Mathematicians' Validity Assessments of Common Issues in Elementary Arguments

Joshua B. Fagan
Texas State University

This study explores how mathematicians view validity in the face of explicit validity issues within written mathematical arguments in the context of the Introduction to Proof (ITP) setting. An internet survey of 30 arguments was constructed leveraging common issues in validity at the ITP level, and widely distributed to research-active mathematicians in the United States. The results suggest that there is no consensus as to the effect of any single validity issue on the overall validity of an argument, lending credence to the notion that argument validity lacks a consistent set of criteria from one mathematician's point of view to the next.

Keywords: Proof, Validity, Mathematician Practice

The nature of a valid, mathematical proof is difficult to define. In fact, several studies have established that even mathematicians have a number of points of disagreement on what constitutes a valid proof (Inglis & Alcock, 2012; Inglis, Mejía-Ramos, Weber & Alcock, 2013; Weber, 2008). This finding corroborates mathematicians' own accounts that there may not be any fixed set of standards for determining what is or is not valid within the mathematical context (e.g., Rav, 2007). The study presented here extends this tract of research by exploring on a large scale how mathematicians judge the effect that specific flaws within an argument can have on the validity of an argument at the Introduction to Proof (ITP) level. The effort is to explore in depth what standards might currently exist and what perceived requirements might lead to disagreement amongst mathematicians. Specifically, this research aims to answer the following questions:

- To what extent do mathematics professors agree about whether basic deductive arguments (at the ITP level) are proofs?
- What characteristics of deductive arguments account for disagreement in mathematician's validity assessment?

Background

In response to the assertion that argument validity is an important criteria when exploring undergraduate mathematics major's understanding of proof (Selden & Selden 2003), researchers have focused on mathematicians' ideas concerning validity to clarify existing standards and determine the consistency and importance of validity within mathematics at large and in the undergraduate mathematics classrooms (Inglis & Alcock, 2012; Inglis, et al., 2013; Weber, 2008). Weber (2008) investigated both the contextual criteria and strategies research-active mathematicians used when validating both elementary and advanced arguments. Weber found that there were a number of extra-mathematical criteria that the eight mathematicians from his study used in considering the validity of the arguments, including who the author of the argument was. One of the most important criteria for many of the mathematicians when looking at elementary proofs was the question of what had been established to be true. This key characteristic hits at the heart of any validity judgement as the building of a specific set axioms, theorems and the like within any setting – or the lack thereof – may require further argumentation on the authors part when constructing an argument. While Inglis and Alcock's (2012) main focus concerned the differences between novice and expert approaches to validating tasks, their findings concerning the 12 mathematicians in the study support the notion that mathematicians do not exhibit a uniform consensus of what might count as valid. Inglis, et al. (2013) expanded upon this idea by exploring how these disagreements in validating might arise in terms of a mathematician's area of expertise within mathematics, as well as exploring mathematicians' assessments of their own validity judgments in terms of their perception of the how other mathematicians would validate a proof. In the end, this study of 109 mathematicians and the two prior studies point to the same overall conclusion that validity is, as yet, a poorly defined construct which is case and individual dependent.

While each of these studies has helped to clarify the relationship between mathematician, context, and expertise and the role the latter pair play in validity judgments, none of them have offered deeper insight into individual, specific criteria relating to argument creation that might affect the validity of an argument. Meaning, for example, it is unclear how mathematicians might react in the face of a warranting issue within an argument or to an argument that begins by assuming the conclusion and showing the antecedent as a direct result. Are mathematicians consistent in their judgments of some set of perceived validity issues, but less consistent in others?

Framing

The idea of proof is nuanced in the mathematics education literature ranging from the overtly mathematical in nature (e.g., Healy & Hoyles, 2000; Knuth, 2002; Mariotti, 2000) where logic and deduction are stressed at the expense of all else, to the cognitive or social perspectives each focusing on aspects of conviction, and communal acceptance (e.g., Balacheff, 1988; Harel & Sowder, 2007). For this study, I adopt Stylianides' (2007) definition:

> *Proof* is a *mathematical argument*, a connected sequence of assertions for or against a mathematical claim, with the following characteristics:
> 1. It uses statements accepted by the classroom community (*set of accepted statements*) that are true and available without further justification;
> 2. It employs forms of reasoning (*modes of argumentation*) that are valid and known to, or within the conceptual reach of, the classroom community; and
> 3. It is communicated with forms of expression (*modes of argument representation*) that are appropriate and known to, or within the conceptual reach of, the classroom community. (p. 291; emphasis in original)

From this definition the understanding is gained that a proof is a mathematical argument which is defined by three distinct characteristics concerning statement, modes of argumentation, and representation. For this research, the scope of what is a proof is limited to statements and representation which are common in the ITP classroom (David & Zazkis, 2016), and modes of argumentation that consist specifically of direct proof[1].

Validity

Selden and Selden (2003) called proof validation, "the reading of, and reflection on proofs to determine their correctness" (p. 5). While correctness may imply some sort of universal standard, the validity standing of an argument is often subjective. I treat proof validation is the act of judgement or evaluation which leads the reader to identify whether a given argument appropriately proves a statement.

Finally, for linguistic clarity, I take *argument* to represent the body of all purported proofs regardless of their validity. Thus, to ascribe a series of logical (or illogical) statements as

[1] Direct proofs in this research included proof by cases.

an argument is to remove any notion of validity from the conversation. Arguments are valid-neutral. On the other hand, identifying an argument as a *proof* is to remove its valid-neutrality and assert that it is valid.

Common Validity Issues

To give context to the validity judgements that mathematicians made, this study leverages the idea of *issues in proof writing* which focuses on the prevalent validity issues amongst undergraduate mathematics major's own written arguments (Hazzan & Leron, 1996; Selden & Selden, 1987, 2003). The validity issues considered for this study fall into one of six categories as presented in Table 1. Each argument which was initially coded as invalid had the inclusion of a single example of one of the six issues where each argument was intended to present a single validity issue to the participants. Though often the validity issues were simple in nature, they were chosen or constructed because they represented issues that are considered common in undergraduate mathematics (Alcock & Weber, 2005; Hazzan & Leron, 1996; Selden & Selden, 1987, 2003; Weber & Alcock, 2005; Weber, 2001).

Table 1. Common Validity Issues

Issue (Abbr.)	Definition
Assuming the Conclusion (AC)	An argument assumes the consequent (conclusion) of the proposition it is claiming to prove and attempts to show that the antecedent is a direct consequence.
Circular Reasoning (CR)	An argument assumes the consequent (or antecedent) of the statement it is claiming to prove and comes to a trivial conclusion, namely the consequent (or antecedent) once again. Equally, within an argument a claim is made and used to argue to trivial ends, the claim itself. ($P \rightarrow Q \rightarrow \cdots \rightarrow P$)
Logical Gap (LG)	An argument omits a portion of reasoning; the argument has a hole. This could be thought of as a lack of explicit warranting where such would seem prudent.
Misuse of Notation (MN)	Within an argument, proper notation or variable naming conventions are not adhered to, or notation and variable naming conventions are used inconsistently.
Warranting (W)	Within an argument, an error in justification is made either explicitly or implicitly. This can take the form of an incorrect explicit warrant, or an incorrect implicit warrant which may emerge as an arithmetic or computational error.
Weakening the Theorem (WT)	An argument proves less than what is implied by the statement being proven or begins by assuming more than is permissible.

Methods

Qualtrics, an internet survey system, was used as the main resource for data collection in this study to obtain a large sampling of mathematicians. The survey itself consisted of 30 arguments to 22 different propositions considered germane in the ITP setting (David & Zazkis, 2016). For each argument, the participants were first asked, "Is the argument for the included proposition a valid proof?" and given the binary option of "Yes - valid," or "No - invalid." Participants were initially warned against grading the proofs as though they were student proofs, but to instead answer for themselves the question, "Does this argument actually prove the proposition in a way that I feel is appropriate, based upon what I believe is requisite for an argument to be valid?" In this way it was left to the participant to infer what they felt was requisite for an argument to be a valid proof.

If the proof was initially coded as invalid and the participant disagreed (i.e., they chose "valid" as their response) the participant was presented with the proposed validity issue and asked how the presence of said flaw affected their initial response, and then given the chance to change their minds about the validity of the argument[2]. If the participant did not change their mind, they were asked to share why they felt the flaw did not invalidate the argument. Additionally, for each argument that was initially coded as invalid, if the participant agreed that it was in fact an invalid argument, they were also presented with the flaw and asked if it was the reason they choose invalid. If it was not the reason, participants were asked to state why they thought the argument was invalid. For all arguments which were initially coded as valid, if the participant disagreed and chose invalid, they were asked to justify their views by stating why they thought the argument was invalid.

The arguments themselves were clustered into one of seven groupings based upon their initial validity coding and issue. Participants were then randomly presented with an argument from each cluster to ensure that they saw an argument whose flaw came from each area of the framework as well as being presented with an argument which was considered to initially be valid. No participant saw the same argument twice. In total, 1528 survey invitations were distributed via email to research-active mathematicians across the United States, of which 228 submitted responses to the survey. Of the 228 participants, 178 completed all 7 argument sets with which they were presented, all others completed no less than 2 argument sets.

All free responses were analyzed using thematic analysis (Braun & Clarke, 2006). The analysis began with open coding of the free responses for each of the 30 arguments independently and categorizing responses relative to each argument in terms of their appropriateness. All nonsensical free-responses led to a cycle of analysis of the quantitative data supplied by the author of said free-response to ensure the author was not supplying malicious data[3]. Malicious data was omitted from further analysis. Following open coding, themes were identified, categorized and condensed for each argument. No cross-argument analysis occurred as the questions for this study do not focus on how responses to one type of validity issue are correlated to responses to other validity issues.

Results

Figure 1 comprises the final validity judgements to all 30 arguments including those of which were initially coded as valid (i.e., arguments V1-V5). The chart represents the percentage of mathematicians that deemed each argument to be invalid calculated by taking the total number of "No - invalid" responses along with the number of mathematicians who changed from "Yes – valid" to invalid and then dividing by the total number of responses. For both the set of valid and invalid arguments, the number of responses for each argument was not uniformly distributed due in part to the random design of the survey and the inclusion of partially completed responses. Disagreements among mathematicians was found in every category, and while there are cases

[2] There were four exceptions where no follow-up was requested. Three of these four arguments contained a logical gap which initially it was unclear if the gap would affect the validity of the argument.

[3] Quantitative data was considered malicious if the entire survey had been completed in under 10 minutes, any one validity set – the initial validity question and all follow-up questions – was completed in less than one minute, or if all validity questions were homogeneous and all *other* follow-up free response questions were left blank.

where 100% of mathematicians agreed that something was invalid (AC1-AC3), no one category was free from disagreement.

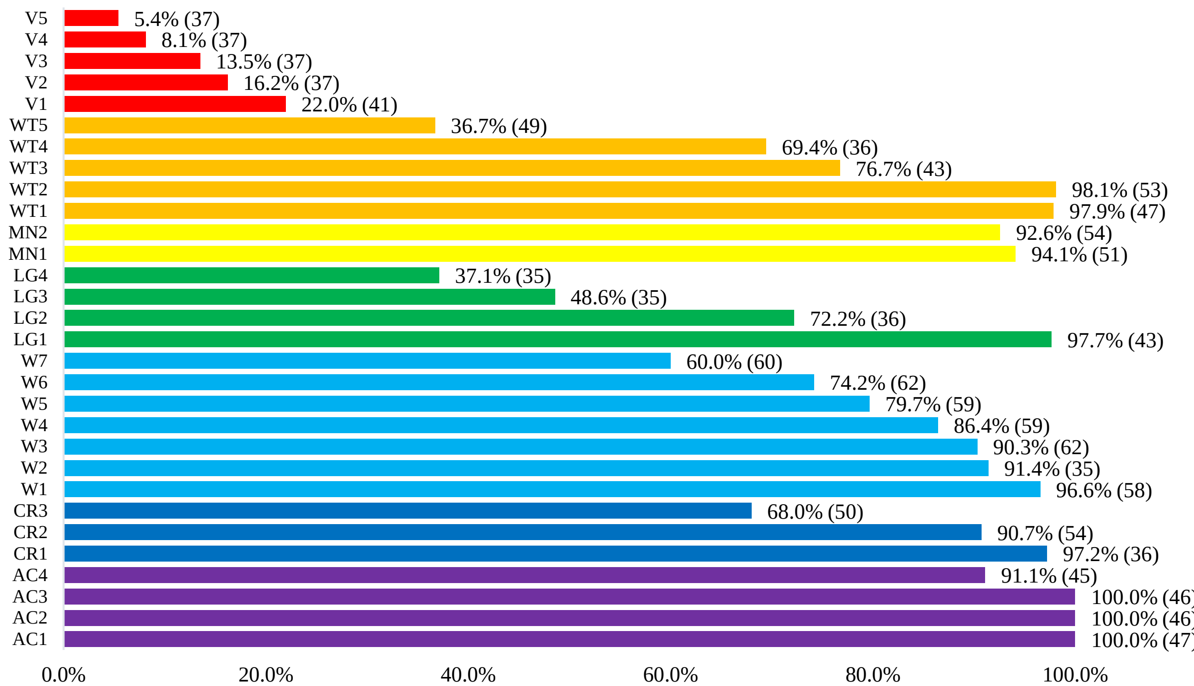

Figure 1. Percentage of mathematicians who thought the argument was invalid (number of responses). Each argument that was initially coded as invalid was given a name and number based on its included validity issue (i.e., WT – weakening the theorem, MN – misuse of notation, LG – logical gap, W – warranting, CR – circular reasoning, and AC – assuming the conclusion).

A Weakening the Theorem Example

One of the more interesting weakening the theorem results, argument WT5's (Figure 2) validity issue was that the argument did not account for the negative integers and zero when defining the parameter x as odd (i.e., "$x = 2a + 1$ for some $a \in \mathbb{N}$" instead of "for some $a \in \mathbb{Z}$"), thus arguing for something weaker than what was intended to be implied by the proposition. Mathematicians who felt this was not enough to invalidate the argument fell into two general groups, the first arguing that the proposition itself does not clearly define odd as to mean odd integers versus odd natural numbers. The second, and perhaps more pertinent group of mathematicians thought that, though the argument failed to account for the negative integers and zero, because the structure and logic of the argument was intact the weakening of the theorem that had occurred did not invalidate the argument. For instance, one mathematician said, "The heart of the argument is understanding that odd numbers are 1 mod 2 and that an odd number squared is 1 mod 2, which remains valid. The error is minor because of its consequence. If this was a proof involving absolute values and the negative numbers [were] not properly dealt with that would be much more damning." Thus, despite the inaccuracy these mathematicians felt the argument was valid.

Proposition: If x is odd, then x^2 is odd.

Argument: Suppose x is odd. Then $x = 2a + 1$ for some $a \in \mathbb{N}$. Thus we have

$$x^2 = (2a+1)^2 = 4a^2 + 4a + 1 = 2(2a^2 + 2a) + 1.$$

Since $2a^2 + 2a \in \mathbb{N}$, then $2(2a^2 + 2a) + 1$ is odd, and therefore so is x^2.

Figure 2. Argument WT5 – 36.7% of mathematicians thought the weakening that occurred was enough to invalidate the argument.

A Warranting Example

The validity issue of warranting lead to unclear results in term of agreement. Arguments W5 (Figure 3) and W7 (Figure 4) argued for the same proposition and had identical arguments save for an explicit warranting issue which occurred at the same point within each argument. The difference came about in the perceived reasonability of the warranting issue which lead 60% of mathematicians to conclude that W7 was invalid, while 79.7% of mathematicians thought the warranting issue in W5 was sufficient to invalidate the argument[4]. Many of the mathematicians who claimed that W7 was valid cited the "minor typo" that occurred did not underscore the soundness of the argument as a whole. On this fact many mathematicians made statements like, "Yes, it is the incorrect term for the property being used; however, the property actually used (multiplication on R is commutative) is certainly true, so the argument is still valid." This contrasts with the general sense that though commutativity was also correctly used in W5, there was a much stronger negative reaction in term of validity to the claim that "multiplication is an equivalence relation in ℝ." It should be noted that no one argued that either was a true statement.

Proposition: For $a, b \in \mathbb{R}$, if $a < b$ and $0 < a$ then $a^2 < b^2$.

Argument: Let $a, b \in \mathbb{R}$ and suppose that $0 < a < b$. Now, consider the following

$$\begin{aligned} 0 < a < b &\Rightarrow a \cdot a < b \cdot a \\ &\Rightarrow a^2 < b \cdot a \\ 0 < a < b &\Rightarrow a \cdot b < b \cdot b \\ &\Rightarrow a \cdot b < b^2 \end{aligned}$$

But since multiplication is an equivalence relation in $\mathbb{R}$, then $b \cdot a = a \cdot b$. Thus we have that $a^2 < b \cdot a = a \cdot b < b^2$, thus $a^2 < b^2$ as required.

Figure 3. Argument W5 – 79.7% of mathematicians thought the claim that "multiplication is an equivalence relation in ℝ" was enough to invalidate the argument.

Proposition: For $a, b \in \mathbb{R}$, if $a < b$ and $0 < a$ then $a^2 < b^2$.

Argument: Let $a, b \in \mathbb{R}$ and suppose that $0 < a < b$. Now, consider the following

$$\begin{aligned} 0 < a < b &\Rightarrow a \cdot a < b \cdot a \\ &\Rightarrow a^2 < b \cdot a \\ 0 < a < b &\Rightarrow a \cdot b < b \cdot b \\ &\Rightarrow a \cdot b < b^2 \end{aligned}$$

But since multiplication is associative in $\mathbb{R}$, then $b \cdot a = a \cdot b$. Thus we have that $a^2 < b \cdot a = a \cdot b < b^2$, thus $a^2 < b^2$ as required.

Figure 4. Argument W7 – 60% of mathematicians thought the claim that "multiplication is associative in ℝ" was enough to invalidate the argument.

An Assuming the Conclusion Example

Even in the case of arguments which assume the conclusion, thus having major structural issues there was some amount of disagreement. Argument AC4 (Figure 5) argues the converse of

[4] The difference between the two proportions is statistically significant with $p = .0196$, 95% CI [0.0195,0.3582] with continuity correction.

the proposition, and despite having this fact pointed out to them, four mathematicians held that the proof was valid making statements like, “I would say this argument is almost correct rather than invalid,” or “It could be modified quite quickly for the proof to be correct. The main idea is still there.” Thus, despite arguing the converse and even though these mathematicians agree the argument is not *correct* they felt it was valid.

Proposition: The sum $x + 4$ is odd whenever x is also odd.

Argument: Assume that $x + 4$ is odd, then there exists an integer n such that $x + 4 = 2n + 1$. Thus we have that $x = 2n - 4 + 1 = 2(n - 2) + 1$. Since $n - 2 \in \mathbb{Z}$, then x is odd.

Figure 5. Argument AC4 – 91.1% of mathematicians thought the argument for the converse was invalid in light of the proposition.

A Logical Gap Example

Finally, mathematician’s sense of the affect of logical gaps lead to an interesting result with argument LG3 (Figure 6). Here, the argument presented trivializes the proving process at many points with unsupported statements, twice using the phrase “which implies” in place of an actual argument. This lack of overt justification divided the mathematicians’ validity stance with 48.6% of mathematicians claiming the argument was invalid.

Definition: The symmetric difference of A and B is defined as $A \triangle B = (A - B) \cup (B - A)$.

Proposition: For any sets A, $A \triangle A = \varnothing$ and $A \triangle \varnothing = A$.

Argument: Let A be a set, then by the definition of symmetric difference $A \triangle A = (A - A) \cup (A - A)$. But $(A - A) \cup (A - A) = \varnothing$ which implies that $A \triangle A = \varnothing$, as required. Furthermore, also by the definition of symmetric difference $A \triangle \varnothing = (A - \varnothing) \cup (\varnothing - A)$. But here $(A - \varnothing) \cup (\varnothing - A) = A$ which implies that $A \triangle \varnothing = A$. Thus we have shown that for any set A, $A \triangle A = \varnothing$ and $A \triangle \varnothing = A$.

Figure 6. Argument LG3 – 48.6% of mathematicians thought the lack of overt justification was enough to invalidate the argument.

Conclusion

In response to the first research question, the data from this study reflects that even in direct, deductive proofs at the elementary level, there is a substantial disagreement amongst mathematicians over validity. This finding corroborates the findings from Inglis et. al. (2013) that mathematicians use different standards in judging an argument’s validity. The divergence was not unexpected in terms of more subjective proof aspects such as the allowable size of a logical gap. However, these divergent validity standards were apparent even when an argument had a major structural issue: assuming the conclusion.

The disagreement over validity has implications for instruction. Particularly, the fact that these inconsistencies may counter the dominant narrative that mathematics is universal. Furthermore, inconsistency across instructors could lead to cognitive dissidence in student’s proof writing and reading as they progress through a tract in undergraduate mathematics, and perhaps beyond.

Finally, taken together with past research, this data suggests that not only do mathematicians have different standards for what is and is not valid, but they might not have a good sense of what valid means generally as such a notion may not exist in a binary sense (Rav, 2007). This in turn leaves some questions about whether we as mathematics education researchers have a good feel for what validity is as well. If nothing else, future studies should be careful in making claims about validity in terms of absolutes, as there may be no absolute standard, at least not for elementary arguments.

References

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

Balacheff, N. (1988). Aspects of proof in pupils' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp. 216-230). London: Hodder & Stoughton.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

David, E. J., & Zazkis, D. (2017) Characterizing the nature of introduction to proof courses: A survey of R1 and R2 institutions across the U.S. Presented at the *20th Annual Conference on Research in Undergraduate Mathematics Education*. San Diego, CA.

Harel, G. & Sowder, L. (2007). Towards a comprehensive perspective on proof. In F. Lester (ed.) *Second handbook of research on mathematical teaching and learning*. NCTM: Washington, DC.

Hazzan, O., & Leron, U. (1996). Students' use and misuse of mathematical theorems: The case of Lagrange's theorem. *For the Learning of Mathematics*, *16*(1), 23-26.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, *31*(4), 396–428.

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

Inglis, M., Mejia-Ramos, J. P., Weber, K., & Alcock, L. (2013). On mathematicians' different standards when evaluating elementary proofs. *Topics in cognitive science*, *5*(2), 270-282.

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

Mariotti, M. A. (2000). Introduction to proof: The mediation of a dynamic software environment. *Educational studies in mathematics*, *44*(1), 25-53.

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.

Rav, Y. (2007). A critique of formalist-mechanist version of the justification of arguments in mathematicians' proof practices. *Philosophia Mathematica, 12*, 291–320.

Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, *34,* 4–36.

Selden, A., & Selden, J. (1987). Errors and misconceptions in college level theorem proving. In *Proceedings of the second international seminar on misconceptions and educational strategies in science and mathematics* (Vol. 3, pp. 457-470). Cornell University New York.

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

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, *48*, 101-119.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, *39*, 431–459.

Weber, K., & Alcock, L. (2005). Using warranted implications to understand and validate proofs. *For the Learning of Mathematics*, *25*(1), 34-51.