Investigating and Improving Undergraduate Proof Comprehension

Lara Alcock, Mark Hodds, Somali Roy, and Matthew Inglis

Undergraduate mathematics students see a lot of written proofs. But how much do they learn from them? Perhaps not as much as we would like; every professor knows that students struggle to make sense of the proofs presented in lectures and textbooks. Of course, written proofs are only one resource for learning; students also attend lectures and work independently or with support on problems. But because mathematics majors are expected to learn much of their mathematics by studying proofs, it is important that we understand how to support them in reading and understanding mathematical arguments.

This observation was the starting point for the research reported in this article. Our work uses psychological research methods to generate and analyze empirical evidence on mathematical thinking, in this case via experimental studies of teaching interventions and quantitative analyses of eye-movement data. What follows is a chronological account of three stages in our attempts to better understand students' mathematical reading processes and to support students in learning to read effectively.

In the first stage, we designed resources we called *e-Proofs* to support students in understanding specific written proofs. These e-Proofs conformed to typical guidelines for multimedia learning resources, and students experienced them as useful. But a more rigorous test of their efficacy revealed that students who studied an e-Proof did not learn more than students who had simply studied a printed proof and in fact retained their knowledge less well. This led us to suspect that e-Proofs made learning feel easier, but as a consequence resulted in shallower engagement and therefore poorer learning.

At the second stage we sought insight into possible underlying reasons for this effect by using eye-movement data to study the mechanisms of mathematical reading. We asked undergraduate students and mathematicians to read purported proofs and found that experts paid more attention to the words and made significantly more back-and-forth eye movements of a type consistent with attempts to infer possible justifications for mathematical claims. This result is in line with the idea that mathematical experts make active efforts to identify logical relationships within a proof and that effective guidance might therefore be needed to teach students to do the same thing.

Lara Alcock is a senior lecturer in the Mathematics Education Centre at Loughborough University in the UK. She was the recipient of the 2012 Annie and John Selden Prize for Research in Undergraduate Mathematics Education, and she is author of the research-based study guides How to Study as a Mathematics Major *and* How to Think about Analysis *(both published by Oxford University Press). Her email address is* l.j.alcock@lboro.ac.uk.

Mark Hodds is Mathematics Support Centre manager at Coventry University in the UK. He completed his PhD at the Mathematics Education Centre at Loughborough University with a thesis entitled "Improving proof comprehension in undergraduate mathematics." His email address is ab7634@coventry.ac.uk.

Somali Roy recently completed her PhD at the Mathematics Education Centre at Loughborough University with a thesis entitled "Evaluating a novel pedagogy in higher education: A case study of e-Proofs." Her email address is S.roy@lboro.ac.uk.

DOI: http://dx.doi.org/10.1090/noti1263

Matthew Inglis is a senior lecturer in the Mathematics Education Centre at Loughborough University, an honorary research fellow in the Learning Sciences Research Institute at the University of Nottingham, and a Royal Society Worshipful Company of Actuaries research fellow. He was the recipient of the 2014 Annie and John Selden Prize for Research in Undergraduate Mathematics Education. His email address is M.J.Inglis@lboro.ac.uk.

The present article draws on work supported by the Royal Society and by the Higher Education Academy's Maths, Stats & OR Network. Self-Explanation Training for Mathematics Students is available at www.setmath.lboro.ac.uk.

Finally, at the third stage, we produced such guidance by adapting *self-explanation training* to form a simple, generic guide to studying mathematical proofs. In a series of three studies we found that students who studied the training gave higher-quality mathematical explanations, exhibited altered eye movements that were more like those of expert mathematicians, and performed significantly better in both immediate and delayed proof comprehension tests. In the remainder of this article we explain this work in detail, giving rationales for our empirical study designs, explaining the nature of the self-explanation training, and expanding the arguments outlined here.

e-Proofs

We began by considering the challenges students face when learning from proofs presented in lectures. One problem, as we saw it, was that live explanations given in lectures are potentially ambiguous and certainly ephemeral: gestures indicating where attention should be focused can be vague, and the professor's additional explanations often go unrecorded so they are no longer available when students engage in independent study of their notes or a textbook. We set out to remedy this by taking advantage of straightforward presentation technology, constructing e-Proofs for several of the more difficult theorems in a course on real analysis (the course covered typical early material on continuity, differentiability, and integrability, with epsilon-delta definitions). Each e-Proof showed a theorem and a complete accompanying proof and was split into 8–10 screens. Each screen (see Figure 1 for an example) focused attention on particular aspects of the proof by graying out some areas and indicating links with boxes and arrows; each had a short accompanying audio file that could be played with a click. Students could navigate freely through the screens, listening to the audio and watching the animations as many or as few times as they wished (for detail see [1]).

Our e-Proofs were designed to capture the additional explanations that a professor might give in a lecture and to improve upon them by ensuring that students' attention was appropriately focused. The design features of e-Proofs meant that they conformed to guidelines typically offered as a consequence of research on multimedia educational resources: they moved some essential processing from visual to auditory channels, they allowed time between successive bite-sized segments, they provided cues to reduce processing of extraneous material, they avoided presenting identical streams of printed and spoken words, and they presented narration and corresponding animation simultaneously to minimize the need to hold representations in memory (cf. [2]). The provision of e-Proofs was popular with students, who saw them as a useful supplement to lectures. Free-form feedback on the course as a whole evoked numerous remarks of the type that are encouraging for educational innovators:

> *I found hearing the lecturer explaining each line individually helpful in understanding particular parts and how they relate to the entire proof.*

> *Having proofs online does make it easier to go at my own pace while still having the lecturer explain each part.*

Unfortunately, it turned out that our e-Proofs did not have the desired effects in terms of improved understanding and learning. We discovered this by conducting an experimental study in which students studied a new theorem (Cauchy's general-

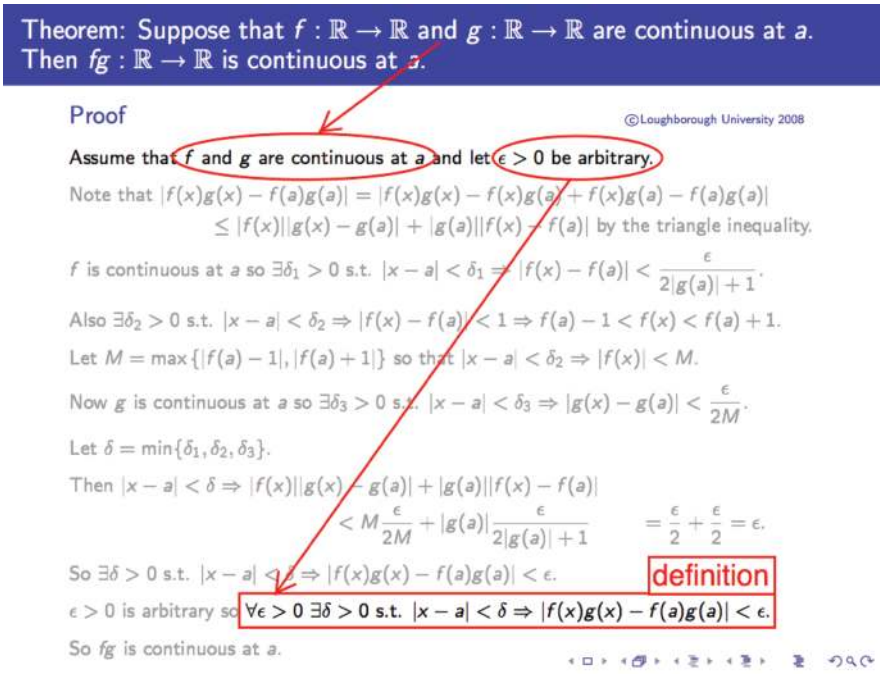

Figure 1. A typical e-Proof screen. The accompanying audio said, "In the first line, we state our assumption that f and g are continuous at a, which corresponds to the premise of our theorem. We also let epsilon greater than zero be arbitrary, because we want to show that fg satisfies the definition of continuity at a, which we will achieve by the end of the proof. Doing so involves showing that something is true for all epsilon greater than zero, so choosing an arbitrary epsilon means that all our reasoning from now on will apply to any appropriate value."

ized mean value theorem) and an accompanying proof. The students were randomly assigned to either an experimental group who studied an e-Proof or a control group who studied the same theorem and proof on paper for the same fixed amount of time. All students then took a comprehension test designed according to the principles outlined in [3]: there were questions testing basic knowledge of algebra and differentiation, understanding of the logical reasoning used in the proof, application of ideas in the proof to examples, and ability to summarize the argument. This immediate post-test was followed two weeks later by an identical delayed post-test that was not announced in advance. The results appear in Figure 2.

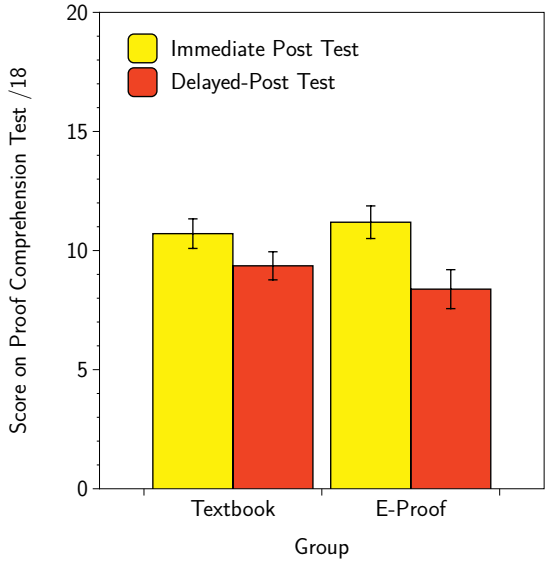

Figure 2. Mean scores for the e-Proof group and the standard presentation group. Error bars show ±1 standard error of the mean. An analysis of variance (ANOVA) revealed a significant main effect of time, $F(1, 47) = 28.213$, $p < .001$, and a significant × time group interaction effect, $F(1, 47) = 5.659$, $p = .021$.

The average scores of the experimental and control groups were not significantly different either at immediate post-test or at delayed post-test. But there was a significant interaction effect: the performance of the students in the e-Proof group dropped more in the intervening time (for details see [4]).

This was a humbling reminder that good pedagogical intentions do not always translate into effective interventions. It does not mean that resources like e-Proofs are never valuable—it could be, for example, that they are not good for first-time learners but are valuable resources for students who have already studied a proof independently and would benefit from clarification on aspects that they have found confusing or difficult. Nevertheless, this result presented a salutary lesson on the limitations of our own understanding of the process of learning from mathematical text: it was clear that we should not construct more e-Proofs or recommend their wider use until we knew more about students' reading processes.

The outcome also raised the broader concern that students might not be accurate reporters on the quality of their own learning. In this case, it seemed likely that students using e-Proofs felt good about their learning because they were able to understand without too much effort, but that this very fact meant that the understanding they acquired was less robust in the longer term. This explanation has been largely confirmed in further studies by the third author—for details see [5]—and is also supported by the remainder of the work presented in this article.

Eye Movements during Mathematical Reading

Our next move was to take a step back and begin a more basic investigation of mathematical reading, studying this process by comparing experts' and novices' eye movements. Eye movements can be studied using technology that allows the researcher to track an individual's focus of attention as that person views information presented on a screen. Modern remote eye-trackers monitor the viewer's pupils using infrared cameras, which are not invasive. Before recording, the tracker must be calibrated by asking the viewer to follow a dot around the screen with their eyes, but the viewer feels nothing, and after calibration the screen looks and behaves exactly like that of an ordinary computer. Eye-tracking is used widely in research on reading (e.g. [6]), and the empirically established close link between fixation location and attention location [7] means that it provides a useful window into the processes involved in reading a text.

Specifically, eye movements lend themselves to quantitative analyses because, although readers experience smooth movement as their eyes shift around a screen, eye movements in fact consist of short *fixations*—typically of around 150–500 milliseconds (ms)—interspersed by very rapid moves known as *saccades* (e.g. [8]). Figure 3 shows a scan path tracing one participant's reading of the instructions for our experiment.

To investigate mathematical reading processes we recruited groups of experts (mathematicians) and novices (first-year undergraduate mathematics students in the UK, roughly the equivalent of US sophomore mathematics majors in terms of mathematical experience). Participants were invited individually to visit our eye-movement lab and were asked to view several purported proofs. For each proof, they were asked to click buttons on a subsequent screen to indicate whether they believed the proof to be valid and how confident they were about their judgment. The first four purported proofs were very short arguments in

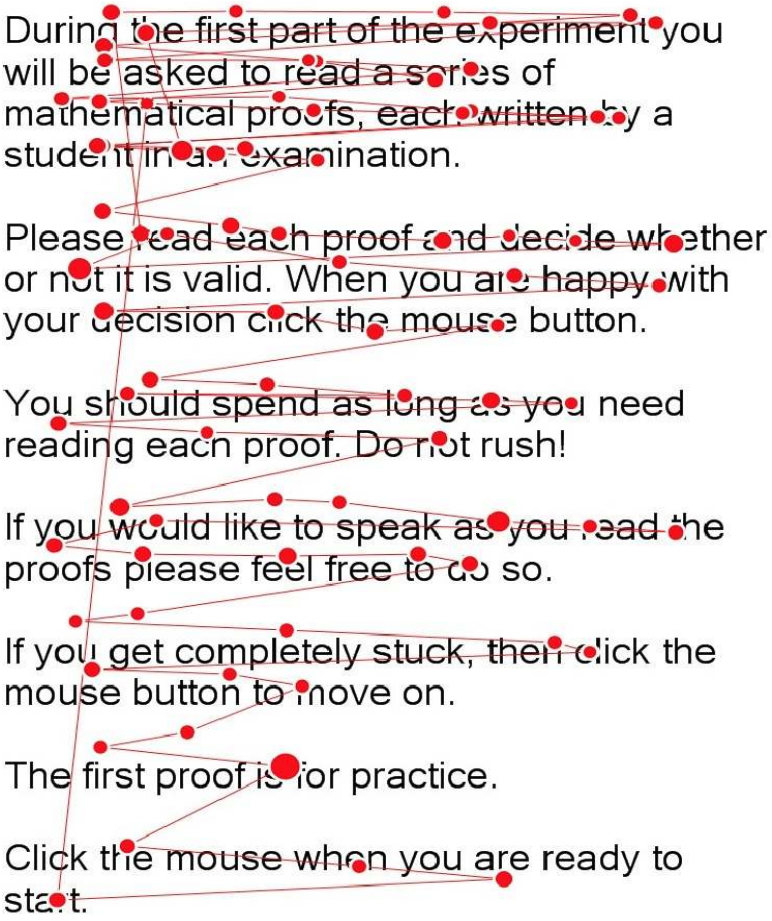

Figure 3. A scan path tracing one participant's eye movements while reading the instructions for the experiment. The discs indicate fixation locations and the straight lines indicate saccades between those locations (these images are produced postrecording by the eye-tracking software and are not visible to the viewer).

elementary number theory that were presented as having been produced by students; the last two were longer and were presented as having been submitted to a recreational mathematics journal (for details see [9]).

We analyzed the eye-movement data in several stages. First, we looked at attention to different features of the proofs. Previous research based on interview studies had led to suggestions that students made poor judgments about proof validity because they tended to focus on the "surface features" of proofs; that they attended adequately to algebraic manipulations but not to the logical structure of an argument as a whole [10]. Our eye-movement data suggested that this might indeed be the case.

Figure 4 (next page) shows one of the longer purported proofs, together with heat maps indicating the degree of attention to different parts of this purported proof by the novices (bottom) and the experts. There is an immediately apparent difference in that the expert mathematicians were very interested in the fifth line of the argument. The validity of the proof depends upon the claim in this line, but the claim is invalid in general and there is no information elsewhere that would make it valid in this context by restricting its applicability. More subtly, the differences do suggest that the students attended more to the algebraic notation.

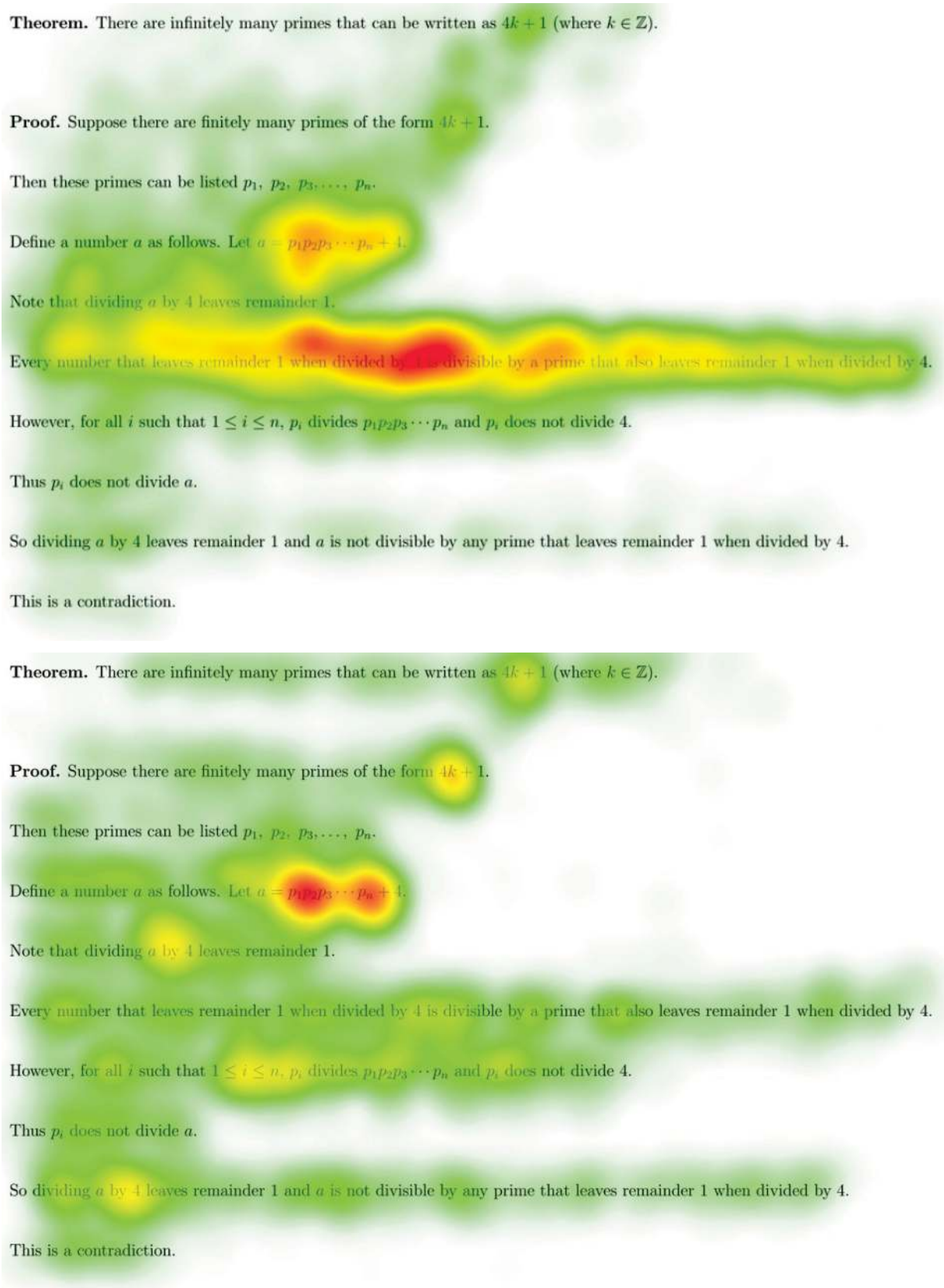

Figure 4. Heat maps showing attention to different parts of an invalid purported proof by mathematicians (top) and undergraduates (bottom) based on data averaged across all participants.

Statistical analyses confirmed this observation. For all six of the purported proofs, we calculated the participants' total *dwell times* on the formulae and on the remaining text (dwell time is calculated by adding the durations of all the individual fixations in a given area of interest; formulae were identified as those parts typeset with math mode in LaTeX). As can be seen in Figure 5, the mean dwell times of the experts and the novices differed: the groups spent about the same amount of time looking at the formulae, but the mathematicians spent more time looking at the words. This provides a measure of empirical support for what many mathematicians suspect: that students at the transition-to-proof level are attentive to algebra but are comparatively unlikely to notice invalid logical reasoning as captured in words.

Next, we looked at another global feature of reading behavior: the pattern of saccades

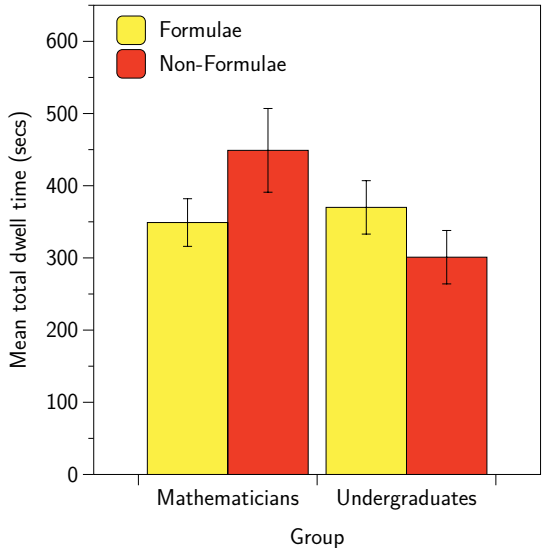

Figure 5. Mean total dwell times on formulae and nonformulae for mathematicians and undergraduates. Error bars show ±1 standard error of the mean. An ANOVA revealed a significant type × status interaction: $F(1, 28) = 8.81$, $p = .006$, $\eta_p^2 = .239$; the students spent proportionately longer fixating on the formulae than did the mathematicians.

around the screen as the reader worked to understand the proof. This required some analytical decisions because it is not practical to describe and meaningfully compare single reading attempts: a five-minute attempt could involve over 1,000 fixations, so general patterns are easily swamped by the detail. We proceeded, therefore, by considering prior theoretical analyses of arguments in general and mathematical arguments in particular.

To a first approximation, a proof can be considered as a sequence of deductions in which each line should follow logically from theorem premises, previous lines, and agreed definitions and theorems (a specific proof might, of course, have a structure more complex than this). In principle it could be that for each deduction there is an explicitly stated *warrant*, a justification for the new claim [11]. In practice, however, many warrants will be left implicit: the author of a proof will expect readers to be able to infer warrants considered to be either common knowledge (in the appropriate context) or otherwise sufficiently obvious from the written material. A reader engaged in a serious attempt to understand a proof therefore has to decide whether a new line requires a warrant and to identify whether and where information relevant to a possible warrant appears elsewhere in the theorem or proof. If individuals do this, we would expect to see it reflected in their eye movements: saccades should take them back and forth between the various lines of the proof.

To obtain a simple measure of this type of behavior, we counted saccades of two types: *within-line saccades* that began and ended within the same line of a proof and *between-line saccades* that began and ended in different lines of the proof (there were of course saccades that began or ended in white space or off the screen; these were not included in our analysis). We found that experts and novices read differently: the experts made significantly more between-line saccades,[1] which is consistent with a search for logical relationships among the lines of the proofs. Figure 6 (next page) illustrates this by showing a scan path of one mathematician's reading of one of the longer purported proofs. Comparing this with the same mathematician's reading of our instructions in Figure 3 highlights an important difference: there is much more back-and-forth movement than one typically sees in ordinary reading.

This result is particularly notable given the mathematical content of the proofs. This content was very straightforward for the mathematicians, necessarily so because our experimental design required the material to be accessible to undergraduates. As a result, one would expect the mathematicians' reading behavior to involve less checking back and forth than would be necessary for the novices. The fact that the experts instead exhibited *more* of this behavior strongly suggests that this is an important feature of expert mathematical reading and one that needs to be developed by typical undergraduates.

Eye-movement data is a rich source of information, and, combined with our other data, it also allowed us to conduct further analyses. Using the validity judgments, we confirmed that undergraduates did not perform well in distinguishing valid from invalid proofs. But we also found that mathematicians did not agree nearly as much as might be expected about the validity of even simple arguments; we have since followed up on this result with a larger study reported in [12].

[1] *78.8 between-line saccades per proof compared with 53.3 per proof, t(28)=2.11, p=.044, d=0.80.*

Using the eye-movement data, we discovered that mathematicians did not conduct initial "skim reads" of the purported proofs, despite routine self-report-based claims that this is a common behavior (e.g. [13]); these results are reported in [9] and [14]. Finally, we examined eye-movement sequences that we considered particularly likely to indicate searches for implicit warrants: shifts from one line of a proof to its predecessor and back again. We found (see [9]) that mathematicians were three times more likely than undergraduates to make such eye movements but that both mathematicians and students were significantly more likely to behave in this way when a warrant was required (when a line required justification rather than simply, say, introducing new terminology). For the purposes of our work on proof comprehension, this indicated a possible way forward.

Self-Explanation Training in Mathematics

We reasoned that if students were aware that they should be looking for justifications but were not doing so very much or very effectively, their comprehension could perhaps be improved via simple training encouraging them to devote more effort to this aspect of mathematical reading. The training approach we took was based on the literature on reading to learn and specifically on a promising intervention commonly termed *self-explanation training*. Self-explanation training is based on early observations, that when learning from texts on Newtonian mechanics, students who showed better subsequent problem-solving performance made more self-explanations: they were more inclined to articulate interpretations that involved information and relationships beyond those explicitly contained in the text [15]. There is a large and growing literature on self-explanation effects (e.g. [16]), and variants on self-explanation training have been used with lower-level mathematics students [17]. But such training had not been adapted for use in undergraduate mathematics.

We adapted a version of self-explanation training from earlier materials used in [18] and [19]. Our training was presented in a series of computer slides for studies conducted in the lab and in a paper booklet for studies conducted in a lecture theatre. The slides and booklet elucidated key principles of self-explanation training as applied to mathematical proofs. Specifically, they:

- instructed students to identify key ideas in each line of a proof and to explain each line in terms of other ideas in the text or in terms of their own existing knowledge;
- noted that self-explanation differs from simply paraphrasing the text without adding new information and from making monitoring statements such as "Okay, I understand that line";
- demonstrated the self-explanation strategy by exhibiting possible student self-explanations in relation to a very short example proof;
- instructed students to generate self-explanations in response to a practice proof.

A full version of the self-explanation training is available at `www.setmath.lboro.ac.uk`; students in our studies spent approximately 15–20 minutes working through it.

Our first study was conducted in the lab. Student participants attended an individual session and were randomly assigned to either an experimental or a control group. Those in the experimental group studied the self-explanation training, and to equalize the time spent in the lab environment, those in the control group were

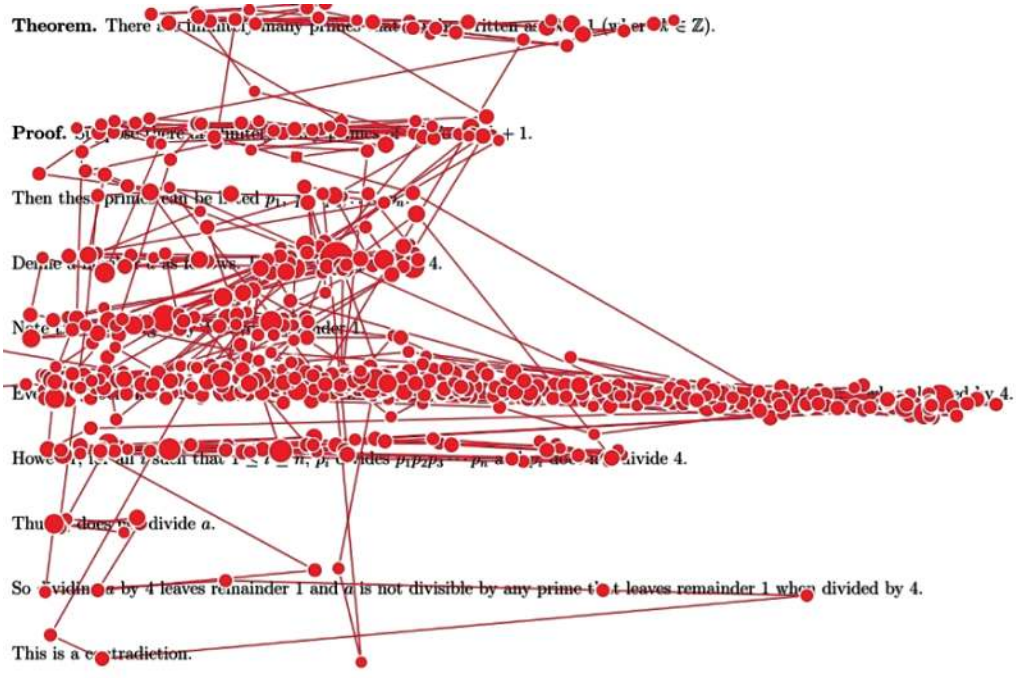

Figure 6. A scan path tracing one mathematician's eye movements as he/she reads an invalid purported proof. Compared with the scan path in Figure 3, this shows more back-and-forth movement during the reading attempt.

asked to read and answer questions on a passage on the history of mathematics. Participants in both groups were then asked to read a proof presented on a screen, first silently and then taking one line at a time and giving explanations out loud. The only difference in this stage was that the self-explanation group was explicitly asked to use its training as a guide when generating these explanations. Finally, each participant completed a fourteen-item free-response proof comprehension test designed according to the principles outlined in [20]. This study design provided us with two sets of data: the participants' verbal explanations and their proof comprehension scores.

Analyses revealed that the self-explanation training had the desired effect. The participants' verbal explanations were classified using a scheme adapted from [19], and we found that students in the self-explanation group gave significantly more high-quality explanations: they produced around twice as many explanations that were classified as *inferring warrants* (articulating justifications), *noticing coherence* (relating lines of a proof to each other), or *being goal driven* (relating a line to the overall goal of proving the theorem). The full range of classification types and numbers is captured in Figure 7.

The comprehension test data required a more nuanced analysis, because time spent studying the proof was correlated with comprehension score and because those in the self-explanation group spent longer on average studying the proof. We were not interested simply in increasing study time; we wanted to know whether students in the self-explanation group learned more effectively. We thus controlled for study time and found that the scores of students in the self-explanation group were significantly higher. Moreover, the size of the effect was large: the students who had received the self-explanation training scored on average almost one standard deviation higher than those in the control group. Finally, we found that this effect was evident across students from all three of the university's academic years, as shown in Figure 8.

Encouraged by this experimental result, we went on to further study its causes by extending our eye-movement work. In particular, we were interested in whether self-explanation training led to observable changes in reading behavior. This required a somewhat complex study design because, as might be anticipated, there is considerable individual variation in eye movements.

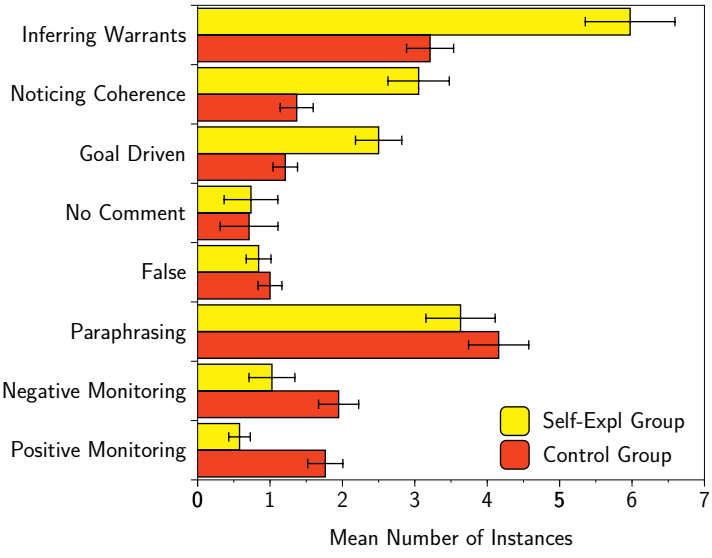

Figure 7. Mean numbers of explanations of different types given by students in the self-explanation training and control groups. Error bars show ±1 standard error of the mean. Bonferroni-corrected Mann Whitney U tests revealed significant differences in numbers of comments classified as principle-based, $U = 386$, $p < .001$, noticing coherence, $U = 399$, $p = .001$, or goal-driven, $U = 407$, $p = .001$ (and as positive monitoring, $U = 400$, $p < .001$ and negative monitoring, $U = 440$, $p = .002$).

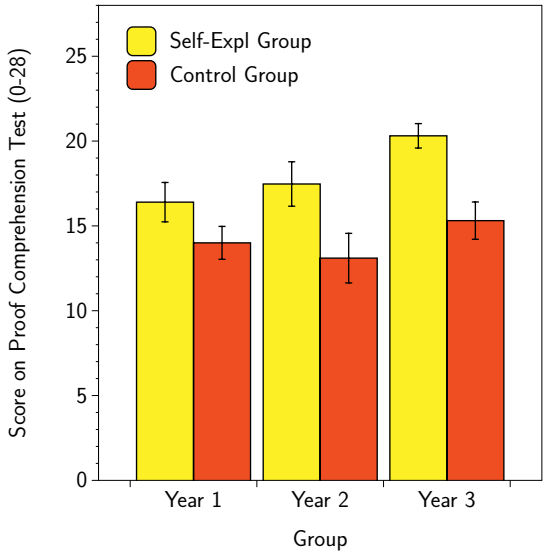

Figure 8. Mean scores on the proof comprehension test, separated by condition and year of study. Error bars show ±1 standard error of the mean. A 3 (year) × 2 (condition) analysis with covariance (ANCOVA) with time as a covariate revealed a main effect of condition, $F(1, 69) = 181.459$, $p < .001$, with those in the self-explanation group outperforming those in the control group. It also revealed a main effect of year, $F(2, 69) = 3.456$, p = .037, with those in Year 3 ($M = 17.8$, $SD = 4.2$) outperforming those in Years 2 ($M = 15.8$, $SD = 5.2$) and 1 ($M = 14.9$, $SD = 3.9$), but no significant year × condition interaction, $p > .2$.

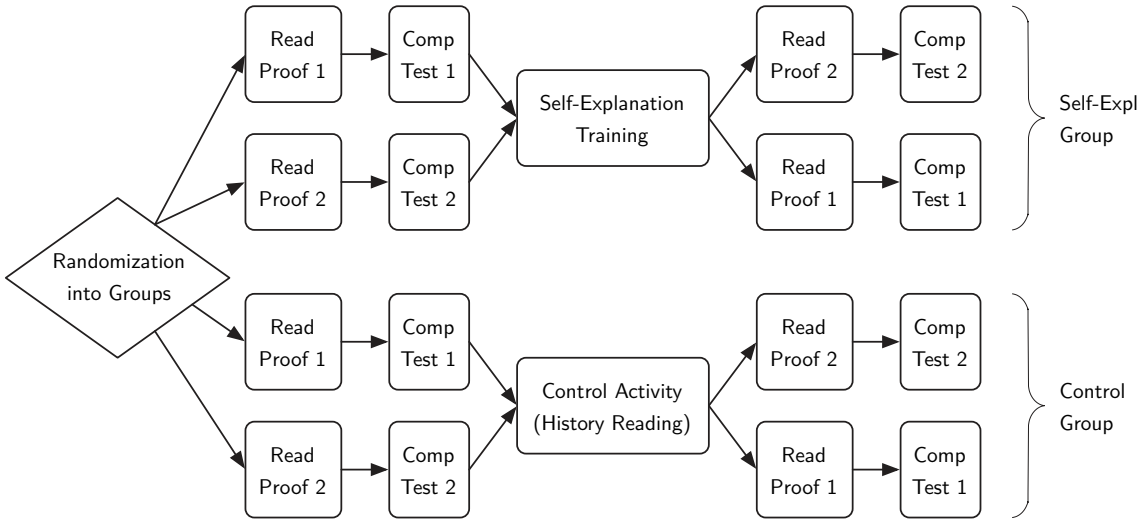

Figure 9. Design for the study of the effects of self-explanation training on eye movements ([22], p. 74).

The design involved four groups and is represented diagrammatically in Figure 9. In this within-subjects design, every participant studied two proofs and completed two proof comprehension tests (multiple-choice tests in this case). This allowed us to study changes in individual reading behavior. The experimental groups received the self-explanation training and the control groups read the alternative text as before, and we were interested in comparing the groups' reading behaviors and comprehension scores for the second proof they read. But it is also conceivable that differences between two proofs would generate systematic differences in reading behaviors and scores, so both the experimental and the control groups were split into two and a counterbalanced design was employed in which half saw one proof first and half saw the other.

Analyzing the comprehension scores again showed that self-explanation training had a positive effect. Independently of which proof was seen second and controlling for comprehension scores on the first attempt, the self-explanation groups outperformed the control groups.[2]

Of more interest in this case, however, was the change in reading behaviors. We investigated these using two separate measures. First, we looked at mean fixation duration, which acts as a measure of intellectual effort: higher mean fixation durations reflect harder concentration (e.g. [21]). We compared the mean fixation durations of students in the experimental and control groups on whichever proof they read second, this time controlling for mean fixation durations on proof read first to account for preexisting individual differences on this measure. This analysis revealed a between-groups difference: regardless of the order in which the participants experienced the proofs, those who received the self-explanation training subsequently concentrated harder.[3]

Second, we looked as before at between-line saccades (see Figure 10). We compared the numbers of between-line saccades for students in the experimental and control groups on the proofs they read second, this time controlling for both the time taken to read this proof (we were effectively interested in between-line saccades per minute, not total saccades) and the number of between-line saccades for the proof read first (again to account for individual differences in reading behavior). This time we found a main effect of proof: some proofs, it seems, do prompt different reading behaviors. For the self-explanation training, we again found a significant difference in the expected direction: regardless of the order in which they experienced the proofs, students who had received the training subsequently made significantly more between-line saccades. This indicates more shifts of attention around the proof and is consistent with more attention to logical relationships between the lines of the proof. In other words, students who had received self-explanation training exhibited reading behaviors more like those associated with expert mathematical reading.

[2] *An ANCOVA with two between-subjects factors (condition: self-explanation, control; proof read second: Proof 1, Proof 2) and one covariate (proof comprehension scores from the first reading attempt) showed a main effect of condition, $F(1,27) = 8.850$, $p = .006$, $\eta_p^2 = 0.247$, but no significant effect of proof order and no significant condition-by-proof-order interaction, both $Fs < 1$.*

[3] *An ANCOVA with two between-subjects factors (condition: self-explanation training, control; proof read second: Proof 1, Proof 2) and one covariate (mean fixation durations for the proofs read first) revealed a significant main effect of condition, $F(1,23) = 14.234$, $p = .001$, $\eta_p^2 = .382$ but no significant main effect of proof order and no significant condition-by-proof-order interaction, $ps > .3$.*

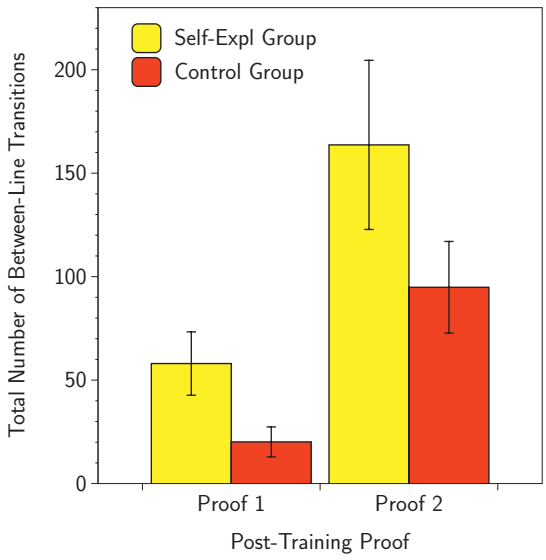

Figure 10. Mean numbers of between-line saccades for the proof read second, split by condition and proof read second. Error bars show ±1 standard error of the mean. An ANCOVA with two between-subjects factors (condition: self-explanation training, control; proof read second: Proof 1, Proof 2) and two covariates (number of between-line saccades made during the first proof reading attempt and the overall duration of the second proof reading attempt) revealed a significant effect of condition, $F(1,22) = 10.394$, $p = .004$, $\eta_p^2 = 0.321$, and a significant effect of proof order, $F(1,22) = 8.449$, $p = .008$, $\eta_p^2 = 0.277$, but no significant interaction between condition and proof order, $p = .742$.

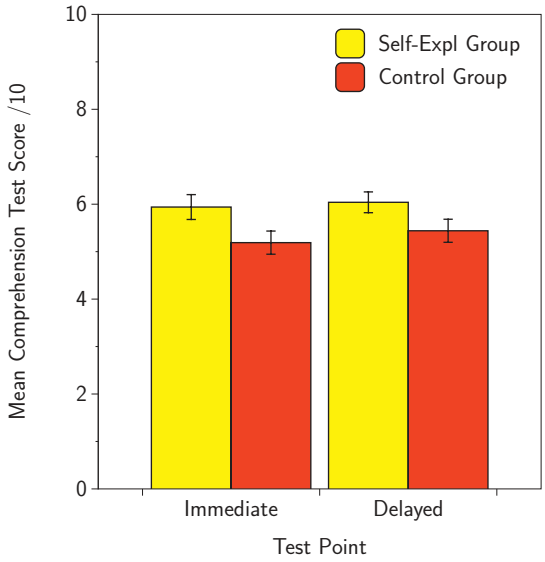

Figure 11. Mean scores at post-test and delayed post-test, split by condition and time. Error bars show ±1 standard error of the mean. An ANOVA with one within-subjects factor (time: immediate post-test, delayed post-test) and one between-subjects factor (condition: self-explanation, control) showed a main effect of condition, $F(1,105) = 6.024$, $p = .016$, $\eta_p^2 = 0.054$, but no significant effect of time and no interaction between condition and time, $F < 1$ in both cases. The differences corresponded to effect sizes of $d = 0.410$ at post-test and $d = 0.350$ at delayed post-test.

Finally, we took our work out of the lab and into the classroom, conducting a larger-scale study of the effects of self-explanation training for students working individually in an ordinary lecture theater. One hundred seven first-year calculus[4] students were randomly assigned to experimental and control groups, where in this case the self-explanation group read a printed version of the self-explanation training and the control group read materials on time management for mathematics students. All students then read a proof and took a multiple-choice comprehension test. In this case, we also followed up twenty days later with a delayed post-test in which all students were asked to read a second proof and take a second multiple-choice comprehension test. The results are shown in Figure 11; they indicated that in both immediate and delayed post-tests, scores of students in the self-explanation group were significantly higher.

Detail on all three of our self-explanation studies can be found in [22].

Discussion

The research reported here has given us improved insight into mathematical reading and expertise, and into the effects of specific research-based teaching interventions. And it leads to a simple implication: undergraduate mathematics students should receive self-explanation training because this can be expected to improve their mathematical reading and consequently their proof comprehension.

However, as is always the case with empirical research, our work has limitations and opens up more questions than it answers. It would be a mistake, for instance, to infer that self-explanation training constitutes a silver bullet: the proofs used in our studies were all fairly short ones drawn from number theory, the experimental groups did not end up with perfect understanding, and certainly there is room for more nuanced research to investigate interactions between self-explanation training and factors like mathematical topic, students' prior knowledge, and alternative pedagogical strategies. It is possible, for instance, that self-explanation effects would be more pronounced for certain groups of students, that the training might be ineffectual for some groups or for some mathematical topics, or that the effects

[4] *In the UK students specialize earlier than they do in US-style systems. Participants in this study had worked on single-variable calculus as part of A-Level Mathematics between the ages of sixteen and eighteen and were taking a course that reviewed this material and extended it into multivariable calculus.*

could be enhanced by opportunities to practice self-explanation strategies in the classroom or by combination with other learning experiences. One important message in this regard is that at this stage we do not know—empirical research is required to investigate these possibilities.

We believe that this message is particularly important in the contemporary educational environment in which much is made of the potential of technology to enhance learning and much value is placed upon innovation. Much less value typically is placed on evaluation, and we think that this is a mistake. The world of the contemporary student is full of apparently useful resources, and access to these is becoming ever easier. This might be good, and it is certainly empowering: students can take charge of their own learning, locating and using resources that provide them with what they feel they need. But many resources are expensive to produce; developing them requires a substantial investment of academic time and technical support. And not all resources will lead to improved learning. As we discovered in our work with e-Proofs, interventions that are designed to make things easier might succeed in that aim and might be well received, but this does not guarantee that they provide effective support for sustainable learning. This, we believe, will always make it risky to evaluate innovations using only self-reporting measures: students might sincerely believe that new resources are of benefit, and they might be right in the sense that those resources make learning easier in the short term, but our results collectively suggest that it might be preferable to leave some resources as they are and focus instead on helping students to engage with them effectively. Perhaps some things should be difficult.

With these comments in mind, we believe that the success of self-explanation training across our three studies is encouraging not only because it appears to be effective but also for two further reasons. First, self-explanation training is extremely light touch: it is generic, it does not rely upon time-intensive adaptation of existing resources, and students can work through it independently in about 15–20 minutes (as noted above, the training is available at `www.setmath.lboro.ac.uk` for readers who might wish to use it). Second, self-explanation training does not require more work from the student; it encourages more effective independent work by simply teaching students to make better use of their existing knowledge and reasoning skills. Studies in education research often highlight what students cannot do, so it is cheering to be able to present positive results based on things that they can.

References

[1] L. ALCOCK and N. WILKINSON, e-Proofs: Design of a resource to support proof comprehension in mathematics, *Educational Designer* **1** (4) (2011), `www.educationaldesigner.org/ed/volume1/issue4/article14/index.htm`.

[2] R. E. MAYER and R. MORENO, Nine ways to reduce cognitive load in multimedia learning, *Educational Psychologist* **38** (2003), 43–52.

[3] K.-L. YANG and F.-L. LIN, A model of reading comprehension of geometry proof, *Educational Studies in Mathematics* **67** (2008), 59–76.

[4] S. ROY, L. ALCOCK, and M. INGLIS, Undergraduates' proof comprehension: A comparative study of three forms of proof presentation, *Proceedings of the 13th Conference on Research in Undergraduate Mathematics Education*, Raleigh, NC.

[5] S. ROY, *Evaluating a novel pedagogy in higher education: A case study of e-Proofs*, PhD thesis, Mathematics Education Centre, Loughborough University, UK, July 2014.

[6] K. RAYNER, Eye movements in reading and information processing: 20 years of research, *Psychological Bulletin* **124** (1998), 372–422.

[7] M. A. JUST and P. A. CARPENTER, A theory of reading: From eye fixations to comprehension, *Psychological Review* **87** (1980), 329–354.

[8] E. MATIN, Saccadic suppression: A review and an analysis, *Psychological Bulletin* **81** (1974), 899–917.

[9] M. INGLIS and L. ALCOCK, Expert and novice approaches to reading mathematical proofs, *Journal for Research in Mathematics Education* **43** (2012), 358–390.

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

[11] S. TOULMIN, *The Uses of Argument*, Cambridge Univ. Press, Cambridge, 1958.

[12] M. INGLIS, J.-P. MEJÍA-RAMOS, K. WEBER, and L. ALCOCK, On mathematicians' different standards when evaluating elementary proofs, *Topics in Cognitive Science* **5** (2013), 270–282.

[13] J.-P. MEJÍA-RAMOS and K. WEBER, Why and how mathematicians read proofs: Further evidence from a survey study, *Educational Studies in Mathematics* **85** (2014), 161–173.

[14] M. INGLIS and L. ALCOCK, Skimming: A response to Weber & Mejía-Ramos, *Journal for Research in Mathematics Education* **44** (2013), 471–474.

[15] M. T. H. CHI, M. BASSOK, M. W. LEWIS, P. REIMANN, and R. GLASER, Self-explanations: How students study and use examples in learning to solve problems, *Cognitive Science* **13** (1989), 145–182.

[16] B. A. FONSECA and M. T. H. CHI, Instruction based on self-explanation, *Handbook of Research on Learning and Instruction* (R. E. Mayer and P. A. Alexander, eds.), Routledge, NY, 2011, pp. 296– 321.

[17] K. DURKIN, *The self-explanation effect when learning mathematics: A meta-analysis*, presented at the Society for Research on Educational Effectiveness (2011), `eric.ed.gov/?id=ED518041`.

[18] K. BIELACZYC, P. L. PIROLLI, and A. L. BROWN, Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving, *Cognition and Instruction* **13** (1995), 221–252.

[19] S. AINSWORTH and S. BURCHAM, The impact of text coherence on learning by self-explanation, *Learning and Instruction* **17** (2007), 286–303.

[20] J.-P. MEJÍA-RAMOS, E. FULLER, K. WEBER, K. RHOADS, and A. SAMKOFF, An assessment model for proof comprehension in undergraduate mathematics, *Educational Studies in Mathematics* **79** (2012), 3–18.

[21] M. A. JUST and P. A. CARPENTER, Eye fixations and cognitive processes, *Cognitive Psychology* **8** (1976), 441–480.

[22] M. HODDS, L. ALCOCK, and M. INGLIS, Self-explanation training improves proof comprehension, *Journal for Research in Mathematics Education* **45** (2014), 62–101.