Brief Reports

On Mathematicians' Proof Skimming:
A Reply to Inglis and Alcock

Keith Weber and Juan Pablo Mejía-Ramos
Rutgers University

In a recent article, Inglis and Alcock (2012) contended that their data challenge the claim that when mathematicians validate proofs, they initially skim a proof to grasp its main idea prior to reading individual parts of the proof more carefully. This result is based on the fact that when mathematicians read proofs in their study, on average their initial reading of a proof took half as long as their total time spent reading that proof. However, inferring individuals' strategy use by averaging their reaction times across multiple trials is problematic. We present a more fine-grained analysis of Inglis and Alcock's data that suggests that mathematicians frequently used an initial skimming strategy when engaging in proof validation tasks.

Key words: Mathematicians, Proof reading, Proof validation

In a recent article published in the *Journal for Research in Mathematics Education*, Inglis and Alcock (2012)[1] analyzed several hypotheses about undergraduates' and mathematicians' proof validation behaviors by studying their eye movements when validating proofs. We believe this research represents a significant advance, both in terms of its novel methodology and its findings. In the article, Inglis and Alcock challenged conjectures that we have made in our research. Specifically, they asserted that mathematicians do not zoom out before validating a proof in a line-by-line manner. In this reply, we challenge that assertion and contend that Inglis and Alcock's data actually provides evidence to the contrary.

---

[1] All Inglis and Alcock references are to the 2012 article, to which we are responding.

The Inglis and Alcock Assertion

Inglis and Alcock challenged claims that we have made with regard to mathematicians' proof reading. In Weber and Mejía-Ramos (2011), we claimed that mathematicians might understand a proof as the application of methods, focusing on the high-level structure of the proof and thinking carefully about not the individual inferences, but the main ideas. We metaphorically referred to this type of understanding as *zooming out* to emphasize that mathematicians would comprehend the proof in terms of its big picture or overarching ideas, rather than its logical details (p. 340). In Weber (2008), the first author observed that some mathematicians claimed they would first check the structure of the proof to obtain an overview of the argument before proceeding to a line-by-line verification (p. 440). We describe these mathematicians as *initially skimming* the proof prior to reading it carefully.

We argue that zooming out and initially skimming are not synonymous. The claim that mathematicians zoom out to understand a proof in terms of its main ideas relates to their comprehension of a proof, whereas the claim about initial skimming relates to the process by which a proof is read. Indeed, in Weber and Mejía-Ramos (2011), we emphasized that zooming out is not a claim about the process of reading a proof, and that we do not know the processes that mathematicians use to zoom out, stating, "It is difficult to say how mathematicians 'zoom out' by encapsulating particular strings of inferences of a proof into methods" (p. 341).

To test the viability of our hypotheses, Inglis and Alcock observed the eye movements of 12 mathematicians as they determined if six proofs were valid. For each proof, the researchers computed a ratio comparing the amount of time that elapsed before the mathematician's first fixation on the last line of the proof to the total amount of time the participant spent reading the proof. From here on, we refer to this ratio as the Initial Reading ratio, or the IR ratio. The mean IR ratio for these mathematicians was 50%. Based on these data, Inglis and Alcock rejected our conjecture that mathematicians use zooming-out strategies because "50% is nonetheless a higher figure than would be expected with an initial zooming-out strategy" (p. 372).

We do not believe that Inglis and Alcock's analysis assesses the viability of whether mathematicians are using a zooming-out strategy. Zooming out refers to understanding the proof in terms of its overarching methods, rather than at the level of particular inferences, and we do not believe it is necessary for mathematicians to read a proof twice to obtain this understanding. For instance, in describing how he read a proof, Thurston (1994) wrote, "I might look over several paragraphs or strings of equations and think to myself, 'Oh yeah, they're putting in enough rigmarole to carry out such-and-such an idea'" (p. 167). In this sense, Thurston describes how he zoomed out, or thought about the proof in terms of methods rather than individual inferences, without necessarily engaging in a second reading of the proof (or, indeed, even completing a first reading). Instead, we believe it would be more appropriate to interpret Inglis and Alcock as testing our conjecture that mathematicians initially skimmed a proof prior to studying it

in detail. As skimming involves the process of studying a proof, we concur with Inglis and Alcock that this issue can be addressed with their eye-tracking study. Nonetheless, for several reasons, we feel the data that Inglis and Alcock obtained do not refute the claim that mathematicians engage in skimming. The purpose of this reply is to articulate these reasons.

The Perils of Identifying Strategies by Aggregating Data

We are not convinced that an IR ratio of greater than 0.5 in Inglis and Alcock's study provides evidence that mathematicians do not use a skimming strategy, for two reasons. First, if a mathematician did an initial skim through a proof to determine its structure and encountered an error, he or she might not feel the need to reread the proof carefully; indeed, the mathematician might not even read the last line of the proof. [2] Second, as Inglis and Alcock acknowledged, the proofs in their study were very short, with one proof being only four lines long, which may have prevented skimming from being observed. However, for the sake of argument, we will accept Inglis and Alcock's operationalization of a skimming strategy and agree that those skimming would exhibit an IR ratio of less then 0.5.

In an influential paper, Siegler (1987) argued that one should not infer what strategies individuals use to complete tasks by aggregating their reaction time data across different tasks. He noted, "If a person uses different strategies on different trials, averaging data generated by those strategies can distort conclusions about numerous aspects of performance" (p. 250) because "just as data aggregated over people may not accurately reflect the behavior of any person (Estes, 1956), so data aggregated over strategies may not accurately reflect the characteristics of any strategy" (p. 250). Siegler demonstrated the dangers of engaging in such analysis by examining the strategies that children use to complete addition tasks. Both for individual and groups of children, their aggregate reaction time patterns suggest they were using the "min" strategy, computing a sum by adapting a counting-on strategy from the larger addend. Because of these data patterns, there was a consensus among developmental psychologists that children predominantly used a min strategy to complete addition tasks. However, Siegler conducted a study revealing that children actually were using five different strategies and no single strategy more than 40% of the time. Collapsing across these strategies happened to lead to a mean reaction time pattern that would be consistent with the min strategy, but that did not imply the min strategy was always being used, and it certainly did not imply the absence of other strategies. As Inglis and Alcock analyzed their data by examining participants' mean IR ratios across different trials, their analysis has a similar threat to validity.

As a thought experiment, imagine a mathematician completing Inglis and Alcock's study. For three of the proofs, she spots an error in her first reading of

---

[2] There were four trials in which mathematicians in Inglis and Alcock's study judged a proof to be invalid without reading the last line of the proof.

the proof and then judges these proofs to be invalid. This would not be too surprising given the blatant errors in some of the proofs used in the study. Her IR ratio for these proofs would be high, say 0.9. For the other three proofs, she adopts a skimming strategy, reading each proof in a minute and then spending four minutes studying each of the inferences within the proof. Her IR ratio for these proofs would be 0.2. Her aggregate ratio score in this study would be 0.55, which Inglis and Alcock would categorize as strong evidence that she was not skimming. However, the descriptive account above illustrates that she was clearly using a skimming strategy for at least three of the proofs and arguably all six. This thought experiment is only a plausibility argument,[3] but we will demonstrate that such an account can also explain Inglis and Alcock's actual data.

A Fine-Grained Analysis of Inglis and Alcock's Data

For the purposes of this section, we describe two hypothetical reading strategies that an individual may use when validating a proof. An individual using an *initial skim strategy* would first read the proof relatively quickly to get a feel for the methods used in the proof without seeking a high level of conviction that the proof was correct. The high level of conviction would be gained by studying the proof after it was initially skimmed. An individual using an *initial careful reading strategy* would try to obtain a high level of conviction that the proof was correct the first time he or she read through it. For the sake of argument, we accept Inglis and Alcock's assumption that the application of the initial skim strategy (from here on, IS strategy) would yield an IR ratio, on average, of substantially less than 0.5. However, we would argue that an application of the initial careful reading strategy (from here on, ICR strategy) should yield an IR ratio substantially greater than 0.5, as individuals should be quicker in subsequent passes through the proof if they checked it carefully the first time.

To see if mathematicians were using an IS strategy, Inglis and Alcock assigned an IR ratio to each mathematician by averaging his or her IR ratios across the six proofs that he or she validated. They then computed the average of the 12 mathematicians' mean IR ratios, obtaining a value of .50. This process is mathematically equivalent to simply averaging the IR ratios across every trial that a mathematician completed. As we note above, we disagree with this analysis, as this is combining mathematicians' IR ratios across different tasks. This analysis ignores the possibility that mathematicians may have been using different reading strategies for the different proofs that they read. Nonetheless, we will argue that this distribution of IR ratios suggests that mathematicians are using two different reading strategies to complete their validations.

---

[3] Wyatt and colleagues' (1993) research on social scientists' domain-specific reading lends credence to the plausibility that there is variance in researchers' use of a skimming strategy. In summarizing the results of their study, Wyatt et al. stated, "Approximately half of the readers surveyed the text before reading it (8 out of 15) and the remaining readers (*n* = 7) did not" (p. 55). If social scientists can selectively use a skimming strategy when reading domain-specific papers, it is plausible that mathematicians might do so as well.

After reading Inglis and Alcock's study, we contacted the authors and asked if we could have access to their data, which they quickly and generously provided. A histogram presenting the distribution of mathematicians' IR ratios across trials is presented in Figure 1.[4]

The mathematicians' mean IR ratio across these trials was .50, which is neither consistent with an IS strategy nor an ICR strategy (as described above). It is possible that mathematicians were using a third strategy to read their proofs. If mathematicians were collectively using the same strategy for each proof that they read, we would expect to see the distribution of their IR ratios to be unimodal and distributed around .50. Figure 1 reveals that this is not the case.

The data in Figure 1 are not normally distributed, and a Shapiro-Wilks normality test rejects the hypotheses that mathematicians' IR ratios were sampled from a normal distribution ($W = .958$, $p = .02$). The sample distribution has an excess kurtosis of -1.01, with this high negative value being consistent with (although not necessarily implying) the distribution being bimodal. A visual inspection of Figure 1 also suggests a bimodal distribution.

One interpretation of Figure 1 is that these IR ratios come from two distributions, a larger distribution with a mean of about .35 and a smaller distribution with a mean of .80. As support for this interpretation, we note that there were 48 IR ratios less than .63, 20 IR ratios greater than .73, and no IR ratios between .63 and .73. If we restrict our analysis to the 48 IR ratios less than .63 (essentially the first seven bars of Figure 1), this distribution has a mean of .348 and passes a Shapiro-Wilks normality test ($W = .975$, $p = .38$), so we cannot conclude that the distribu-
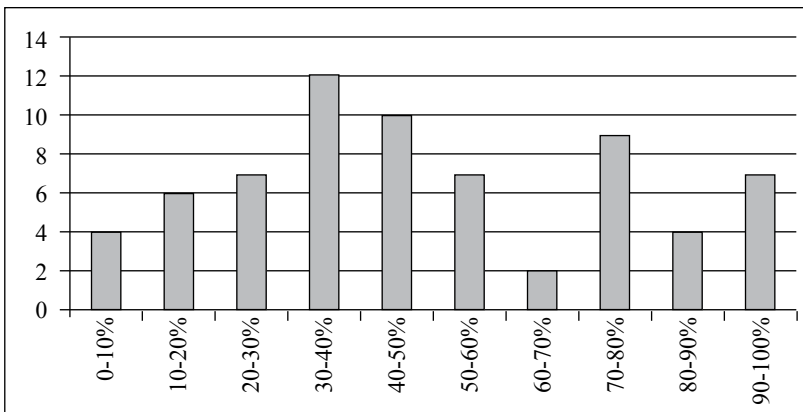

Figure 1. Histogram presenting distribution of mathematicians' IR ratios.

---

[4] Although the 12 mathematicians collectively read 72 proofs, the histogram contains only 68 IR ratios. For four trials, an IR ratio could not be computed as the participant never fixated on the last line of the proof. Inglis and Alcock did not include these trials in their analysis of IR ratios. For the sake of consistency, we also did not use these trials in our analysis.

tion of this data departs from normality. As the mean of this distribution is substantially less than .5, the cumulative IR ratio is consistent with the account that this distribution was generated from mathematicians using an IS strategy. We are not saying that this account is necessarily correct, only that the mathematicians' IR ratio distribution can hardly be considered evidence that mathematicians were not frequently using an IS strategy.

An important weakness of this analysis is that, like Inglis and Alcock, we combined the IR ratios of mathematicians across different tasks. To address this weakness, we conclude this section by focusing on mathematicians' strategies for reading Proof 5 from Inglis and Alcock's study. We chose this proof as we believe it to be the one where skimming was most likely to occur, due to its length, complexity, and the absence of blatant errors. The distribution of mathematicians' IR ratios on this task is presented in Figure 2. If mathematicians were not using an IS strategy to complete this task, we would expect few IR ratios to be substantially less than .50. However, eight of the 12 mathematicians had IR ratios less than .35. We believe that Figure 2 is strikingly consistent with our account that mathematicians are using two strategies to validate proofs, with the more common strategy being the IS strategy. Indeed, it is difficult to look at Figure 2 and conclude that no mathematician was using an IS strategy.

In conclusion, Inglis and Alcock claimed mathematicians were not using an IS strategy, but this is because they simply averaged mathematicians' IR ratios without carefully considering how the IR ratios were distributed. As Siegler (1989) demonstrated, this type of analysis is problematic, as participants might have been using different strategies for different tasks. Inspecting the distribution of the IR ratios, both cumulatively and for Proof 5 (the proof we argue is the most appropriate to search for evidence of mathematicians' proof skimming), revealed that mathematicians might be using two different strategies. In both Figure 1 and
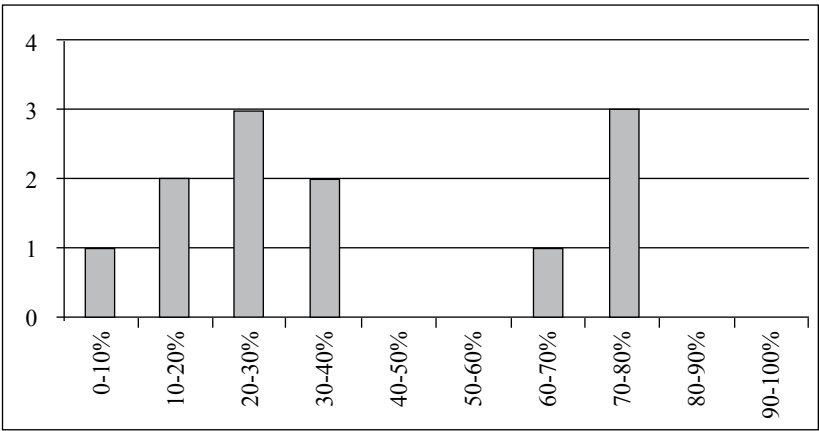

Figure 2. Histogram presenting distribution of mathematicians' IR ratios for Proof 5.

Figure 2, we found evidence of bimodal distributions, with the larger part of the distribution having a mean of substantially less than .5 and the less common one having a mean of substantially greater than .5. If we assume that using an IS strategy to validate a proof would yield an IR ratio of substantially less than .5 and an ICR strategy would yield an IR ratio of substantially greater than .5, Figure 1 and Figure 2 are both consistent with the claim that mathematicians use two initial reading strategies to validate proofs, with the more common one being an IS strategy and the less common one being an ICR strategy. Of course, a demonstration of consistency does not necessarily mean that our account is correct, but it does imply that Inglis and Alcock failed to refute our conjecture that mathematicians skim proofs when validating.

Implications for Future Research

Inglis and Alcock presented data that they believed challenged our claims that mathematicians use skimming strategies when validating proofs. We argue that their data actually provides support for our claims. We conclude by discussing what might be learned from Inglis and Alcock's data about skimming and what research questions our analysis suggests.

First, Inglis and Alcock's data challenge the assertion that mathematicians use an IS strategy with every proof that they read. Although we have argued that a high IR ratio does not necessarily imply the absence of an IS strategy, we did note that one mathematician judged Proof 5 to be valid, reading each line sequentially, with an IR ratio of 0.74. It is difficult to claim that this mathematician was using an IS strategy given his or her eye-movement data. In both Weber (2008) and Inglis and Alcock, the authors treated the question, "Do mathematicians skim proofs when validating?" as a dichotomous question. If our conjecture that mathematicians may often, but not always, use an IS strategy is correct, this suggests that a more appropriate research question would be, "When do mathematicians skim proofs for validation purposes?"

Second, we believe a weakness both to Inglis and Alcock's and our analysis is that all of the data came from proofs that were rather short; even Proof 5, which we chose to focus on, was only eight lines long, and one proof was only four lines long. Conducting similar studies with substantially longer proofs may provide better insight into mathematicians' actual professional practice with respect to proof reading.

Third, although we found evidence that mathematicians were using an IS strategy, we suggested that their mean IR ratios when this strategy was applied was about .35. This is substantially less than .50, but still perhaps greater than one would predict for someone using a skimming strategy. If a mathematician spent, say, three hours reading a proof, it might seem strange to call her first pass through the proof as skimming if it took an hour to complete. It might be the case that the .35 estimate is artificially high due to the short proofs in the Inglis and Alcock study. Support for this hypothesis comes from mathematicians' IR ratios on Proof 5, where half of the mathematicians had IR ratios of less than 0.3. It is possible

that if mathematicians read more proofs such as these, lower IR ratios would have been observed. Nonetheless, these data do suggest that we may not fully understand what mathematicians mean when they say they skim proofs. If so, this suggests that prior to conducting eye-tracking studies to see if mathematicians actually do as they allege when validating proofs, it might be more worthwhile to first understand what mathematicians mean when they say they skim proofs, including what they would consider evidence for or against the claim that a mathematician was engaging in skimming.

References

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education, 43*(4), 358–391. doi:10.5951/jresemath educ.43.4.0358

Siegler, R. (1987). The perils of averaging data over strategies: The case of children's addition. *Journal of Experimental Psychology: General, 116*(3), 250–264. doi:10.1037/0096-3445.116.3.250

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education, 39*(4), 431–459.

Weber, K., & Mejía-Ramos, J. P. (2011). Why and how mathematicians read proofs: An exploratory study. *Educational Studies in Mathematics, 76*(3), 329–344. doi:10.1007/s10649-010-9292-z

Wyatt, D., Pressley, M., El-Dinary, P., Stein, S., Evans, P., & Brown, R., (1993). Comprehension strategies, worth and credibility monitoring, and evaluations: Cold and hot cognition when experts read professional articles that are important to them. *Learning and Individual Differences, 5*(1), 49–72. doi:10.1016/1041-6080(93)90026-O

Authors

Keith Weber, Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ 08901; keith.weber@gse.rutgers.edu

Juan Pablo Mejía-Ramos, Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ 08901; pablo.mejia@gse.rutgers.edu

Skimming:
A Response to Weber and Mejía-Ramos

Matthew Inglis and Lara Alcock
Loughborough University, United Kingdom

We recently reported a study in which undergraduate students and research mathematicians were asked to read and validate purported proofs (Inglis & Alcock, 2012). In our eye-movement data, we found no evidence of the initial skimming strategy hypothesized by Weber (2008). Weber and Mejía-Ramos (2013) argued that this was due to a flawed analysis of eye-movement data and that a more

fine-grained analysis led to the opposite conclusion. Here we demonstrate that this is not the case, and show that their analysis is based on an invalid assumption.

Weber and Mejía-Ramos (2013) suggested that our analysis was flawed because, after calculating what proportion of reading time a mathematician took to reach the last line of a proof (which they called an Initial Reading [IR] ratio), we took means across different tasks. Considering means, they argued, obscures reading strategy variation. Clearly, this is true in principle, and at the end of this response, we discuss what exactly is obscured in our data. First, however, we respond to Weber and Mejía-Ramos's more specific criticisms.

Neither by Subjects Nor by Items

Weber and Mejía-Ramos argued that our parametric analysis was invalid because the distribution of IR ratios is not normal; however, they took task-individual pairs as their unit of analysis. Our unit of analysis was different, as we conducted a traditional by-subjects analysis, finding that the distribution of participants' mean IR ratios *was* approximately normal. Thus our analysis, with inference to the mean time it takes mathematicians to first fixate on the last lines of these purported proofs, is valid.

Knowing this, of course, does not tell us directly about reading strategies on individual proofs, and Weber and Mejía-Ramos aimed to investigate these strategies using single IR ratios. But their method results in a statistically problematic mixture of by-subjects and by-items analyses: It allows generalization neither to the population of mathematicians nor to the population of proofs. It does not even permit conclusions about the IR-ratio distribution in the population of task-individual pairs, because the sampling then grossly violates the assumption of independence (each participant appears in six task-individual pairs; each task
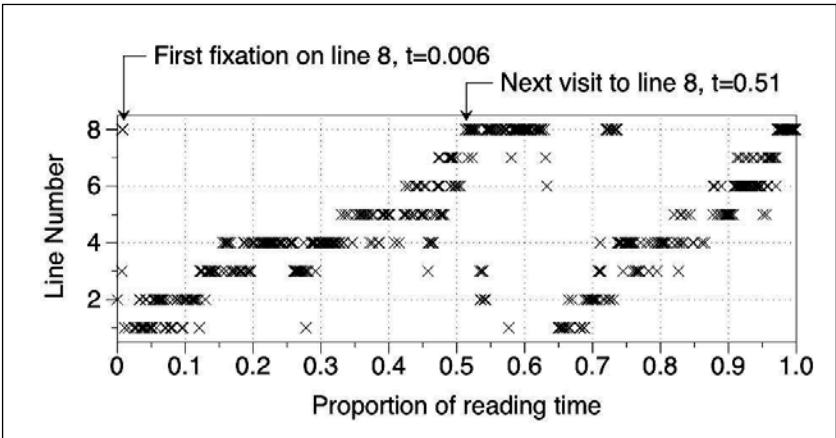

Figure 1. One mathematician's time versus line-number fixation plot for Proof 5. A point at (x, y) indicates that the participant fixated on line y at time x.

in 12). Thus, single IR ratios do not allow us to make desirable inferences, even though they exhibit more detail of the data.

The Absence of Fixations Versus the Presence of a Fixation

One strength of eye tracking is that it provides a lot of detail; however, a corresponding risk is that the data is noisy. The noise creates problems for Weber and Mejía-Ramos's argument that low IR ratios—early first fixations on last lines—indicate the use of a skimming strategy. Whereas we asserted that the absence of early last-line fixations implies that a skimming strategy was not used, Weber and Mejía-Ramos assumed the inverse, that the presence of early last-line fixations implies that a skimming strategy was used. This assumption is invalid. Although the absence of fixations is meaningful, the presence of a single fixation can easily be meaningless—perhaps the result of an irrelevant head movement or of returning from a blink or from an off-screen fixation. This is almost certainly the case for the very early last-line fixation in Figure 1 (from a mathematician's reading of Proof 5, as analyzed in detail by Weber and Mejía-Ramos). This fixation occurred after 0.6% of the total reading time, giving an IR ratio of 0.006. However, the point (0.006, 8) is clearly an outlier, and this participant did not skim, but rather adopted an approximately linear reading strategy for the first 60% of his or her reading time. Considering participant means reduces such misinterpretations because it decreases distortions caused by noise. Indeed, it does so in a way that is conservative: Noise can only reduce first fixation times, so "false" fixations bias the data against our conclusion.

Participant means are, of course, a crude measure of behavior. But to study that behavior directly, we can use all the eye-tracking data. Studying the time versus line-number fixation plots for Proof 5 for all twelve mathematicians[1] indicates that it is unreasonable to conclude, as Weber and Mejia-Ramos do, that half of the participants completed an initial read-through within 30% of their reading time. In fact, around 30% is the lowest proportion of reading time that any participant can be said to have taken for their initial read-through. We agree with Weber and Mejía-Ramos that to describe such behavior as skimming would "seem strange" (p. **xxx–xxx.**).

Conclusion

Weber and Mejía-Ramos reported that many mathematicians claim to skim proofs, and observed that this is surprising if the behavioral evidence suggests otherwise. We agree. We also agree that the discrepancy could indicate that we do not understand what mathematicians mean when they talk about skimming, or that this behavior appears only for more complicated proofs. We are investigating the second possibility in our ongoing work.

---

[1] High-resolution versions of these plots are available at http://hdl.handle.net/2134/10725.

References

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

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education, 39*(4), 431–459.

Weber, K., & Mejía-Ramos, J. P. (2013). On mathematicians' proof skimming: A reply to Inglis and Alcock. *Journal for Research in Mathematics Education, 44*(2), **xxx–xxx.**

Authors

Matthew Inglis, Mathematics Education Centre, Loughborough University, Loughborough, LE11 3TU, United Kingdom; m.j.inglis@lboro.ac.uk

Lara Alcock, Mathematics Education Centre, Loughborough University, Loughborough, LE11 3TU, United Kingdom; l.j.alcock@lboro.ac.uk