What Makes a Proof a Proof?: It’s Your *Job* to Decide

Andre Rouhani
Arizona State University

Dov Zazkis
Arizona State University

This study investigates students' conceptions of what makes a particular written mathematical communication a proof. To that end, we implemented one-on-one task-based interviews involving mathematical communications. These communications were presented as trios, and we called each communication a job. We intentionally designed some of these jobs to sit in the vague space between proof and calculation. Our study investigated the reasons why participants believed certain jobs were proofs, allowing us to develop a model of students' conceptions of proofiness. This model revealed differences among students in terms of which attributes they thought contributed to or detracted from proofiness, as well as those they thought were and were not relevant to whether a job is a proof. These differences and how they relate to normative conceptions of proofiness point to aspects of proof that may require further attention in these students' introduction to proof courses.

Keywords: Proof, Introduction to proof students, Remote data collection.

What follows is the novice portion of a larger planned expert novice study (e.g., Inglis & Alcock, 2012). The overall project addresses the question “What makes a proof a proof?” We use the term *mathematical communication* to broadly capture any written work that establishes a mathematical fact or property. For example, a calculation that finds the roots of a polynomial might be considered a mathematical communication that establishes what the roots of the polynomial are. Similarly, a written proof is a mathematical communication that establishes the truth of some theorem.

This portion of the research addresses the research question: “What are undergraduate mathematics students’ conceptions of what makes a mathematical communication a proof?” To facilitate our discussion, we borrow a construct name from Seife (2010), *proofiness*. In his work “Proofiness: the dark art of mathematical deception,” Seife defines proofiness to be manipulations of mathematical norms that, “make falsehoods look like numerical fact” (Seife, 2010, p. 18). Instead, we use *proofiness* to refer to a person’s personal conceptions of the attributes that make a written communication a (correct mathematical) proof; in other words, *proofiness* refers to the qualities of a text that distinguish proofs from non-proofs. Those are the attributes that keep a purported proof from being an instance of proofiness in Seife’s sense of the word. So, from our perspective, our use of this construct name is more precise than Seife’s usage and more consistent with the mathematical concept of proof (Weber, 2014).

To simplify the conveyance of ideas, we refer to mathematical communications as *jobs*. We used our *jobs*-based tasks, and data from clinical interviews with undergraduate mathematics students, to develop a *proofiness* framework. This framework can be implemented to classify the attributes of a mathematical communication which a person thinks are relevant to certify or discount it as a proof. In the future, this might allow us to implement more informed learning trajectories for guiding students toward normative conceptions of proof. As alluded to earlier, determining mathematicians’ conceptions of *proofiness* is part of a planned followup to the current study. Thus, the utility of this framework will be demonstrated in a second order followup study. The students’ conceptions of *proofiness* will inform the design of a learning trajectory (Simon, 2020; Simon & Tzur, 2012) that is intended to advance students' understanding of proof toward observed mathematicians' conceptions.

Literature Review

Proof is, out of necessity, not well defined (Inglis & Aberdein, 2016, Weber & Czocher, 2019). This means that, even among expert mathematicians, there is often no consensus on whether a particular mathematical communication is a proof (Weber & Czocher, 2019). Czocher and Weber (2020) addressed this fuzziness by describing proof as a cluster concept (Lakoff, 1987), which is when "a number of cognitive models combine to form a complex cluster that is psychologically more basic than the models taken individually" (p. 74). The cluster concept nature of proof means that novice provers' conceptions of proof are, inherently, a fuzzy picture of an already fuzzy picture. However, whether a mathematical communication is a proof is also context dependent (Stylianides, Stylianides, & Weber, 2017) and students, especially those in introduction to proof (ITP) courses (David & Zazkis, 2020), have only experienced proofs in the single context of their course. This means that asking an ITP student the question "is this a proof?" implicitly asks, "is this considered a proof within the context of your ITP course?" The existing research on conceptions of what a proof is and what counts as a proof is primarily focused on expert mathematicians (e.g., Inglis, & Aberdein, 2016, Weber & Czocher, 2019). Here, we shift toward studying students, since a good grasp of student conceptions of proof is needed to use this knowledge as part of well-informed learning trajectories for proof.

We argue here that, especially given both how much variety there is in approaches to ITP course work (David and Zazkis, 2020) and the importance of proof in mathematics, there is value in exploring ITP students' impressions of what makes a proof a proof. However, the nebulous nature of proof necessitates a shift away from traditional tasks and interview questions (Zazkis & Hazzan, 1998).

Context: For $f: R \to R$ defined by $f(x) = x^{\tilde{}} + x - 6$,

Job (1): Justify that f has two roots.

Claim: f has roots -3 and 2.

Justification: Note that $f(x) = x^2 + x - 6 = (x + 3)(x - 2)$. If $(x + 3)(x - 2) = 0$, then $x = -3$ or $x = 2$. Thus, f has roots -3 and 2.

Job (2): Justify f is not one-to-one.

Claim: f is not one-to-one.

Justification: Suppose to the contrary that f is one-to-one. This means that if $f(x_1) = f(x_2)$, then $x_1 = x_2$. Note that $(-3, 0)$ and $(2, 0)$ are ordered pairs that satisfy f. In this case, we have $f(-3) = 0 = f(2)$, but $-3 \neq 2$. And this is a contradiction, since f is one-to-one. Thus, our supposition that f is one-to-one must be false. Therefore, f is not one-to-one.

Figure 1. An example task which presents two jobs for the subject to compare.

Method

Participants were undergraduate students, each of whom completed a zoom-based think-aloud one-on-one task-based clinical interview (Clement, 2000). Remote data collection was needed because of the COVID-19 pandemic. The ten participants were students in the same section of an ITP course offered at an Anonymous State University (ASU) in the Southwestern United States. Six of the ten were pursuing a degree in mathematics, and the rest were engineering or science majors.

Students were presented trios (or sometimes pairs) of *jobs* that spanned the range from proof to calculation (from our perspective). Throughout the interview, students were asked which of the *jobs* in front of them were proofs, and then asked to justify their assertions by referring to the *jobs* and their attributes. A pair of example *jobs* are shown above in Figure 1. As a reminder, most of the *jobs* in our study occurred in trios, but we include this pair due to space limitations.

Working in the spirit of grounded theory (Strauss and Corbin, 1997) allowed us to generate a framework of the components of *proofiness* that participants attended to. We refer to our organization and categorization of these components as the Personal Conceptions of *Proofiness* framework (PCP). Documenting which students attended to which segments of the PCP allowed us to create a personalized *proofiness* profile (PPP) for each participant.

Results

First, we present the components and definitions of our Personal Conceptions of *Proofiness* (PCP) framework. Then, we utilize the PCP to create personalized *proofiness* profiles (PPP) for each of our participants. We then narrow our focus to the parts of students' PPP that differ from student to student. These highlight differences in our participants' conceptions of what makes a *job* a proof. The differences in students' PPP are particularly relevant, since all participants were recruited from the same ITP section at the same institution, and thus had ostensibly comparable experiences with proof.

The PCP is divided into two *domains of proofiness*: *norms of communication* (how a mathematical communication is written and presented) and *purposes of communication* (why a mathematical communication is written, and how and whether it meets its mathematical aims). Figure 2 presents the *domains, subdomains, and dimensions of proofiness* that make up the PCP. Figure 2(a) displays the *norms of communication domain*. The *subdomains* in the left table of Figure 2(a) refer to mathematics-specific conceptions, while the *subdomains* on the right describe conceptions that pertain to argumentative writing in general. The primary difference between the left and right tables in Figure 2(a) is that the conceptions on the right are broadly applicable in written argumentation that is not mathematical (e.g., legal argumentation), while those on the left are more closely related to mathematical argumentation. Figure 2(b) details the *purposes of communication domain* of the PCP. Since the PCP was generated from student remarks in aggregate, the relative size of Figure 2(b) compared to Figure 2(a) is an important finding: it highlights that participants spoke relatively little about *purposes of communication*; most students focused their attention on *norms of communication.*

We used the PCP to classify each student's PPP. A subset of each student's PPP can be seen in Figure 3. The table is written so that + means the student believes that dimension *of proofiness* is necessary (though not sufficient) for a *job* to be a proof. We use ++ to indicate when the student was emphatic in their statement of a *dimension*'s importance that enthusiasm is not part of the analysis in this paper, we maintain the convention to be consistent with future work. We use - to indicate the student believes the *dimension* detracts from a *job's proofiness*. We use the symbol / to indicate that the student explicitly indicated that the *dimension* has no effect on whether a *job* is a proof. Finally, the symbol ~ means the student was inconsistent in their assessment of the *dimension*.

Domain	Sub-domain	Dimension	Description of Dimension
Norms of Communication	Sufficient Detail	Explicit Reference to Definitions	Stating "By definition of…" and especially followed by a verbatim statement.
		Assumptions Stated	Explicit statement of assumptions. As opposed to taking them as shared between author of claim and writer of Job. Even if assumptions are listed one line before, ensuring that assumptions are stated as a necessary first line.
		Fastidiousness	Thouroughness. Writing things out for the sake of being thorough, with attention to the detail in reproducing text they have read or written before, especially with definitions.
		Explicit Statement of Warrant / Backing	Stating why a certain inference was made, though no participant used that language. Explicit statement of warrant / backing (no participant used that language either). Participants talked about explaining why a step was taken (though not necesssarily why the outcome is true). If a function was factored, stating something like "By factoring, …". If the associate property was used, stating that it was. Ensuring that steps do not "come out of nowhere".
	Adherence to Norms of Mathematical Communication	Generality	Addressing more than a single case, preferably many or all cases.
		Arbitrariness	Acknowledgement of arbitrary choice.
		Variables	Letters as stand ins for either one or multiple numbers.
		Unidentified variables	Letters that are used but not intoruduced with words, described, or named.
		Specific Numbers	Using known, non-arbitary numbers with no variable label. Like "5".
		Symbols	Set notation. Elementhood / inclusion symbol. Blackboard R, N, and Z.
		Use of Examples	As stated by the participant, the presence of and use of an example. There were times when students called things examples that normatively were not.
	Mathematical Content and Context	Familiar Context	Contexts that participants have seen in their transition to proof course, including functions (but only special kinds like 1-1, and not polynomials), sets, divisibility, relations, number theory, etc.
		Unfamiliar Context	Contexts that participants have not seen in their transition to proof course that, importantly, remind them of their previous mathematics experiences. Especially calculus. No instances of mathematics contexts that the participants had never seen before. For instance, a Job whose mathematical context was in topology or graph coloring.
		Algebra	Presence of one or multiple lines of algebraic work like manipulation of equations, factoring, expansion of polynomials, etc.
		Higher Sophistication of Math Context	Lower sophistication: quadratic functions and calculus. Higher sophistication: number theory (divisibility) and classifying functions as 1-1. Mathematical concepts that were treated in their transition to proof courses felt more sophisticated to some participants.

Domain	Sub-domain	Dimension	Description of Dimension
Norms of Communication	Structure	Givens, Work, Goal	A structural concern. Starting with a statement of some assumptions and then taking steps (doing work) to somehow arrive at a goal.
		Proof Framework / Outline	The Job makes use of a recognizable proof framework or outline.
		Transition Words to Link Steps	Words like "hence", "thus", "therefore", "yields", "which means", "and then".
	Visual Cues	Paragraphs	As opposed to one block of text, the Job is written in paragraphs.
		Typesetting	Underlining of words like "Claim:","Justification:" and "Proof:", followed by indented text that contains the body of the Job.
		Centered Equations	As opposed to in-line equations.
		Pictures	The presence of pictures, especially graphs.
	Adherence to Norms of Written Argumentation	Conventionally Proofy Language	Words like "let", "assume", "suppose", "thus", and "therefore".
		Complete Sentences	The text of the Job is written in complete sentences.
		Long Length	More words. More lines. The Job takes up more physical space.
		Efficiency / Concision	The Job is lean. Does not take more steps than necessary. Brevity.
		The Word "If"	Presence of the word "if" in the Job.
		Voice	A composite dimension that is composed of various written English dimensions of *proofiness* . An understanding that proofs have or are written in
	Tone	Helps Reader	Effort is made by the Job to help or guide the reader, or otherwise advance the understanding of the reader.
		Convincing	The Job works to convince the reader of the truth of the claim or persuade the reader of something.

Figure 2(a): Norms of communication domain of PCP.

Domain	Sub-domain	Dimension	Description of Dimension
Purpose of Communication	Normative Mathematical Value of Proof	Inevitability	The claim is the logical conclusion that follows from the work done in the Job.
		Coherence	The conclusion of the Job matches the claim.
		Unpacking and Repacking	Definitions are used to make statements about some fact. Manipulation of those statements yields something of a sub-conclusion. This sub-conclusion is presented according to a definition, and this is the conclusion. Unpacking definition: writing a fact in a more useful form. Repacking definition: using that more useful form to make a conclusion, which is achieved by refrencing the language of a definition. Note that participants that explored this dimension of *proofiness* did not necessarily explicitly reference logical inference.
		Worthwile	The Job proves something that is nontrivial or proves something that requires a proof.
		Circular Logic	Assuming the conclusion. Or using rules that depend on other rules.
	Types of Claims	If / Then	The Job is an attempt ot justify a conditional statement.
		Answer-Getting	The Job is an attempt to solve a calculational problem. The Job amounts to a calculation.

Figure 2(b): Purposes of communication domain of PCP.

Participant	Major	Explicit Reference to Definitions	Inevitability	Coherence	Unpacking & Repacking	Worthwhile	Circular Logic	If / Then	Answer-Getting
Redmond	Mech E	++							
Megan	Math	++							-
Gem	Math			+					
Javon	Math	+	+		++	+	-		
Aidan	Math	++							
Eric	EE & Math			+			-	+	-
Claudette	Comm & Math	+							
Estrella	EE	/	++						
Han	Chemistry								-
Sabrina	*Professional*	~							

Figure 3: A subset of the participants PPP. Note the thick border, separating domains.

To illustrate these distinctions, consider the following two quotations from Han and Estrella, both of which are relevant to the Generality *dimension of proofiness* in the PCP framework:

Han: "Broader case feels *proofier* than talking about one specific case."

Estrella: "Not all proofs have to be arbitrary. You can still prove something for like a specific situation."

Han indicates he holds the conception that Generality is a dimension of *proofiness* that increases *proofiness*, while Estrella states that she does not believe that Generality affects *proofiness*. This indicates differences in their personal conceptions of proof. Since the mathematically normative conception of Generality is that a proof can be general or specific, depending on the claim, we see that not all students agree with that normative conception: two of them stated explicitly that to be more general was to be more *proofy*.

We further explore the PCP by comparing Claudette and Redmond's remarks associated with the Long Length dimension of *proofiness*, which relates to a *job* having more words, more lines, and taking up more space on the page. Claudette said she thinks proofs should have "lots of words", even referring positively to its "sheer amount of words" when claiming one *job* was *proofier* than another. Redmond, on the other hand, stated he did not like one of the *jobs* because, in his words, "It's too long." Therefore, Claudette has a + in the relevant space, while Redmond has a -.

The table in Figure 4 allows us to glean patterns and differences in the students' *personal proof profiles* (PPP). For example, we can notice that seven of the ten participants mentioned that they hold the belief that the inclusion of Conventionally *Proofy* Language contributes to *proofiness*. The three participants that did not verbalize this view on language did not bring up language at all, indicating de facto consensus. This stands in contrast to the group's conception of Algebra (the presence of one or multiple lines of algebraic work like manipulation of equations, factoring, expansion of polynomials, etc.)—some students believe that Algebra decreases a mathematical communication's (*job's*) proofiness, and others believe that Algebra is irrelevant to the *proofiness* of a mathematical communication. This indicates disagreement regarding whether the presence of algebra contributes to *proofiness*. More broadly, we found that, taken as a whole, there was consensus among our participants with respect to Assumptions Stated, Explicit Statement of Warrant / Backing, 'Givens, Work, Goal', Proof Framework / Outline, and Conventionally *Proofy* Language as positive contributors to *proofiness*. Moreover, we saw strong support for the detrimental effect to *proofiness* that Unidentified Variables, Unfamiliar Context, and Answer-Getting yielded (see Figure 2 for descriptions of these *dimensions of proofiness*). Further, students broadly disagreed about the *proofiness* of Long Length, Algebra, and Generality. Lastly, the Explicit Reference to Definitions *dimension of proofiness* is noteworthy: most participants agreed enthusiastically that this *dimension* contributed positively to *proofiness*, with only one stating it did not matter, and another being inconsistent. This is intriguing, if not worrisome, since many mathematicians believe that, across contexts, full statements of definitions should not be included in proofs (Lew and Mejia-Ramos, 2020). This discrepancy about whether to include definitions is related to lack of clarity (Ibid.) on what student-produced proofs should look like, according to mathematicians.

Dimension of Proofiness	Redmond	Megan	Gem	Javon	Aidan	Eric	Claudette	Estrella	Han	Sabrina
Explicit Reference to Definitions	++	++		+	++		+	/		~
Assumptions Stated	+		+		++		+			+
Fastidiousness			+		+		+			++
Explicit Statement of Warrant / Backing	+	+			+		+		+	+
Generality			+					/	+	/
Arbitrariness		+								
Variables	+		+				+			
Unidentified variables			-	-		-		-	-	-
Specific Numbers	-		-			/	-			
Symbols	+									
Use of Examples	~	-	-				-		~	
Familiar Context					+	+			+	
Unfamiliar Context			-	-	-	-			-	-
Algebra	-					-		/	-	/
Higher Sophistication of Math Context			+							/
Givens, Work, Goal	+				+		+	+	+	+
Proof Framework / Outline		+	+	+	+				+	+
Transition Words to Link Steps			+							
Paragraphs							+			
Typesetting							+			
Centered Equations							~			
Pictures				-			-			
Conventionally Proofy Language	+		+	+	+	+	++			+
Complete Sentences					+					+
Long Length	-	-	+			+	+		-	+
Efficiency / Concision			+	+			-	+	+	
The Word "If"		-								
Voice				/		+				
Helps Reader	+	+			+		++			~
Convincing		+					+			
Inevitability				+				++		
Coherence			+			+				
Unpacking & Repacking				++						
Worthwhile				+						
Circular Logic				-		-				
If / Then						+				
Answer-Getting		-				-			-	

Figure 4: PCP framework (applied to students). Note the thick border, separating domains.

One final observation merits consideration: Figure 4 illustrates sparse attention among the pool of participants to *purposes of communication.* In fact, only four participants had a + listed in any *dimension of proofiness* within this *domain*. Put another way, six out of the ten participants did not attend at all to the *purposes of communication domain.* This is understandable given that the participants are at the ITP stage in their mathematics education. However, we hope that future study can inform ways to support students like these to do more than "show their work" nicely, and in doing so they may consider more deeply exactly what the "work" of proof is.

Discussion and Conclusions

We used *job*-based tasks to study individual students' conceptions of *proofiness*. This allowed us to develop a framework for which attributes emerged in discussion of what makes something a proof. Compiling all these conceptions into one organized list of *dimensions of proofiness* is what we call the Personal Conceptions of *Proofiness* framework. Applying this to each of our participants creates a personalized *proofiness* profile, which allows us at a glance to notice patterns and trends in our data related to student beliefs about *proofiness*. For example, we were able to notice differences in our student participants regarding conceptions of generality in proof, and that there was consensus regarding students' belief that the use of a proof framework / outline was relevant to determining whether something is a proof–that is, it increased the *proofiness* of a *job*. Given that all participants were enrolled in the same introduction to proof (ITP) section, this information provides a kind of profile of conceptions of proof that emerged from that class as a whole.

In this connection, we could also envision using these same tools (our *jobs*-based tasks and the PCP) to compare the class profiles of two separate sections of the ITP course taught by different instructors with different instructional approaches. Documenting this could lead to richer understandings of what kinds of instructional approaches lead to conceptions of proof that are more/less aligned with normative mathematical perceptions of *proofiness*.

Finally, we plan on using and further developing both our *jobs*-based tasks and the PCP and getting a better sense of the range of mathematicians' PCP. This followup work should aid us in creating curricula aimed at, among other things, developing students' PCP and simultaneously tying this research to currently available research on mathematicians' conceptions of *proofiness*.

References

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. *Handbook of research design in mathematics and science education*, *547*, 341-385.

Czocher, J. A., & Weber, K. (2020). Proof as a cluster category. *Journal for Research in Mathematics Education*, *51*(1), 50

David, E. J., & Zazkis, D. (2020). Characterizing introduction to proof courses: a survey of US R1 and R2 course syllabi. *International Journal of Mathematical Education in Science and Technology*, *51*(3), 388-404.

Inglis, M., & Aberdein, A. (2016). Diversity in proof appraisal. In *Mathematical Cultures: The London Meetings 2012-2014* (pp. 163-179). Springer International Publishing.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, *43*(4), 358-390.

Lakoff, G. (1987). Women, fire, and dangerous things. Cambridge: Cambridge University Press.

Lew, K., & Mejía Ramos, J. P. (2020). Linguistic conventions of mathematical proof writing across pedagogical contexts. Educational Studies in Mathematics, 103(1), 43-62.

Seife, C. (2010). Proofiness: The dark arts of mathematical deception. *AMC*, *10*, 12.

Simon, M. (2020). Hypothetical learning trajectories in mathematics education. In *Encyclopedia of mathematics education* (pp. 354-357). Cham: Springer International Publishing.

Simon, M. A., & Tzur, R. (2012). Explicating the role of mathematical tasks in conceptual learning: An elaboration of the hypothetical learning trajectory. In *Hypothetical learning trajectories* (pp. 91-104). Routledge.

Strauss, A., & Corbin, J. M. (1997). *Grounded theory in practice*. Sage.

Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. *Compendium for research in mathematics education*, 237-266.

Weber, K. (2014). Proof as a Cluster Concept. *North American Chapter of the International Group for the Psychology of Mathematics Education*.

Weber, K., & Czocher, J. (2019). On mathematicians' disagreements on what constitutes a proof. *Research in Mathematics Education*, *21*(3), 251-270.

Zazkis, R., & Hazzan, O. (1998). Interviewing in mathematics education research: Choosing the questions. *The Journal of Mathematical Behavior*, *17*(4), 429-439.