.

Coding Matrix – Explanatory Notes

(Strickland & Rand, 2016)

1. Error-caused Omissions: at times, an early error causes the student to veer off-track and spend quite a lot of energy barking up the wrong tree. In these cases, the rater may feel it is worth noting that any substantial progress towards a correct proof is absent. The code for Error-caused Omissions is unique in that it must be used in conjunction with a second code, which describes the error at onset of the divergent path. The type - Fundamental, Content, Rhetorical, or Ambiguous - would also be

determined according to the error of onset. As a hypothetical example, consider a student who has committed the error “Vocabulary and Grammar” by claiming that the definition of a rational is a number that is even or odd. In their proof, they therefore failed to consider all non-integer rationals, and only explored cases of even and odd integers. The Vocabulary error triggered the Error-Caused Omission.

2. Content Assertion: student fails to realize that a statement requires a logical leap. For example, suppose a student asserts, within a larger proof, that if n2 is even then n is also even. Although the claim is true, they have claimed the assertion as though it is obvious, which it is not.

3. Rhetorical Assertion: student appears to understand the logical leap required to make the statement, but it is appropriate to provide the reader with more details. This is particularly contextual, according to whether the student is supposed to take a fellow student as the intended audience or whether they are learning to write for the professional community.

4. Content Misuse of a Theorem: the error was commonly seen in the linear algebra course, when the material was more advanced than what the students had previously seen. An example would be a student who cites a theorem, where that theorem’s conditions do not actually apply to the context.

5. Fundamental Vocabulary and Grammar: we do not find Fundamental Vocabulary and Grammar errors to be a useful category. The concept of having definitions or grammar itself is elementary. When students misuse fundamental mathematical words like “implies,” “therefore” or “since,” it is doubtful that they are actually wrestling with the word itself. More often, they are either struggling with the parts of an implication or with the phrasing of an implication. The former is better described by the category of Flow of Proof, while the latter is better described as Rhetorical error.

6. Content Vocabulary and Grammar: an example, again from linear algebra, is a student who is asked to prove that a matrix is non-singular. In the course of the proof, the student reveals deep misunderstandings about what it means for a matrix to be non-singular.

7. False Implication: False Implication must be distinguished from the error False Statement. If a student asserts that P implies Q, the code False Implication applies only to the implication itself, not the accuracy of P or Q. Often we applied this code to implications where both P and Q are necessary statements for the proof, but Q does not follow from P.

8. Fundamental Extraneous Detail: a student may commit this error if they do not appear to understand the question, and are just throwing all possible details at the wall in hopes that something sticks. For Content and Rhetorical versions of this error, the proof may be correct as stated. Nevertheless, community norms of proof-writing call for mathematicians to cut out useless tangents and details, and so addressing extraneous detail is an aspect of teaching students to write proofs.

9. Notation: elsewhere, we mentioned the heuristic “would a mathematician commit this error?” for determining which errors are Content errors. This heuristic fails to apply to Notation, however. Generally mathematicians have mastered all conventional forms of notation, but there is plenty of content with which a student can wrestle.

10. False Statement: although all errors are technically False Statements, this is reserved for errors which are not otherwise described. Many of these are computational errors, but this category also serves as a general catch-all.

References

Andrew, L. (2009). Creating a Proof Error Evaluation Tool for Use in the Grading of Student-Generated "Proofs". *PRIMUS : Problems, Resources, and Issues in Mathematics Undergraduate Studies, 19*(5), 447-462. https://10.1080/10511970701765070

Brown, D. E., & Michel, S. (2010). Assessing proofs with rubrics: the RVF method. *Proceedings of the 13th Annual Conference on Research in Undergraduate Mathematics Education,* http://sigmaa.maa.org/rume/crume2010/Archive/Brown_D.pdf

Moore, R. C. (2016). Mathematics Professors' Evaluation of Students' Proofs: A Complex Teaching Practice. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 246-278. https://10.1007/s40753-016-0029-y

NESA. (2017a). *Mathematics extension 1 stage 6 syllabus (2017).* NSW Government Education Standards Authority (NESA). https://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/stage-6-learning-areas/stage-6-mathematics/mathematics-extension-1-2017

NESA. (2017b). *Mathematics extension 2 stage 6 syllabus (2017).* NSW Government Education Standards Authority (NESA). https://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/stage-6-learning-areas/stage-6-mathematics/mathematics-extension-2-2017

NESA. (N.D.). *Working Mathematically.* NSW Government Education Standards Authority (NESA). https://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/stage-6-learning-areas/stage-6-mathematics/mathematics-life-skills-2017/working-mathematically/

Rice, J., Wheeler, A., Edwards, D., & Marshall, D. (2016). *Predicting students' approach to knowledge by analyzing misconceptions and errors students make writing proofs*. ProQuest Dissertations Publishing.

Selden, A., & Selden, J. (1987). *Errors and Misconceptions in College Level Theorem Proving*. Unpublished. https://10.13140/2.1.3012.0641

Strickland, S., & Rand, B. (2016). A Framework for Identifying and Classifying Undergraduate Student Proof Errors. *PRIMUS : Problems, Resources, and Issues in Mathematics Undergraduate Studies, 26*(10), 905-921. https://10.1080/10511970.2016.1199615