Observing Students' Moment-by-Moment Reading of Mathematical Proof Activity

Paul Christian Dawkins
Northern Illinois University

Dov Zazkis
Arizona State University

This study presents findings from a series of interviews in which we observed undergraduate students' moment-by-moment Reading of Mathematical Proof (ROMP) activity. This methodology is adapted from a validated assessment of narrative reading comprehension developed by cognitive psychologists. We demonstrate the fruitfulness of the method by describing four relatively novel phenomena that we observed in our interviews, and highlight ROMP activities that seemed to distinguish less productive and more productive readers.

Keywords: Proof, Reading, Transition to Proof, Systemic Functional Linguistics

Much of students' apprenticeship in advanced mathematics at the tertiary level involves learning how to read and write mathematical proof. Mathematics educators have studied this transition in terms of students' ability to comprehend proofs after reading (e.g. Mejia-Ramos, Lew, de la Torre, & Weber, 2017), validate proofs (e.g. Alcock & Weber, 2005; Inglis & Alcock, 2012), and write proofs (e.g. Weber, 2001). Fewer studies have investigated the reading of mathematical proof (ROMP) process itself (Weber, 2015). In this paper, we present findings from our adaptation of a moment-by-moment reading assessment method developed by psychologists for studying narrative text reading (Magliano & Millis, 2003; Magliano, Millis, Team, Levinson, & Boonthum, 2011). That methodology of read aloud interview protocols and line-by-line presentation provides different insights into narrative reading than those provided by end-reading comprehension tests. Similarly, we argue here that our method reveals a different set of sense-making activities than has previously been documented. We also contribute to the literature by comparing the ROMP behaviors of novice readers and more experienced readers.

Relevant Research Studies

Our study builds directly on the work of cognitive psychologists Magliano, Millis, and their team, who developed the Reading Strategy Assessment Tool (RSAT) (Magliano et al., 2011). RSAT is a validated measure of reading comprehension. It presents students with single lines of text and asks students to think aloud about each line. The nature of the inferences that students make indicate their relative competence as a reader in the following way: students who connect the given line to previous lines (*bridging inferences*) or to their outside knowledge (*elaborating inferences*) tend to have higher comprehension than students who merely restate lines (*paraphrasing inferences*). The quality of the inferences is less salient to this assessment compared to an end-reading comprehension test. RSAT focuses on qualitative differences in reading behavior rather than post-reading understandings. Forming bridging inferences and elaborating inferences correlates with measures of end-reading comprehension.

Fletcher, Lucas, and Baron (1999) adapted this moment-by-moment reading assessment methodology to ROMP, using secondary geometry proof texts. They directly compared the observed behavior to reading of narrative text. They reported that ROMP was more effortful than reading narrative texts and elicited a different constellation of reading activities. The primary reading activity novel to ROMP was *forward elaboration* in which students anticipate later lines of the text, which was less common in reading narrative text.

Mejia-Ramos, Fuller, Weber, Rhoads, and Samkoff (2012) present a framework for the various kinds of understanding students might develop from ROMP, which built heavily on Yang and Lin's (2008) framework. Those authors successfully adapted their framework into a validated, multiple-choice assessment of end-reading comprehension (Mejia-Ramos et al., 2017). Our study and methodology differ in large part because we seek to investigate moment-by-moment ROMP activities and our analysis focuses more on sense-making activities rather than kinds of understanding to be constructed.

Relatively few undergraduate mathematics education studies focus on reading processes. Shepherd and van de Sande (2014) compared undergraduate student reading of textbooks to faculty reading. They found salient differences regarding the way their subjects articulated equations; experts referred to parts of equations in terms of their meaning or role while novices read the names of the symbols in sequence. A couple of studies compare expert and novice ROMP behaviors using eye-tracking technology (Inglis & Alcock, 2012; Panse, Alcock, & Inglis, 2018). An interesting finding from those studies is that novices attend more commonly to equations in proof texts while experts spend more time examining the connecting statements that state logical inferences. Weber (2015) reports some reading behaviors of very successful undergraduate students by which they made sense of a novel proof text.

Two studies report on interventions aimed at improving student ROMP activity. Hodds, Alcock, and Inglis (2014) adapted self-explanation training from other reading domains to the context of proof and found that self-explanation training was successful in improving student comprehension of proofs they read. Samkoff and Weber (2015) reported findings from trying to train students in the effective reading behaviors reported in Weber (2015). They had modest success, though students needed guidance in using the strategies effectively.

Analytical Framework

Our analysis of the reading process is informed largely by the tradition of Systemic Functional Linguistics (Halliday, 1994; Schleppegrell, 2004). As suggested in its name, SFL emphasizes how language functions to make meaning, either in articulation or interpretation of language. From this standpoint, choice is a key aspect of all language use. In particular, Halliday argues that linguistic choices are made to achieve three metafunctions: *ideational* – what is being talked about, *interpersonal* – who are the interlocutors and how are they positioned, and *textual* – what kind of text is being constructed. In this study, we particularly attend to the first and third metafunctions (though interpersonal metafunctions influenced the observed ROMP activity). The ideational metafunction (which Halliday at times subdivided into experiential and logical) for mathematical proof naturally involves discussion of mathematical objects, properties, and relationships. One of the novel contributions of this study consists in observing how the textual metafunction became salient in students' ability to make meaning of the proof texts they read.

Methodology

Adapted Assessment Method

To select proof texts for students to read, we searched introduction to proof textbooks and asked for ideas from mathematician colleagues. We sought proofs that were at least 10 lines (to increase opportunities to respond), were accessible to novice readers, and that were less likely to appear in common undergraduate courses (to minimize prior exposure). We selected four proof texts, proving the statements listed in Figure 1. Mirroring RSAT, we developed both general and specific response prompts for each line of text. Like RSAT, the two authors began by coding each line of text for all of the connections we expected students might make. This informed our choice of specific response prompts for each line. Students were always asked to think aloud, but the more specific response prompts included:

- "Why is this line justified?" inviting identification of data, definitions, and warrants.
- "What is the purpose of introducing d?" probing student recognition of goals.
- "What does this line accomplish?" assessing achievement of proof goals.
- "What do you expect in the following line[s]?" inviting forward elaboration.

The final prompt was used when we expected that students would be able to elaborate forward based on a proof frame that had been introduced (cases, universal generalization, contradiction, induction, dual inclusion between sets) or because a stated goal was nearly accomplished.

For every primitive Pythagorean triple (a, b, c) there exist some numbers s and t with no common factors such that $s > t \geq 1$ where $a = st$, $b = \frac{s^2 - t^2}{2}$, and $c = \frac{s^2 + t^2}{2}$.

Euclid's Fifth Postulate implies Playfair's Parallel Postulate.

For any two sets A and B, $A^c \cap B^c = (A \cup B)^c$.

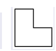
For any positive integer n, a $2^n \times 2^n$ square grid with any one square removed can be covered with L-shaped tiles. 

Figure 1. Statements of the four theorems proven in the texts presented to students.

Study Participants and Interview Methodology

We recruited from courses at one medium and one large public research university in the US. To sample students with varying experience, we recruited from differential equations, introduction to proof, real analysis, and topology courses. We classified our participants in three groups: novice readers who had completed no proof-oriented courses at university (6), experienced readers who had completed at least one such course (9), and graduate students (2). Interviews were conducted outside of class time for 1-2 hours, and students were modestly compensated. All interviews were audio recorded and any student work was retained.

Each proof text reading began by students reading definitions, previous theorems assumed true, and the statement of the theorem to be proven. The interviewers answered questions about mathematical facts and clarified the theorem statement if needed (e.g. L-shaped tiles each covered 3 squares). We generally avoided explanations that would affect the reader's construal of proof text itself. Students could always see all prior lines of the text as they read and had the definitions and theorem statements available on paper. In addition to scripted response prompts, interviewers could ask elaboration questions at their discretion.

Analysis Methods

Interview coding proceeded in three stages. Upon watching the interview recording, the researcher first described the student response to each response prompt for each line, transcribing quotes that seemed significant or relevant. Upon completing these detailed *field notes*, the researcher then compiled a list of notable patterns in each student's *ROMP activity* on each proof. Some organizing categories emerged for this stage of analysis, but these were meant to guide the researcher's noticing more than serving as research constructs. In particular, we always tried to focus on ways students sought to make meaning of the text, regardless of how normative their interpretations were. Initially, both authors completed these first two stages of coding for the same two interviews. Once we compared and reach some agreement about the process, the rest of the interviews were partitioned and each analyzed by only one of the two authors.

The third stage of analysis followed thereafter when we created general categories of ROMP activities that could be assessed on all the tasks by specific indicators. This proved challenging because students did not exhibit particular ROMP activities uniformly throughout each text and students' ability to construe each proof normatively did not appear to correlate with their ROMP experience. The final goal of the analysis is to identify parent *categories of ROMP activities* that can be assessed for each proof text along with *indicator activities* specific to each proof that can be used to represent each student's reading of that text. It is beyond the scope of this report to present these categories and indicators. Rather, in the following section we present some representative reading phenomena observed that demonstrate the fruitfulness of this assessment methodology and the complexity of student ROMP activities. Figure 2 presents the first proof that students interpreted, that will be referenced in the data presented.

1. Let (a, b, c) be a primitive Pythagorean triple.
2. Then $a^2 = c^2 - b^2 = (c+b)(c-b)$.
3. We want to show that $(c+b)$ and $(c-b)$ are both squares and share no common factors.
4. Suppose that d is a common factor of both $(c+b)$ and $(c-b)$.
5. Then d is also a factor of $(c+b)+(c-b) = 2c$ and a factor of $(c+b)-(c-b) = 2b$.
6. So both $2b$ and $2c$ are multiples of d.
7. If d is factor of both b and c, then d is also factor of a.
8. This contradicts (a, b, c) being a primitive Pythagorean triple.
9. Since d is not a factor of both b and c, d divides 2.
10. Hence d is 1 or 2.
11. d divides a, which is odd, thus d cannot be 2.
12. Thus $d = 1$ and $(c+b)$ and $(c-b)$ share no factors greater than 1.
13. This means that $(c+b)(c-b) = a^2$ implies that $(c+b)$ and $(c-b)$ are both perfect squares.
14. So there are some numbers s^2 and t^2 such that$(c+b) = s^2$ and $(c-b) = t^2$.
15. s and t share no common factors as established above.
16. Clearly, $s > t \geq 1$.
17. Finally, $c = \frac{s^2+t^2}{2}$, $b = \frac{s^2-t^2}{2}$, and $a = \sqrt{(c+b)(c-b)} = st$

Figure 2. Proof characterizing primitive Pythagorean triples (adapted from Rotman, 2013).

Results

In this section we exemplify of four ROMP phenomena that we observed in our interviews: 1) computational and inferential orientations, 2) low-level construal of proof claims, 3) ongoing revision of proof construal, and 4) patterns of identifying and stating warrants.

Computational and Inferential Orientations

We observed that some novice readers interpreted the proof texts using what we call a *computational orientation* while more experienced and effective readers exhibited an *inferential orientation*. These two constructs relate to the textual metafunction. That is, they relate to the student's sense of what kind of text is being constructed and what kinds of activities are relevant in such a text. The distinction was most prominent with regard to how students interpreted equations in proofs. We have reported more fully on this distinction elsewhere (Dawkins & Zazkis, 2018), so we shall merely describe this phenomena without extensive data.

The first proof used the equation $a^2 = c^2 - b^2 = (c+b)(c-b)$ in multiple ways. First, it is used to infer that if $(c+b)$ and $(c-b)$ are both multiples of d, then a is also (L7, L11). Later, it was used to infer that since $(c+b)$ and $(c-b)$ had no common factors they are both perfect squares (L13). Students who exhibited a computational orientation saw the equation and the introduction of d as a factor of $(c+b)$ and $(c-b)$ as an opportunity to substitute into the equation and solve for certain variables. They made meaning of the text using practices that were native to the mathematics courses they had thus far completed in college (calculus and differential equations). We understood this as construing the proof as a different kind of mathematical text than was actually being produced. These students often exhibited great perturbation in sense making, and articulated desire to deal with the equations in familiar "plug-and-chug" ways. Students exhibited an inferential orientation when they interpreted the equation as a means of inferring the properties of the various quantities in the equation (as is intended).

Low-Level Construal of Proof Claims

Low-level construal of proof claims refers to the quality of the mental model students build of the information presented in the text. This relates both to the model of the line currently being read and how students' interpretation/recall of previous lines affects their reading of the current line. We report Novice 1's (Nov1) ROMP activity to exemplify this construct.

A number of steps in the primitive Pythagorean triples (PPTs) proof (Fig 2) related to which numbers shared common factors (definition of PPT, L3, L4, L7, L12, L15). This relation thus appears in the proof with reference to at least four sets of numbers: (a, b, c), $((c+b),(c-b))$, $(2b, 2c)$, (s, t). Some relations are assumed by hypothesis (L1), some are assumed toward a contradiction (L7), and others are inferred from other properties (L15). When Nov1 read the definition of PPT, he said, "Run of the mill Pythagorean triple that I've learned since high school." He showed no sign of attending to the word "primitive" or how it modifies the meaning of Pythagorean triple by incorporating an additional no common factors stipulation.

When Nov1 read L4, he correctly noted that d would be used to accomplish the goals stated in L3, likely using proof by contradiction. Nov1 justified L5 by imagining factoring d out of the expressions $(c+b)$ and $(c-b)$ and then factoring again to show that both sides of the equations are multiples of d. He made a similar argument for L7, except applied to the equation in L2. His reasoning suggests his meaning of "factor of" in terms of being able to factor a term out of an expression was productive in helping Nov1 justify certain inferences. He also seemed aware of the goal stated in L3 regarding "no common factors" and how d would be used to accomplish that goal. After reading L7, the interviewer asked what Nov1 expected to follow:

> A little up in the air because of the assumption it would be proof by contradiction because in the assumption of the, it said that "with no common factors," even though that, in the next coming line we are going to be moving towards "d does not work for both b and c." [The interviewer asked him to elaborate.] Because the theorem being proven it says that there are some numbers with no common factors, but then again that's, yeah. But that's for s and t and I just transferred that assumption to a and b, but I don't know. If s and t have no common factors, oh, but b and c already have a common factor of 2 because they are both being divided by 2, or ½ I should say. So the assumption that, from what I derived from the theorem being proven, it's being misassigned to b and c and not necessarily to $\frac{s^2-t^2}{2}$ and $\frac{s^2+t^2}{2}$. So I am excited to see what this next line says.

This marked a shift in Nov1's ability to track the inferences being made. He began trying to interpret L7 in terms of the properties of s and t (part of the theorem's conclusion). He also inferred that b and c are divisible by 2 based on the equations in the theorem's conclusion.

After reading L8, Nov1 questioned his prior claims and decided that L8 was referring to the "no common factors" claim in the definition of PPT. He did not elaborate further on how this revised his interpretation of the proof. When Nov1 read L9, he was able to explain the claim with reference to L6. He exemplified this inference when $2b = 10$ and $2c = 14$. The interviewer asked what the rest of the proof needed to accomplish, and part of Nov1's reply was: "What the last couple of lines have been is giving the evidence and basically proving in a more theoretical way that a, b, and c share no common factors, and so the next part of the proof will be defining b and c in terms of s and t so that they will have no common factors." When Nov1 read L12 that explicitly refers back to the goal in L3, he again concluded that this line verified that a, b, and c share no common factors.

Nov1 read the last part of the proof frequently using the conclusions of the theorem to justify proof claims. He used the equations in the theorem statement to justify L14. Reading L15, Nov1 said that it was self-explanatory because it was stated in the theorem. In his explanation, he referred to factors of b and c, s and t, and $(c + b)$ and $(c - b)$, but he showed no sense of dependence among these claims. Rather, he said this line simply reminded the reader of what had been done, since everything was being redefined in terms of s and t. He similarly noted that L16 was "a statement made in the theorem." After he had read the entire proof, Nov1 reflected, "I would have plugged and chugged would have to worked to get this expression from that expression. But I would have skipped all the $2b$ and $2c$ and the common factor stuff."

To summarize, in Nov1's ROMP activity he was quite successful at using equations to show that if some constituents had a factor of d, then others would also. He used his meaning for factor to connect L9 to L6 using particular examples (c.f., Weber, 2015). He recognized the beginning of proof by contradiction in light of the goals stated in L3. Less productively, it appeared that he only became aware of the "no common factors" stipulation in the definition of PPT when it was used in L8. He initially tried to make sense of that line in terms of the "no common factors" claim in the theorem's statement. In this middle section of the proof, he seemed to lack a clear sense of "no common factor" claims were known and which required justification. As a result, his emerging construal of the proof began to completely reverse the intended relationship between hypotheses and conclusions. Nov1 reached the point of claiming that L12 proved that property held for (a, b, c) rather than for $((c + b), (c - b))$.

We argue that Nov1's weak image of what was taken as hypothesis in the proof influenced the way that he confused the various "no common factor" claims. For lines that clearly stated the hypotheses and conclusions, he produced valid justifications. However, he never developed a clear sense of what the overall proof began assuming and how the set of claims proven grew over the course of the text. This is why we describe this as a *low-level construal of proof claims*. This account of Nov1's sense making of the text helps explain why he ended the reading unable to explain the necessity of the middle section of proof.

We observed other forms of this construct, especially among novices. This often seemed to result from a weak understanding of the underlying concepts. For instance, students who thought about "d is a factor of $(c - b)$" in terms of the process of dividing, rather than being made up of units of d, and students who had trouble thinking of $(c - b)$ as a unit all tended to have trouble building a mental model of what was assumed and what needed to be shown. Like Nov1, such students ended up trying to draw inferences from the equations in the theorem's conclusion because they seemed to provide richer resources for sense making.

Ongoing Revision of Proof Construal

This construct represents a complex form of bridging inference (Magliano et al., 2011). It describes when students revised their existing model of the proof's prior claims in light of later lines. As an example, Experienced 5 (Exp5) could not recall which claim was assumed as true in the wording "Euclid's Fifth Postulate (EFP) implies Playfair's Parallel Postulate (PPP)." Because the proof begins with the hypotheses of PPP (Zandieh, Roh, & Knapp, 2014), he inferred that "implies" meant to assume PPP and prove EFP. Exp5 initially interpreted that L1 assumed PPP was true. It was not until L6 when the proof applied EFP that the student decided EFP was the hypothesis and PPP the conclusion. He supported this by revising his understanding of L1 as assuming only the hypotheses of PPP rather than assuming the entire claim.

Some novice readers exhibited less productive examples of ongoing revision when they read L8 of the Pythagorean triples proof. They inferred that the contradiction denied the hypothesis in L1 rather than the hypothesis "d is a factor of both b and c" from L7. Once they concluded that L8 stated that (a, b, c) is not a primitive Pythagorean triple, they rightly expressed difficulty making sense of the argument when the object in question was not in the relevant category. Our moment-by-moment methodology uniquely provides access to this type of ongoing forming and reforming of models for what proofs claim to be true.

Identifying and Stating Warrants

The final notable pattern of ROMP activity we present in this report dealt with the ways students sought and stated warrants for inferences made in proofs. The interview protocol often invited students to explain why particular lines were justified, which for us meant to identify warrants. More experienced readers tended to be more adept at this practice and we observed key differences among the kinds of warrants sought and produced. Nov1's reasoning about L5 above exemplifies an *enacted warrant* in which he justified the inference by describing how particular manipulations could be made to show that d would be a factor of an expression. This constituted a mini-proof of the relevant warrant. Nov1's reasoning about L9 above is an example of *justification by example*, which does not constitute a valid warrant, but nevertheless provides some support for the claim. More experienced students were more often observed trying to *state warrants in general form*. For example, they articulated that L5 is justified because the sum [or difference] of any two multiples of d is also a multiple of d. Finally, Graduate 1 was able to *cite a relevant warrant* for L5, namely that any linear combination of multiples of d is also a multiple of d. Across our interviews, we observed a range of ROMP activities within which students with more experience exhibited greater tendency to seek warrants and where more adept at identifying particular inferences as instances of a general mathematical fact.

Discussion

This paper presents findings from our adaptation of the moment-by-moment reading assessment methodology to the reading of mathematical proof. We identified several novel ROMP activities that emerged in our interviews that justify the value of the methodology. The first phenomena distinguishes between the kinds of practices that students used to make sense of the proof texts and relates to the textual metafunction of mathematical texts. We anticipate that this finding that novice readers try to make sense of proofs using expectations from other mathematical texts could be fruitfully explored in the context of introduction to proof instruction. This pattern of ROMP sense making may help explain why Inglis and Alcock (2012) found that novice readers attended more closely to equations in proofs while experts attended to the surrounding text, which contains logical connectives. We hope that these other ROMP activity constructs can be further harnessed in later investigations to better understand how students make meaning of proofs they read and how that process develops over time. Ongoing work intends to find ways to adapt this methodology into an efficient assessment tool that can be more quickly administered and coded. This will contribute more insights about the process of reading to supplement the existing assessments of end reading comprehension and proof validation.

References

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

Dawkins, P.C. & Zazkis, D. (2018). Computational and Inferential Orientations: Lessons From Observing Undergraduates Read Mathematical Proofs. *Proceedings of the 40th Annual PME-NA Conference.* Raliegh, SC.

Fletcher, C.R., Lucas, S., & Baron, C.M. (1999) Comprehension of Mathematical Proofs. In S.R. Goldman, A.C. Graesser, & P. van den Broek (Eds.) *Narrative Comprehension, Causality, and Coherence: Essays in Honor of Tom Trobasso* (pp. 195-207). Mahwah, NJ: Lawrence Erlbaum Associates.

Halliday, M.A.K. (1985). *An Introduction to Functional Grammar*. London, UK: Edwards Arnold.

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

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, *21*(3), 251-283.

Magliano, J. P., Millis, K. K., The RSAT Development Team, Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and learning*, *6*(2), 131-154.

Mejia-Ramos, J. P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics*, *79*(1), 3-18.

Mejía-Ramos, J. P., Lew, K., de la Torre, J., & Weber, K. (2017). Developing and validating proof comprehension tests in undergraduate mathematics. *Research in Mathematics Education*, *19*(2), 130-146.

Panse, A., Alcock, L., & Inglis, M. (2018). Reading Proofs for Validation and Comprehension: an Expert-Novice Eye-Movement Study. *International Journal of Research in Undergraduate Mathematics Education*, 1-19.

Rotman, J. J. (2013). *Journey into mathematics: An introduction to proofs*. Courier Corporation.

Samkoff, A., & Weber, K. (2015). Lessons learned from an instructional intervention on proof comprehension. *The Journal of Mathematical Behavior*, *39*, 28-50.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ; Lawrence Erlbaum Associates.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, *48*(1), 101-119.

Weber, K. (2015). Effective proof reading strategies for comprehending mathematical proofs. *International Journal of Research in Undergraduate Mathematics Education*, *1*(3), 289-314.

Yang, K. L., & Lin, F. L. (2008). A model of reading comprehension of geometry proof. *Educational Studies in Mathematics*, *67*(1), 59-76.

Zandieh, M., Roh, K. H., & Knapp, J. (2014). Conceptual blending: Student reasoning when proving "conditional implies conditional" statements. *The Journal of Mathematical Behavior*, *33*, 209-229.