Why and how mathematicians read proofs: further evidence from a survey study.

Juan Pablo Mejia-Ramos and Keith Weber

Running title: Mathematicians' proof reading

Abstract

In a recent paper (Weber & Mejia-Ramos, *Educational Studies in Mathematics* 76:329-344, 2011) we reported findings from two small-scale interview studies on the reasons why and the ways in which mathematicians read proofs. Based on these findings we designed an Internet-based survey that we distributed to practicing mathematicians working in top mathematics departments in the United States. Surveyed mathematicians (N=118) agreed to a great extent with the interviewed mathematicians in the exploratory studies. First, the mathematicians reported that they commonly read published proofs to gain different types of insight, not to check the correctness of the proofs. Second, they stated that when reading these proofs, they commonly: (1) appeal to the reputation of the author and the journal, (2) study how certain steps in the proof apply to specific examples, and (3) focus on the overarching ideas and methods in the proofs. In this paper, we also report findings from another section of the survey that focused on how participants reviewed proofs submitted for publication. The comparison of participant responses to questions in these two sections of the survey suggests that reading a published proof of a colleague and refereeing a proof for publication are substantially different activities for mathematicians.

Key words: Mathematical practice, Proof, Proof reading,

1. Introduction

In advanced mathematics courses, proof is a primary way in which teachers and textbooks justify and explain mathematical statements to students. Thus, students spend a significant amount of time reading proofs in these courses. However, research has shown that when reading proofs, students find them to be confusing or pointless (e.g., Harel, 1998; Porteous, 1986; Rowland, 2001) and cannot distinguish a valid proof from an invalid argument (Selden & Selden, 2003; Weber, 2009). In an attempt to delineate what good proof reading looks like, some researchers have recently turned to the investigation of proof reading in expert mathematical practice.

In a recent article, Weber and Mejia-Ramos (2011) coordinated theoretical accounts of mathematical practice with findings from two recent empirical studies—interviews with nine mathematicians about proof reading and observations of eight mathematicians evaluating proofs for correctness—to form a model of how and why mathematicians read the published proofs of their colleagues. In particular, they attended to the role of non-deductive evidence (e.g. empirical and authoritative evidence) in mathematicians' proof reading. However, due to the relatively small sample size of these studies, they presented these findings as tentative hypotheses. The purpose of this report is to assess the viability of our results with a large number of research-active mathematicians.

2. Background

2. 1. Why mathematicians read proofs

When asked what they hoped to gain by reading the published proofs of others, mathematicians in Weber and Mejia-Ramos's (2011) interview study indicated that they did not only read proofs for correctness, but also to gain different types of mathematical insight. In fact, some participants claimed they did not check published proofs for correctness at all. One particular type of insight that all participants valued highly was discovering techniques or ideas that might be applicable to their own research. Some participants went so far as to question the value of a proof of a new theorem that does not contain new ideas.

2. 2. How mathematicians read proofs

Weber and Mejia-Ramos (2011) noted that the interviewed mathematicians talked about understanding a proof in three different ways, which corresponded to their description of what they did when they read proofs, and closely matched theoretical accounts of mathematical practice related to proof.

Proof as cultural artifact. A published proof can be viewed as an artifact with a history in a mathematical culture. At this level, reading a proof focuses on the contextual aspects of its construction and publication. Accordingly, some of the interviewed mathematicians claimed that if a proof is written by an authoritative source with a reputation for making careful arguments, or if the proof is published in a prestigious journal with a reliable reviewing process, they are convinced that the theorem is true without reading its proof at all. In this case, these mathematicians appear to be relying upon the authority of the author of the proof or the journal's reviewers to obtain conviction.

Proof as a sequence of inferences. According to Rav (1999) and Azarello (2007) the proof of a theorem can also be conceptualized as a series of claims of the form $A_0 \rightarrow A_1 \rightarrow A_2 \rightarrow A_{3\ldots} \rightarrow A_n$, where A_n is the theorem. In this conceptualization, reading a proof focuses on how each new inference in a proof is derived from previous inferences. Our research revealed that while participants often studied how a new inference followed from previous inferences by constructing a sub-proof, they also did so by considering carefully chosen examples— e.g., to see if every integer n has the property that $n^2 \equiv 0(mod\ 4)$ or

$n^2 \equiv 1(mod\ 4)$, one participant in Weber (2008) checked this assertion for *n*=5, 6, 7, 8, and 24. This suggests empirical evidence plays an important role in mathematicians' conviction.

Proof as the application of methods. Participants in Weber and Mejia-Ramos's (2011) interview study also talked about understanding proofs in terms of the main ideas or methods being applied in the proof. In this case, reading a proof focuses on these overarching ideas and methods. Indeed, some participants claimed that understanding a proof in these terms was as important, *if not substantially more important,* than understanding how each step followed logically from previous steps. Surprisingly, participants claimed they focused on these high-level ideas and methods, not only when attempting to understand a proof, but also when evaluating its validity. One participant indicated that he did not always perform a line-by-line check, even when refereeing a paper for publication. Similarly, some participants indicated that if they were convinced that the main ideas or methods in the proof were valid, they would be convinced that the proof was correct, even if they did not verify every line of the proof.

2. 3. Limitations of the qualitative study

The analysis of Weber and Mejia-Ramos (2011) has three limitations that threaten the validity of the findings. First, due to the relatively small sample sizes of the study, it is possible that the interviewed mathematicians' views and practices are not representative of those of the larger mathematical community. Indeed, mathematicians are not always homogeneous in their views and practices (e.g., Burton, 2004, 2009; Geist, Löwe, & Van Kerkhove, 2010). Thus, it is plausible that the quotes we excerpted were indicative of the perspectives of *some* of the mathematical community, but a sizeable percentage of this community might have the opposite viewpoint. Second, although the interviewees in Weber and Mejia-Ramos (2011) were instructed to discuss the practice of reading published proofs in journals, some nonetheless described the process of refereeing papers. Hence, the results reported by Weber and Mejia-Ramos might have conflated the different activities of reading proofs in published journals and refereeing proofs for publication. The present study addresses both of these threats to validity and provides further evidence of the findings reported in Weber and Mejia-Ramos (2011).

Finally, as both the data presented in Weber and Mejia-Ramos (2011) and the surveys presented in this paper rely on self-report, we can make strong generalizable claims about what mathematicians *claim* to do when reading a proof. However, as Inglis and Alcock (2012) argued, this might not be indicative of what mathematicians *actually* do. Consequently we believe that studies examining mathematicians' actual practice when reading proofs would complement and strengthen the claims put forth in this paper.

3. The survey study

In this study, we adapted the methodology of Heinze (2010), who recommended complementing qualitative data and philosophical analyses with quantitative studies to build a more robust understanding of mathematical practice. Heinze constructed a survey to explore the different criteria that mathematicians employed to accept mathematical arguments. Our survey is similar, and explores why and how mathematicians read proofs.

We argue that the current survey study addresses the first two weaknesses listed above: we have a tenfold increase in our sample of mathematicians, and our instructions explicitly separate the contexts of reading published proofs in journals and refereeing proofs for publication. However, we note there are still shortcomings that could be addressed by further research. First, the current survey study relies on mathematicians' retrospective reports about whether or not some events are uncommon in their reading of proofs. In general, this type of report could inaccurately describe participants' actual behavior for a variety of reasons:

participants may be unaware of certain behaviors or events, they may be unable to recall a perceptible event (and its frequency), and they may report something different from what actually occurred (Ericsson & Simon, 1980; Nisbett & Wilson, 1977). Thus, although verbal reports have provided valuable information on reading research (Afflerbach & Johnston, 1984), these findings must be corroborated by studies employing other methodologies.

Second, while we have differentiated the contexts of reading published proofs in journals and refereeing proofs for publication, and identified two different goals for reading on these contexts (i.e. checking correctness and gaining insight), further research may focus on mathematicians' proof reading strategies for more specific sub-goals. For instance, one could inquire about more specific reasons why mathematicians want to comprehend a proof, and study whether different sub-goals influence their use of specific proof reading strategies. More generally, further research should address the why and how mathematicians read, not only proofs (the focus of this paper), but also the manuscripts and books that contain them.

3.1. Method

Following the methodology employed by Inglis and Mejia-Ramos (2009), we collected data through the Internet in order to maximize our sample size. Recent studies have examined the validity of Internet-based experiments by comparing this type of studies with their laboratory equivalents (e.g. Kranz & Dalal, 2000; Gosling et al., 2004). The notable degree of congruence between the two methodologies suggests that, by following simple guidelines, Internet data has comparable validity to more traditional data.

Reips (2000) notes one practical threat to the validity of Internet-based studies is multiple submissions from the same individual. In order to address this threat, we adopted the strategy advocated by Reips, and implemented by Johnson-Laird and Savary (1999), and logged the IP address of each participant and the time they submitted their response. Under the assumption that each IP address was associated with a unique individual, these data were used to screen for possible cases of multiple submissions. Given our adherence to Reips' (2000) guidelines, and the impracticality of obtaining large samples of research-active mathematicians in any other fashion, we believe our methods were justified.

Participants. We recruited mathematicians to participate in this study as follows. Twenty-four secretaries from top-ranked mathematics departments in the United States[1] were contacted and asked to distribute an email to the mathematics faculty, post-doctoral researchers, and PhD students of that department. A total of 118 mathematicians agreed to participate.

When participants clicked on the link to the survey website, they were taken to a webpage that described the purpose of the study and asked for demographic information, including their status (doctoral student, post-doc, or mathematics faculty), their level of experience in mathematics research (0-3 years, 3-6 years, 6-9 years, 9-12 years, or more than 12 years), their primary area of research (according to the Mathematics Subject Classification), whether they had refereed a mathematics research paper submitted for publication in an academic journal, and how many of these papers they refereed per year (0, 1 to 2, 3 to 4, 5 or more).

Of the 118 participants, 55 participants stated they had refereed a paper (10 doctoral students, 14 post-docs, and 31 mathematics faculty), while the remaining 63 participants said they had not (55 doctoral students, 5 post-docs, 2 faculty members, and 1 non-response).

Materials and procedures. After completing the demographic information, participants were shown a screen saying they "will be asked about what you do when you are reading a proof

[1] As ranked by the USNews.com "Best Graduate Schools" list of "top mathematics programs".

that a colleague *published in a respected academic journal*" (the italics appeared in the text to participants). They were then asked to declare the extent to which they agreed (strongly disagreed, disagreed, neither agreed nor disagreed, agreed, or strongly agreed) with each one of 17 statements about why and how they read published proofs. Fourteen of these statements were based on hypotheses generated in Weber and Mejia-Ramos (2011) and are presented in Table 1.[2] Except for statements M2 and C1, all statements began with: "When I read a proof in a respected journal, …"

***** Insert Table 1 About Here *****

After answering these questions, the 63 participants who had not yet refereed a mathematical paper were shown a screen thanking them for participating in the survey. The remaining 55 participants who had refereed a mathematical paper were then shown a screen saying, "For the next set of questions, you will be asked what you do when you are reading a proof in a manuscript that has been submitted to a journal for publication. *Your role in this case is as a referee*" (italics were in the instructions to the participants). Participants were then shown the statements in Table 1, except that the phrase "When I read a proof in a respected journal" was replaced with "When I referee a manuscript". Questions C1, C2, M1, and M4 were not included in this second section of the survey.

For both sets of questions, we included three "foil" questions of behaviors that we did not think mathematicians would engage in (e.g., reading a proof to explore the writing styles of

[2] The items in Table 1 are organized and named by general theme (not any particular construct): questions regarding the purpose of reading proofs are named as P_, items related to the use of examples are named E_, questions regarding proofs as application of methods are named M_, items related to contextual/cultural aspects of proof are named C_, and foils are named F_. Participants completing this survey were not aware to which theme the questions belonged to. We did not necessarily anticipate a high correlation between survey items for each theme as in some cases they were discussing ideas that seemed to be very different.

academics from different countries). These were included to verify that participants would not agree to saying it was not uncommon that they engaged in any plausible behavior.

3. 2. Results

Participants' responses when asked about reading published proofs in journals

The participants' responses to reading published proofs in journals are presented in Table 2. For each item, we performed a one-sample Wilcoxon signed-rank test to see if participants' median responses differed significantly from neutral. Because there were 14 separate tests conducted, to assess significance we used a Bonferroni adjustment and set the critical alpha level at .003 (.05/14). The data in Table 2 strongly confirm the hypotheses that Weber and Mejia-Ramos (2011) advanced based on their interviews with nine mathematicians. For each of the 14 survey items used in this study, the large majority of mathematicians agreed with them (for 12 of the items, the level of agreement was over 70%), less than 20% of mathematicians disagreed with them, and their median responses differed significantly from neutral. We also note that our foils had their intended effect: for each foil item, less than 12% of participants agreed with the item and the majority of participants disagreed with it, indicating that most of the participants were not simply agreeing to any assertion. In summary, these data considerably strengthen the claims of Weber and Mejia-Ramos (2011) and limit the possibility that their findings were due to a small sample size or their participants conflating reading a proof for comprehension with refereeing.

***** Insert Table 2 around here *****

Of the 118 surveyed mathematicians, 69 claimed to have less than six years experience as a research mathematician and 49 claimed to have more than six years experience. To examine the effects of experience on participants' responses, we conducted post-hoc Mann-Whitney tests comparing participants with less than and more than six years experience, again using a Bonferroni-adjusted alpha-level of 0.003. Only two tests satisfied this standard of significance, with less experienced participants showing a higher level of agreement with P2 (reading a proof to learn new proving methods) and C3 (judging a proof to be correct because it was written by a trustworthy authoritative source). We note that in both cases, over 70% of both groups of participants expressed agreement with both assertions. These results are consistent with Heinze (2010), who also found only small differences between less experienced and more experienced mathematicians in how they obtained conviction.

Participants' responses when asked about refereeing a manuscript

Table 3 presents the responses of the 54 participants[3] who had experience reviewing to questions on how they read proofs in manuscripts that they were refereeing. For each item, we performed a one-sample Wilcoxon signed-rank test to see if participants' median responses differed significantly from neutral. Because there were 10 separate tests conducted, to assess significance we used a Bonferroni adjustment and set the critical alpha level at .005 (.05/10).

***** Insert Table 3 around here *****

Perhaps the most noteworthy feature of Table 3 is the extent of disagreement between these mathematicians on some of the items. In particular:

- 43% of participants agreed with E4—that they would sometimes gain sufficiently high levels of conviction in an assertion in a proof by empirical evidence to accept the statement as true—but 28% disagreed;

[3] One of the 55 participants who stated he/she had refereed a mathematics research paper submitted for publication did not answer any questions in the refereeing section of the survey.

- 39% of participants agreed with M3—that they would sometimes accept a proof as correct if they were convinced the main idea was correct without needing to check every line in the proof—but 44% disagreed;
- 39% of participants agreed with C3—that they would sometimes be convinced a proof was correct if it came from a trustworthy authoritative source—but 41% disagreed.
- 35% of participants agreed with M5—that they would sometimes not check every line in a proof that they were refereeing—but 52% disagreed.

Findings such as these reveal that mathematicians' behavior when refereeing might not be homogeneous. It is also noteworthy that for some mathematicians refereeing might be a less rigorous process than is commonly believed.

To explore if there was a difference in the proof-reading behaviors when mathematicians read published proofs in journals and when they were refereeing, we conducted a related-samples Wilcoxon signed-rank test comparing participants' responses to the items about refereeing and the analogous items about reading published proofs in journals, again using a Bonferroni adjustment to set an alpha level for significance at .005. We found significant differences for the purposes for which they read proofs (P1, P2), how they read proofs (M3, M5), and the weight they gave to cultural or social factors when they read proofs (C3). Not surprisingly, participants were more likely to check proofs they were refereeing for correctness (P1) and were less interested in identifying methods they could use for their own work (P2). Participants were more likely to check every line in a proof that they were refereeing (M5). They were also less likely to believe a proof was correct because it was written by a trustworthy authoritative source (C3). These findings suggest that reading a published proof of a colleague and refereeing a proof for publication are substantially different activities, where mathematicians use different processes to achieve different goals.

Finally, we note that the foils that we used again achieved our desired goals, with the majority of participants disagreeing with each of the three foils. However, we were surprised that 22% of mathematicians agreed with F3, indicating that they sometimes would check every reference of a proof that they refereed.

4. Discussion

On reading published proofs

We believe the noteworthy findings from this survey support the following hypotheses from Weber and Mejia-Ramos (2011) about how mathematicians read published proofs. First, a significant majority of mathematicians (74%) agreed that they sometimes do not check proofs in published journals for correctness (P1) and most (90%) agreed that a primary reason for reading published proofs is to identify methods that might be useful in their own work (P2).

Second, our survey data supports the hypothesis that mathematicians consider the cultural history of a proof when evaluating the proofs they read in a journal. Indeed, 72% of participants declared to be commonly convinced that a proof is correct because it appeared in a reputable journal (C1), 83% agreed to sometimes being highly confident that a proof is correct if it came from an authoritative and reliable source (C3), and 67% said they considered the quality of the journal when evaluating the correctness of the proof (C2). These data challenge Shanahan, Shanahan, and Misischia's (2011) claim that mathematicians treat proofs as context-free text where all that matters is the text on the page. Instead, like other scientists, mathematicians seem to use the source of an argument as a critical factor in determining its correctness (Weber & Mejia-Ramos, 2013b).

Third, the data provide evidence that mathematicians try to understand the proof in terms of its main ideas (91%, M1) and overarching methods (77%, M4). In fact, mathematicians claimed that understanding the methods is sometimes sufficient to judge the proof to be correct (75%, M3), sometimes to the point that they do not check that every line of the proof is correct (77%, M5).

Further, these data suggest that participants try to understand a proof in terms of its main ideas *before* performing a line-by-line check. More than 90% of participants indicated that they sometimes skimmed lengthy proofs to understand their main ideas prior to reading it line-by-line (M2), both when they read a proof in a respected journal and when then referee. Based on their data from an eye-tracking study, Inglis and Alcock (2012) questioned Weber's (2008) finding that mathematicians skimmed a proof at the beginning of a proof validation attempt. The data presented here rule out the possibility that the mathematicians in Weber's (2008) study who claimed to initially skim a proof were outliers that were not representative of the mathematical community. Whether mathematicians actually skim proofs or whether they only claim to do so remains an open question (Weber & Mejia-Ramos, 2013a; Inglis & Alcock, 2013).

Finally, these survey data provides further evidence that for mathematicians the consideration of examples plays a pivotal role in understanding a proof and gaining confidence that it is correct. The strongest statement indicates empirical evidence is sometimes sufficient for the majority of participants to accept a claim in a proof as correct (56%, E4), although it is possible that they were treating these examples generically and consequently saw this as a deductive justification.

As argued by Weber and Mejia-Ramos (2011), the empirical study of expert mathematical practices has significant implications for mathematics education, particularly with regards to proving practices at the university level. Several researchers have argued that mathematics instruction should have students engage in justification and argumentation activities in a manner that is consistent with how mathematicians perform these activities (Moschkovich, 2002; Harel & Sowder, 2007; RAND Mathematics Study Panel, 2003), calling for more empirical research to be conducted on expert mathematical practices.

Although there are important differences between why mathematicians read the published proofs of their colleagues and why mathematics majors read proofs in their textbooks, these activities also contain important similarities. In both cases, a primary reason that the individual studies the published proofs of others is to expand his or her knowledge about a mathematical domain. Unlike mathematicians, mathematics majors do not seem to try to understand proofs in terms of their main ideas or their overarching methods (e.g., Samkoff et al., submitted) and many are reluctant to resolve difficulties they experience with reading a proof by the consideration of examples (e.g., Weber, 2009). If undergraduates attempted to do these things when studying proofs, their comprehension of these proofs might improve.

On refereeing proofs for publication

The purpose of Weber and Mejia-Ramos (2011) did not concern mathematicians' behavior when refereeing a proof. Hence the character of survey questions on this topic is of an exploratory character. Geist et al. (2010) contended that "many mathematicians accept results from the literature as black boxes in their own research" (p. 158) and Auslander (2008) noted that "this is the case even if we haven't read the proof, and even if we don't have the background to follow the proof" (p. 64). Our data support these claims. Given the importance on trusting the work of other mathematicians, Geist et al. (2010) argued that investigating the mathematical reviewing process is central to understanding and assessing the reliability of mathematical knowledge, but little work of this type has been done.

In her study of the disciplinary practices of mathematicians, Burton (2002, 2004, 2009) noted that while mathematicians are homogeneous about their views on certain aspects of their practices (e.g. the importance of making connections between different areas of mathematics), they are heterogeneous in their observations about other aspects of their work (e.g. the importance of aesthetics in mathematics). Similarly, our data shows that while there is a high level of agreement among mathematicians regarding the ways in which they read the published proofs of their colleagues, there is considerable disagreement regarding the ways in which they referee proofs for publication. This implies that we need to be wary of making broad generalizations regarding disciplinary practices in mathematics, based on the views and behavior of a small sample of mathematicians.

Some mathematicians have suggested that the reviewing process in mathematics is less rigorous than is commonly believed. For instance, Nathanson (2008) claimed that, "Many (I think most) papers in most refereed journals are not refereed… some referees check line by line, but many do not". Our data also support these claims. Fallis (2003) noted that readers of mathematical proofs often leave *untraversed gaps*—gaps in a proof that are intentionally left unchecked by the reader. Geist et al. (2010) concluded that this implies that some published and accepted proofs contain gaps that were not checked by any mathematician (what Fallis called *universally untraversed gaps*). Our data offer qualified support for this position, with many mathematicians saying they sometimes do not check every line in a proof that they referee, but the majority disagreeing with this assertion.

We tentatively suggest that the types of evidence used to validate a proof depend upon the importance of the theorem and the context of the sub-claim presented. Theorems of fundamental importance, such as Perlman's purported proof of the Poincare conjecture, will be checked at a line-by-line level by multiple mathematicians, but every year, hundreds of less important theorems will be checked by a single referee (which Geist et al. argue is standard journal practice in mathematics), many of whom will not read every line of the proof. This suggests that most published proofs have a lower standard of acceptance (i.e., a lower probability threshold that the proof is valid) and implies that gaining absolute certainty in the truth of a theorem or the validity of a theorem is not the goal of the reviewing and publication process most of the time. Indeed, many published proofs do contain errors (e.g., Davis, 1972; Hanna, 1991). For these less important proofs, many mathematicians will use different types of evidence that they believe are reliable in that context. For instance, in Weber (2008), one mathematician was confident that an assertion was correct after checking for several examples, but indicated that he only did this since these types of checks were reliable in the context of number theory; he would not do the same in topology, his area of research. Similarly, our data suggest that some mathematicians do not feel the need to do a line-by-line check if they are convinced the main methods of the proof are correct. We conjecture that mathematicians are aware of which main methods can be judged to be usually correct and which methods are more problematic and potentially contain subtle mistakes; they likely read the details of the latter methods more closely. Consequently, in the refereeing process, high levels of mathematical confidence are developed from the strategic use of deductive, empirical, and testimonial evidence depending upon the context being studied and the level of confidence that is desired.

Finally, Mejia-Ramos and Inglis (2009) hypothesized that proof validation (reading a purported proof with the goal of checking its validity) and proof comprehension (reading a proof with the goal of understanding it) are mathematical activities that engender different behaviors and, as such, should not be lumped together by mathematics education researchers studying the ways in which students and mathematicians read proofs. The data presented in Table 3 offer support for this hypothesis. The data demonstrate that to mathematicians, the way in which they read a published proof of a colleague (an activity they more closely associate with reading for comprehension) differs significantly from the way in which they referee a proof for publication (an activity they more closely associate with validation). This also challenges Geist et al.'s (2010) suggestion that for many mathematicians refereeing a proof does not involve validation; indeed, 76% of our participants disagreed that they did not check a proof for correctness when refereeing.

While we distinguished between reading a published proof and refereeing a proof for publication, it might be useful in future research if different proof reading contexts and purposes were further categorized (e.g., do participants read a proof to learn new methods in the same way they read a proof for aesthetic appreciation?). We agree with Mejia-Ramos and Inglis (2009) that mathematics education researchers need to be aware of these differences when designing studies on, and discussing behaviors about, the reading of mathematical proofs. Finally, we also believe it would be worthwhile for undergraduates to be aware of the variety of goals and methods associated to mathematicians' proof reading, as this would help these students have a better understanding and appreciation for this fundamental activity in advanced mathematics.

References

Afflerbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Literacy Research, 16*(4), 307-322.

Auslander, J. (2008). On the roles of proof in mathematics. In B. Gold and R. A. Simons (Eds.), *Proof & Other Dilemmas: Mathematics and Philosophy* (pp. 61–77). Mathematical Association of America.

Azarello, F. (2007). The proof in the 20th century: From Hilbert to automatic theorem proving. In P. Boero (Ed.), *Theorems in schools: From history, epistemology, and cognition to classroom practice*. Rotterdam: Sense Publishers.

Burton, L. (2002), Recognising commonalities and reconciling differences in mathematics education. *Educational Studies in Mathematics*, *50*(2), 157-175.

Burton, L. (2004). *Mathematicians as enquirers: Learning about learning mathematics.* Dordrecht: Kluwer.

Burton, L. (2009). The Culture of Mathematics and the Mathematical Culture, in O. Skovsmose, P. Valero and O. R. Christensen (Eds.), *University Science and Mathematics Education in Transition* (pp. 157-173). New York: Springer.

Davis, P. (1972). Fidelity in mathematical discourse: Is one and one really two? *American Mathematical Monthly*, *79*, 252–263.

Ericsson, K.A., & Simon, H.S. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251.

Fallis, D. (2003). Intentional gaps in mathematical proofs. *Synthese, 134*, 45-69.

Geist, C., Löwe, B., & Van Kerkhove, B. (2010). Peer review and knowledge by testimony in mathematics. In B. Löwe and T. Müller (Eds.), *Philosophy of Mathematics: Sociological Aspects and Mathematical Practice* (pp.155-178). London: College Publications.

Hanna, G. (1991). Mathematical proof. In D. Tall (Ed.), *Advanced mathematical thinking.* Dordrecht: Kluwer.

Harel, G. (1998). Two dual assertions: The first on learning and the second on teaching (or vice versa). *American Mathematical Monthly, 105*, 497-507.

Heinze, A. (2010). Mathematicians' individual criteria for accepting theorems as proofs: An empirical approach. In G. Hanna, H.N. Jahnke, & H. Pulte (Eds.), *Explanation and Proof in Mathematics: Philosophical and Educational Perspectives* (pp. 101-111). New York: Springer.

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

Inglis, M., & Alcock, L. (2013). Skimming: A response to Weber and Mejia-Ramos. *Journal for Research in Mathematics Education,44,* 471-474.

Inglis, M., & Mejia-Ramos, J. P. (2009). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction, 27*, 25-50.

Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition, 71,* 191-229.

Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35–60). San Diego: Academic Press.

Mejia-Ramos, J. P., & Inglis, M. (2009). Argumentative and proving activities in mathematics education research, in F.-L. Lin, F.-J. Hsieh, G. Hanna, and M. de Villiers (Eds.), *Proceedings of the ICMI Study 19 conference: Proof and Proving in Mathematics Education* (Vol. 2, pp. 88-93), Taipei, Taiwan.

Moschkovich, J. N. (2002). An introduction to examining everyday and academic mathematical practices. In J. N. Moschkovich, and M. E. Brenner (Eds.), *Everyday and Academic Mathematics in the Classroom, Journal for Research in Mathematics Education Monograph* (Vol. 11, pp.1-11). Reston, VA: NCTM

Nathanson, M. (2008). Desperately seeking mathematical truth. *Notices of the American Mathematical Society, 55*(7), 773.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Porteous, K. (1986). Children's appreciation of the significance of proof. In *Proceedings of the Tenth International Conference for the Psychology of Mathematics Education* (pp. 392-397). London, England.

RAND Mathematics Study Panel (2003). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education.* Santa Monica, CA: RAND Corporation.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica, 7*, 5–41.

Reips, U. D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89–117). San Diego: Academic Press.

Rowland, T. (2001). Generic proofs in number theory. In S. Campbell and R. Zazkis (Eds.), *Learning and teaching number theory: Research in cognition and instruction* (pp. 157-184). Westport, CT: Ablex Publishing.

Samkoff, A., Weber, K., & Mejia-Ramos, J. P. (submitted). What's the big idea?: Mathematicians' and undergraduates' proof summaries.

Selden, A., & Selden, J. (2003). Validations of proofs written as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education, 36*(1), 4–36.

Shanahan, C., Shanahan, T., & Misischia, C. (2011). Analysis of expert readers in three disciplines: History, mathematics, and chemistry. *Journal of Literacy Research, 43*, 393-429.

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

Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning, 12*, 306-336.

Weber, K. (2009). Mathematics majors' evaluation of mathematical arguments and their conception of proof. In *Proceedings of the 12th Conference for Research in Undergraduate Mathematics Education*. Available for download from: http://sigmaa.maa.org/rume/crume2009/proceedings.html

Weber, K. & Mejia-Ramos, J. P. (2011). Why and how mathematicians read proofs: An exploratory study. *Educational Studies in Mathematics, 76*, 329-344.

Weber, K., & Mejia-Ramos, J. P. (2013a). On mathematicians' proof skimming: A reply to Inglis and Alcock. *Journal for Research in Mathematics Education, 44,* 464-471.

Weber, K., & Mejia-Ramos, J. P. (2013b). The influence of sources in the reading of mathematical text: A reply to Shanahan, Shanahan, and Misischia. *Journal of Literacy Research*, *45*(1), 87-96.

Table 1. Survey statements.

Purposes:
P1: … it is not uncommon that I do not check the proof for correctness. Rather I read the proof to gain some other type of insight.
P2: … it is not uncommon that an important reason for my reading the proof is to gain some insights into how I can solve problems that I am working on.

Examples
E1: … it is not uncommon for me to see how the steps in the proof apply to a specific example. This helps me understand the proof.
E2: … it is not uncommon for me to see how the steps in the proof apply to a specific example. This increases my confidence that the proof is correct.
E3: … and I am not immediately sure that a statement in the proof is true, it is not uncommon for me to increase my confidence in the statement by checking it with one or more carefully chosen examples.
E4: … and I am not immediately sure that a statement in the proof is true, it is not uncommon for me to gain a sufficiently high level of confidence in the statement by checking it with one or more carefully chosen examples to assume the claim is correct and continue reading the proof.

Proof as application of methods:
M1: … it is not uncommon for me to try to understand the proof in terms of its main ideas and not only in terms of how each step is justified.
M2: … it is not uncommon that I skim the proof first to comprehend the main ideas of the proof, prior to reading the proof line-by-line.
M3: … it is not uncommon for me to judge the proof to be correct if I am sure the main idea or method is correct (as opposed to having to check every line of the proof).
M4: … it is not uncommon for me to try to understand the proof in terms of the overarching method the author used.
M5: … if I understand the main idea of the proof and think it is correct, it is not uncommon that I do not check that every line of the proof is correct, but trust that the logical details are correct.

Proof as cultural artifact:
C1: It is not uncommon for me to believe that a proof is correct because it is published in an academic journal.
C2: … it is not uncommon for the quality of the journal to increase my confidence that the proof is correct.
C3: … it is not uncommon for me to be very confident that the proof is correct because it was written by an authoritative source that I trust.

Foil:
F1: … it is not uncommon that I gain confidence that the proof is correct because the author cited my work.
F2: … it is not uncommon that an important reason that I read proofs is to explore the writing styles of academics from different countries.
F3: … it is not uncommon that I also read each of the references cited by the author.

Note: Except for statements M2 and C1, statements related to reading the published proofs of others began with: "When I read a proof in a respected journal, …" In this first section of the survey, statement M2 began with: "When I read a lengthy proof in a respected journal, …"
Except for statement M5, statements related to refereeing a proof for publication began with "When I referee a manuscript, …" In this second section of the survey, statement M5 simply said: "When I referee a manuscript, it is not uncommon that I do not check that every line in the proof is correct." Statements M1, M4, C1, and C2 were not included in this second section of the survey.

Table 2. Mathematicians' level of agreement with statements related to reading the published proofs of others (*N*=118).

Statement	Agree	Neutral	Disagree	*p*-value of one-sample Wilcoxon signed-rank test
P1*	74%	14%	13%	<.001
P2* +	90%	7%	3%	<.001
E1*	81%	8%	11%	<.001
E2*	83%	8%	9%	<.001
E3*	79%	13%	8%	<.001
E4*	56%	30%	14%	<.001
M1*	91%	3%	6%	<.001
M2*	92%	4%	3%	<.001
M3*	75%	15%	10%	<.001
M4*	77%	18%	5%	<.001
M5*	77%	14%	9%	<.001
C1* +	72%	16%	12%	<.001
C2*	67%	16%	17%	<.001
C3*	83%	10%	7%	<.001
F1*	6%	40%	54%	<.001
F2*	5%	7%	88%	<.001
F3*	11%	11%	78%	<.001

*- Indicates the one-sample Wilcoxon signed-rank test reveals the median response differed significantly from "neutral" with an alpha-level of p=.003.
+- Indicates a Mann-Whitney reveals significantly different response patterns from participants with over six years experience and participants with under six years experience with an alpha-level of p=.003.
(The order of the questions: C1-C2-P1-F1-E2-C3-P2-F2-M1-M2-F3-M3-M4-E1-M5-E3-E4)

Table 3. Mathematicians' level of agreement with statements related to refereeing a proof for publication (N=54).

Statement	Agree	Neutral	Disagree	p-value of one-sample Wilcoxon signed-rank test	p-value of related-samples Wilcoxon signed-rank test
P1^{*+}	13%	11%	76%	<.001	<.001
P2^{+}	28%	28%	44%	.041	<.001
E1*	93%	4%	4%	<.001	.827
E2*	74%	22%	4%	<.001	.021
E3*	83%	11%	6%	<.001	.857
E4	43%	30%	28%	.285	.026
M2*	93%	6%	2%	<.001	.035
M3^{+}	39%	17%	44%	.169	<.001
M5^{+}	35%	13%	52%	.040	<.001
C3^{+}	39%	20%	41%	.578	<.001
F1^{*+}	4%	9%	87%	<.001	<.001
F2*	0%	9%	91%	<.001	.052
F3^{+}	22%	26%	53%	.009	.001

*- Indicates that a one-sample Wilcoxon signed-rank test reveals the median response differed significantly from "neutral" with an alpha-level of p=.005.

+- Indicates that a related-samples Wilcoxon signed-rank test reveals different response patterns by the participants on the refereeing items and the analogous items for reading a proof in a journal with an alpha-level of p=.005.

(The order of the questions: P1-F1-C3-P2-F2-M2-F3-M3-E1-E2-M5-E3-E4)