



ELSEVIER

Contents lists available at SciVerse ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures



Reasoning-and-proving opportunities in elementary mathematics textbooks

Kristen N. Bieda*, Xueying Ji, Justin Drwencke, Andrew Picard

Michigan State University, United States

ARTICLE INFO

Article history:

Received 29 November 2012
Received in revised form 10 June 2013
Accepted 18 June 2013
Available online xxx

Keywords:

Elementary grades
Reasoning
Proof
Curriculum analysis

ABSTRACT

Over the past two decades, standards documents have emphasized the importance of developing students' abilities to generate and critique mathematical arguments across all grade levels. However, little is known about the opportunities elementary textbooks provide for students to learn mathematical argumentation. We analyzed seven upper elementary (ages 9–11) mathematics textbooks published in the U.S., focusing specifically on reasoning-and-proving opportunities in written tasks, and found that the average percentage of such tasks was 3.7%. Further, analyses of the task purpose and type of justification warranted revealed distinctions between the text materials in terms of the kinds of reasoning-and-proving activities prompted and the placement of tasks in the lesson sections. Specifically, textbooks developed based on research and written to align with curriculum and instruction standards were more likely to have reasoning-and-proving tasks within the narrative and student exercise sections than other texts. We discuss implications for the opportunities to learn reasoning-and-proving in elementary classrooms.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Advocates of reform in school mathematics often declare that developing mathematical proficiency involves gaining capacity to engage in mathematical habits of mind (Cuoco, Goldenberg, & Mark, 1996) that promote not only procedural fluency but also conceptual understanding, adaptive reasoning, and strategic competence (Kilpatrick, Swafford, & Findell, 2001). The most recent mathematics education reform in the U.S., the Common Core State Standards for Mathematics (CCSS-M) (CCSSI, 2010), now adopted by 45 states and three territories, aims to achieve greater coherence in the topics taught across grade levels so that students may learn content at a sufficient depth and level of mastery as to not need to be re-taught the content again in the next grade level (Schmidt & Prawat, 2006). However, coherence in students' learning of mathematical processes, such as reasoning and proving, across grade levels has not garnered the same attention as procedural and conceptual understanding of mathematical content.

The CCSS-M presents a list of eight mathematical practices that are described to be “varieties of expertise that mathematics educators at all levels should seek to develop in their students” (p. 6). Several of the practices in the CCSS-M were inspired by Cuoco et al. (1996) mathematical habits of mind, and relate to reasoning and proving, namely: (1) “Reason abstractly and quantitatively”; (2) “Construct viable arguments and the reasoning of others”; (3) “Look for and make use of structure” and; (4) “Look for and express regularity in reasoning” (pp. 7–8). Each of these practices entails mathematical reasoning, which is the process of making sense of and understanding mathematical ideas and concepts inherent to

* Corresponding author. Tel.: +1 517 432 9925.
E-mail address: kbieda@msu.edu (K.N. Bieda).

procedures. Students use reasoning when they engage in mathematical argumentation, a process that involves making and justifying mathematical claims. Proof, a specialized form of argumentation, is a process to show that a claim is always true (or false), governed by disciplinary norms establishing the modes of reasoning and representational forms are appropriate for valid proof.

The CCSS-M *Standards for Mathematical Practice* are similar to aspects of the Reasoning and Proof Standard from the 2000 *Principles and Standards for School Mathematics* (NCTM). Like the CCSS-M, the NCTM *Principles and Standards* positioned the Reasoning and Proof Standard as applying to grades K-12. However, the broad aims of the CCSS-M practices and the Reasoning and Proof Standard, and the lack of research-based learning trajectories for reasoning and proving, make it difficult to operationalize what counts as appropriate proof at a particular grade level or grade band.

Given that engaging in mathematical proof is a complex practice that demands conceptual understanding and strategic competence, among other aspects of mathematical proficiency (Kilpatrick et al., 2001), it is not surprising that studies of students' proficiency to engage in reasoning and proving indicate significant weaknesses in students' abilities to reason deductively and abstractly at all grade levels (Chazan, 1993; Harel & Sowder, 1998; Healy & Hoyles, 2000; Hoyles & Küchemann, 2002; Knuth, Choppin, & Bieda, 2009; Lannin, 2005). One possible factor contributing to these findings is a lack of consistent, coherent experiences with reasoning and proving across grades K-8 that may lead to difficulties in learning to construct formal proofs in high school geometry and beyond.

This paper presents results from a study investigating the nature of opportunities to engage in reasoning-and-proving in elementary mathematics textbooks. The aims of this investigation were to learn more about the existing opportunities students have to understand the kinds of reasoning that yield valid justifications and the opportunities presented to engage in reasoning-and-proving. Specifically, the general aim of this study was: *What opportunities exist in student text materials for students to engage in reasoning-and-proving, such as making claims, justifying claims, and evaluating claims?*

2. Empirical and conceptual foundations

2.1. A definition of reasoning-and-proving in elementary mathematics

Researchers interested in the teaching and learning of argumentation in school mathematics have long been concerned with trying to accurately define the phenomena that are the foci of their study. Balacheff (1988) asserted a definition of what constitutes proof in school mathematics, while in 2002 posited that the field is hampered by a lack of consensus over what the words “proof” or “justification” in school mathematics means. Stylianides (2007a) aimed to address this lack of consensus by presenting the following definition, now widely cited in research on reasoning-and-proving in school mathematics:

“Proof is a mathematical argument, a connected sequence of assertions for or against a mathematical claim, with the following characteristics: it uses statements accepted by the classroom community... it employs forms of reasoning that are valid... and it is communicated with forms of expression that are appropriate...”(p. 291).

We acknowledge that an important component of Stylianides' (2007) definition is that it emphasizes the knowledge and beliefs of the classroom community in determining whether a mathematical argument was a proof within the classroom context. As Zack (1999) eloquently wrote, the nature of children's discourse around justification and proof seems informal and colloquial on the surface but a closer examination reveals that it involves quite sophisticated mathematical thinking. Even without explicit instruction about mathematical proof, elementary grades students engage in generalization, justification, and refutation spontaneously when determining the validity of answers with their peers (Zack, 1999).

This study, however, is far removed from the enactment of reasoning and proving tasks in classrooms in its focus on classifying the kinds of tasks related to mathematical argumentation present in elementary textbooks. Our study was informed by the conceptualization of *reasoning-and-proving* (RP) (Stylianides, see Editorial in this issue) as a wide range of activities involved in mathematical argumentation, including generalizing the mathematical relationship in a given pattern, producing a conjecture, generating a justification or proof, and evaluating a given justification or proof. For our study, we employed the following conceptions to operationalize our understanding of RP tasks at the elementary grade levels: *reasoning* involves engaging in processes to generalize mathematical phenomena and/or conjecturing about mathematical relationships, whereas *proving* involves justifying a mathematical claim to be true for the domain to which the claim applies, using logically valid reasoning. Defining the activities of RP in this way allows for the intent of tasks to be interpreted and classified in ways that honor the use of proof in the discipline of mathematics, but, following Gardiner (1992) as cited in Zack (1997), do not constrain these activities to particular kinds of contexts, discourse, or representational forms. We argue that adopting a different, but complimentary, definition for analyses of RP tasks in curriculum is necessary as curriculum, student thinking, and teaching are distinct yet interconnected parts of the instructional triangle (Cohen, Raudenbush, & Ball, 2003) and the instantiation of RP within each part of the instructional triangle is uniquely designed for the functions enacted by that part.

2.2. Learning to prove in elementary mathematics

There is a relative wealth of research on RP in secondary mathematics when compared to elementary mathematics. Unlike many of the studies of reasoning-and-proving at the secondary level, which assess students' capacities to generate

mathematical proofs through interviews and written assessments, the majority of the work at the elementary level focuses on the development of elementary students' abilities to produce generalizations and reason both inductively and deductively about those generalizations through design-based research, case studies, and within the context of classroom lessons (Maher & Martino, 1996; Reid, 2002; Stylianides, 2007b; Zack, 1997). Maher and Martino conducted a 5-year longitudinal case study of six students through grades 1–5, and found that students were able to construct proofs by cases within the context of combinatorics problems, such as stacking different color cubes to make towers, without receiving explicit instruction in generating proof. The researchers identified the role of the classroom culture, which encouraged an open sharing of ideas, as essential to fostering mathematical argumentation in this elementary school classroom. Likewise, Stylianides (2007a) examined the development of 3rd grade students' conceptions of valid justification in a classroom taught by Deborah Ball, and found that students were capable of producing arguments that represented valid mathematical proof although the representational forms used to communicate the proof did not follow conventions of representing proof in the discipline of mathematics. However, there also seems to be reluctance in the literature on elementary students' reasoning to label classroom activities as "doing proof" or what elementary children produce as "proofs." Zack (1999) describes the nature of elementary students' discussions while proving as having the quality of not being formal and rigorous, but nonetheless children are capable of engaging in deductive reasoning about mathematics. The existing literature, both tacitly and explicitly, reveals how RP as a mathematical activity is conceptualized in terms of its role more in the discipline of mathematics rather than as a way for students to gain a deeper understanding of school mathematics.

While research shows that students are developmentally capable of generating proof even at the elementary school level, teachers must create an environment that nourishes this capability. Studies of proving in elementary classrooms describe lessons featuring proof-related tasks with instruction that utilizes effective forms of questioning which press students to produce claims and justify those claims (Zack & Reid, 2003, 2004). Existing research has not, however, focused on identifying the features of written tasks that are especially important in building students' capacities to generalize and justify in the elementary mathematics curriculum, nor upon what RP opportunities currently exist in elementary textbooks. Considering renewed interest in developing students' abilities to produce and critique mathematical arguments (CCSSI, 2010), we argue that one area of school mathematics in need of research is on the kinds of tasks that are used to engage elementary students in RP in both the intended and enacted curricula.

3. Research questions

This study investigates one aspect of learning to engage in RP in elementary mathematics: the opportunities to learn present in elementary mathematics textbooks. Specifically, we seek to answer the following general questions: *What opportunities to learn about mathematical reasoning-and-proving exist in elementary grades curricula? What aspects of reasoning-and-proving are prompted by tasks in elementary grades curricula and how frequently and consistently do those opportunities appear in the text materials?*

To address these questions, we focused specifically on determining the frequency of such tasks as compared to the total tasks in the student text materials and consider the implications of the placement of such tasks within the curricular lessons and units. We also focused on how the tasks written to elicit RP prepared students to engage in the types of proving tasks they might be expected to do in middle school and beyond, such as a two-column proof in high school geometry. We acknowledge that different results might have been obtained if we had considered elementary students' engagement with tasks that focus on broader range of activities in the proving process, such as classifying mathematical objects according to given definitions, rather than the more complex RP tasks that they might be asked to do in secondary mathematics. We chose to focus on students' experiences with tasks that engage students in more aspects of the proving process so that our findings could be compared and considered with similar curriculum analyses at the middle and high school levels.

4. Methods

We chose to conduct *content analysis* (Shapiro & Markoff, 1997) to analyze and compare the content of written textbooks using our conceptualization of RP. In this section, we describe the sample of textbooks analyzed, followed by a discussion of the unit of analysis and the analytic framework of the study, and finally describe the procedure used to conduct the content analyses.

4.1. Sample

4.1.1. Textbook series

We conducted analyses of seven elementary mathematics textbooks, namely *Everyday Mathematics* (UCSMP, 2007), *Investigations in Number, Data, and Space* (TERC, 2008), *Foresman-Wesley Grade 5 Math* (Charles et al., 1999), *Silver Burdett Ginn Mathematics* (Fennell et al., 2001), *Math Expressions* (Fuson, 2009), *Trailblazers* (University of Illinois-Chicago, 1998), and *enVisionMATH* (Charles et al., 2009), focusing our analyses on the student text materials and, when needed, the teacher's guide or support materials. At present, there is no single elementary textbook series widely used in the United States. Dossey, Halvorsen, and McCrone (2008) state that the three most widely used textbook series emerging from data collected in 2004–2005 only accounted for little more than 40% of the books used in grades K–5. Thus, we selected texts for our sample as they represented a variety of approaches to aligning with *Standards-based* recommendations (NCTM, 1989, 2000).

The U.S. National Science Foundation funded the development of *Investigations in Number, Data, and Space* (for short, *Investigations*), *Everyday Mathematics*, and *Trailblazers* to be aligned with the 1989 NCTM *Curriculum and Evaluation Standards* (Senk & Thompson, 2003). In addition, *Investigations* explicitly states that the development team utilized research and recommendations from the *Principles and Standards for School Mathematics* (NCTM, 2000) and *Adding It Up* (Kilpatrick et al., 2001) when creating and revising the curriculum. Also, Houghton Mifflin Harcourt explicitly states in their promotional materials for *Math Expressions* that the curriculum is *Standards*-based and was developed with funding from the U.S. National Science Foundation. The two most widely used elementary *Standards*-based texts were *Everyday Mathematics* and *Investigations*, but they only accounted for 18% of the total books in use for grades K-5 (Dossey et al., 2008).

Our analyses also considered commercially published texts that were not explicitly *Standards*-focused in their development, namely Scott Foresman-Addison Wesley *Grade 5 Math*, Silver Burdett Ginn *Mathematics*, and Scott Foresman-Addison Wesley *enVisionMATH*. Scott Foresman-Addison Wesley is a division of Pearson, and Silver Burdett Ginn was acquired by Pearson in 1998 but no longer publishes texts with the Silver Burdett Ginn label. *enVisionMATH* is the latest mathematics text to be published by Pearson under the Scott Foresman-Addison Wesley label. Considering that 67% of schools are using textbooks published prior to 2001 (Dossey et al., 2008), we also included the now-obsolete Silver Burdett Ginn *Mathematics* and Scott Foresman-Addison Wesley *Grade 5 Math* in our analysis to increase the likelihood of representing the nature of the written curriculum in use across elementary mathematics classrooms in the United States.

4.1.2. Grade level

We chose to analyze texts from grade 5 based on the assumption that grade 5 content represents the most advanced mathematics content of the upper elementary mathematics curriculum, and the likelihood that students would be engaged in RP tasks would be greater in grade 5 materials. In some districts within the U.S., students may be in an elementary school in grade 6, but in other districts they would be in a middle school or junior high. Since many middle school curricula include a text for grade 6, we chose to name grade 5 as an upper elementary grade level. With that determination, grade 5 represents a transition point to middle school content in applicable districts, and it may prepare students for work with a textbook series at the middle school level like *Connected Mathematics Project* (CMP) (Lappan, Fey, Phillips, Fitzgerald & Friel, 2003). Stylianides (2009) found that approximately 40% of tasks in CMP engaged students in RP. Thus, our analyses attempted to discern the degree to which the tasks in selected grade 5 texts were written to engage students in learning to reason and prove in ways similar to those in CMP.

4.2. Data analysis

4.2.1. Unit of analysis

Given that our aim was to determine what opportunities to learn about RP were present in elementary schools, and our sample spanned seven curricula, we chose to code every other lesson or investigation in each unit or chapter in each textbook starting with the first chapter. Within each unit, we identified the placement of RP tasks in one of four possible sections of a lesson (for these sections and how they were identified, see Section 4.2.3). Within each of the sections considered, coders identified problems considered to be opportunities to learn RP. One problem may have contained multiple, lettered parts. Each part was coded as a separate problem. The total number of possible problems in the sample (counting lettered parts of problems as one problem) was 28,210.

4.2.2. Analytic framework

We adapted our analytic framework from existing frameworks (Thompson, Senk, & Johnson, 2012; Stylianides, 2009). As Table 1 shows, problems were coded for *Type of Problem*, as well as *Purpose of RP Problem*, *Intended Outcome of RP Problem*, and *Type of Argument Elicited*. The *Purpose of RP Problem* code describes the type of RP process elicited by the problem, and our coding scheme evolved from a synthesis of Thompson, Johnson and Senk's and Stylianides' frameworks. The *Intended Outcome of RP Problem* code was designed to classify whether a problem as intended solicited students to provide a mathematical proof. The "proof-type argument" and "non-proof argument" are similar to Stylianides' *proof precursor* and *non-proof precursor*. Finally, codes assigned for *Type of Argument Elicited* differentiated RP problems by the kind of justification elicited using four sub-codes developed by Stylianides (2009). As will be discussed in Section 4.2.3 below, the first three categories (*Type of Problem*, *Purpose of RP Problem*, and *Intended Outcome of RP Problem*) are mutually exclusive

Table 1
Analytic framework for curriculum analysis.

Categories	Type of Problem	Purpose of RP Problem	Intended Outcome of RP Problem	Type of Argument Elicited
Codes	•Narrative	•Making claims	Proof-type Argument	•Demonstration •Generic Example
	•Student exercise •Extensions	•Justifying claims •Making claims & justifying claims	Non-proof Argument	•Empirical •Rationale
	•Assessment	•Evaluating justifications		

sets of codes, however codes assigned in the fourth category (Type of Argument Elicited) depended to some extent on the code given for Intended Outcome of RP Problem.

4.2.3. Procedure

The first phase of coding involved identifying problems from a page of text that contained the following keywords and phrases: explain, describe, predict, show, write a rule, tell why, tell how, justify, and prove. Once problems were identified by a keyword, coders first assigned the Type of Problem code (narrative, student exercises, extensions, assessments). *Narrative* sections were considered to be any section where the exposition of the lesson content was presented. Only the narrative sections available in the student text materials were considered, and within those sections, only the text stated as a question or a command (such as “Find the GCF of 12 and 24) was considered as a “problem” for coding purposes. The text for these sections preceded *Student Exercise* sections. Any sections labeled “Practice”, “Homework”, “Student Exercise” or other related classifications, and were not embedded within the exposition of the lesson (if provided in the student materials), were considered to be sections of student exercises. Sections that were labeled “Enrichment”, “Extra Practice”, “Reflection”, “Extension” or other related terms, that contained content related to the lessons in the unit but appeared either between two lesson sections or at the end of a chapter or unit were considered to be *Extension* sections. Finally, if a section was labeled “Test” or “Assessment” and occurred at the end of a series of lessons, chapter, or unit, problems coded from those sections were labeled as *Assessment* problems.

Following the assignment of a Type of Problem code, we interpreted contextual cues in a problem to determine the Purpose of RP Problem code. For problems coded as *Making Claims*, the problem prompted students either to generate a rule for a pattern, a conjecture, or provide a response indicating whether or not they agreed with a stated claim. For example, a question like “Josie says that the sum of two even numbers will be even. Is Josie correct?” would have been coded as *Making Claims* because the students are being asked to make a claim about whether Josie’s conjecture is true or false. We included a category for *Justifying Claims*, as a problem could provide a claim and ask students to generate a proof for the claim. For example, a problem such as “Josie claims that the sum of two even numbers is even. Provide a justification to support Josie’s claim” would have been coded as *Justifying Claims*. If a problem included directions for students to produce a claim and explain their reasoning or justify their response, the code *Making Claims and Justifying Claims* was assigned. Finally, we also included a category for *Evaluating Justifications*. This category encompassed any problems asking students to review either one or multiple justifications for a claim and determine whether the justifications were convincing and/or valid. It is worth noting that not all problems that contained the relevant keywords and phrases engaged students in making or justifying claims that applied to a wide, or infinite, set of cases (i.e. the sum of two odd numbers is even). Many problems we included asked students to make a claim about a finite case, such as making a claim about a solution to a word problem or determining whether a conjecture was true for numbers between 1 and 10.

After coding for Purpose of RP Problem, coders assigned codes for Type of Argument Elicited and Intended Outcome of RP Problem. Table 2 provides archetypes of problems coded across all possible codes for each general category.¹

To code for Intended Outcome of RP Problem, that is, to code a problem either as eliciting a proof argument or a non-proof argument, we interpreted the written text to determine the type of response being elicited and, when necessary, consulted the teacher’s guide materials to infer what type of response was expected from students. In some cases, the problem might indicate in its formulation that students were to generate a particular type of response. Statements such as “Provide a proof that your conjecture is true” or “Generate three examples to support your answer.” For other problems, the teacher’s guide was consulted to determine what the intended correct response. We then coded the suggested answer in the teacher’s guide based on whether or not it was a proof-type argument, satisfying our definition of proof. In the examples presented in Table 2, we determined one example to be a proof-type argument because the answer provided in the teacher’s guide referenced a key definition that students would need to use to justify the claim more generally. However, for the example coded as a non-proof argument, the teacher’s guide solution only provided a generalized description of the cases to which the pattern applied, but did not provide any statements that students would be expected to use when justifying that the pattern would hold for such cases. There were several cases where the teacher’s guide provided no suggested answer, or listed that “Answers may vary.” For these problems, we assigned the code “Non-proof argument” as the materials did not specifically state that the expected response should be in the form of a proof-type argument.

The last phase of analysis involved applying the definitions provided by Stylianides (2009) of the sub-codes for Type of Argument Elicited to further categorize problems identified as RP opportunities. For problems coded as Non-proof arguments, the explicitly or implicitly intended responses received one of two codes: *Empirical* or *Rationale*. Problems coded as *Empirical* either contained explicit instruction for students to generate example cases in their response, or the answer suggested in the teacher’s guide included examples. In contrast, problems coded as *Rationale* had solutions in the teacher’s guide that were explanations of how the problem should be solved, or justifications that referenced applicable statements, theorems or properties but did not logically validate the claim for all relevant cases.

For problems coded as Proof-type argument, we assigned either *Generic Example* or *Demonstration* codes to describe the responses either explicitly requested in the written task or described in the teacher’s guide. Problems receiving a

¹ Problems did not receive more than one code within a single category. However, since the three categories are distinct, an example may also have been assigned another code from a different category.

Table 2

Examples of codes for Purpose of RP Problem, Intended Outcome of RP Problem, and Type of Argument Elicited.

Category	Code	Example
Purpose of Problem	Making claims	"Write a rule that tells for which shapes the curve will work and for which shapes it will not work." (Charles et al., 1999, p.19)
	Justifying claims	"Are both Felicia's and Lee Yah's formulas for the perimeter of a rectangle correct? Explain your reasoning." (Math Trailblazers, 1998, p. 488)
	Making and justifying claims	"Suppose you have sheets of grid paper that are 5×5 . Would you need more or less of these sheets than the 10×10 sheets to show one million squares? Explain." (Charles et al., 1999, p. 52)
	Evaluating justifications	Does Jamirah's argument for why the sum of two odd numbers is even convince you that this is true for all whole numbers? Why or why not? (None found in actual sample)
Intended Outcome of RP Problem	Proof-type argument	"Examine your model of a cube. Does the cube have more edges than vertices, the same number of edges as vertices, or fewer edges than vertices? Is this true for all polyhedrons? Explain". Teacher's Guide solution: "Yes. At least three edges are needed to form one vertex." (UCSMP, 2007, p. 369)
	Non-proof argument	"Do you think this pattern also works for the problems like $1/8 + 1/3$? Explain." Teacher's Guide solution: "Yes. This pattern will work whenever two unit fractions are added." (UCSMP, 2007, p. 225)
Type of Argument Elicited	Empirical	"How many pairs should you test to see if your rule works?" (Charles et al., 1999, p.22)
	Rationale	"Without finding the quotient, tell which quotient is greater. Explain your choice. $251/7$ or $175/8$." (Fennell et al., 2001, p. 186)
	Generic Example	"Using any two numbers, is it possible to make a quotient with three digits? Explain. (Numbers given: 10, 20, 8240)" (Fennell et al., 2001, p.240)
	Demonstration	"Examine your model of a cube. Does the cube have more edges than vertices, the same number of edges as vertices, or fewer edges than vertices? Is this true for all polyhedrons? Explain."Teacher's Guide solution: "Yes. At least three edges are needed to form one vertex." (UCSMP, 2007 pp. 369)

Demonstration code had solutions in the teacher's guide with justifications that would address the entire set of cases for which the claim applied and used complete, logical and valid reasoning without referencing example cases. On the other hand, problems receiving a *Generic Example* code did have solutions in the teacher's guide that contained empirical evidence, but also provided a logical explanation justifying why the claim would hold for other cases for which the claim applied.

In assigning the Type of Argument Elicited code, there were situations where problems had no suggested response in the teacher's guide. In those cases, the coding team met to discuss what responses would be plausible for a fifth grade student to produce, and then coded the problem based on the most mathematically sophisticated response probable. For instance, with Non-proof Argument problems, we utilized the Rationale code when we believed students were capable of producing some sort of verbal explanation, following Stylianides (2009) formulation of rationales as a transition between empirical reasoning and proof-type reasoning. Such instances included problems written to elicit explanations of why a conjecture was true, but did not suggest in any way that students draw upon definitions, accepted statements, or properties or validate the conjecture is true for the entire domain of cases for which the claim applied.

4.2.4. Inter-rater reliability

To obtain inter-rater reliability, each of the two coders on the research team independently coded the first two textbooks analyzed for the project (Foresman-Wesley *Grade 5 Math* (1999) and Silver Burdett Ginn *Mathematics* (2001)). Another member of the research team selected a sample of sections coded in the and compared coders' ratings for (1) the problems identified by justification and proof keyword (see Section 4.2.3) by each coder; (2) the assignment of Type of Problem code; (3) the assignment of Purpose of RP Problem code; and (4) the assignment of Type of Argument Elicited code. The remaining textbooks were not analyzed until coders reached acceptable agreement. After discussing instances of initial disagreement in the coding and revising initial codes, the inter-rater reliability on identifying RP problems was approximately 90%, assigning Type of Problem codes was 100%, assigning Purpose of RP Problem was 83% and the inter-rater reliability was 75% for Type of Argument Elicited codes assigned.

5. Results and discussion

This study investigated the prevalence of RP tasks in elementary text materials, and found that such tasks were a very small percentage of the total number of potential tasks for the materials we reviewed. Based on a review of the literature about curriculum analyses of secondary mathematics texts, we hypothesized this would be the case. However, our findings reveal important differences between the texts as to where RP tasks appeared as well as the proportion of proof-related tasks to overall possible tasks. In this section, we begin by presenting overall findings across the seven text materials that were analyzed. Overall results are followed with a presentation of detailed data from the text materials that offered the most RP tasks an examination of the findings regarding task placement within the textbooks.

Table 3
Frequency of reasoning-and-proving problems by textbook.

Name of text materials	Total number of proof-related problems	Percentage of total possible problems (%)
<i>Everyday Mathematics</i>	76	2.8
<i>Foresman-Wesley</i>	363	4.2
<i>Math Expressions</i>	198	3.9
<i>Investigations</i>	50	2.9
<i>Silver Burdett Ginn</i>	144	2.9
<i>Trailblazers</i>	78	3.8
<i>enVisionMATH</i>	141	4.6
Total	1050	3.7 (1050 problems out of 28,210 total possible problems)

Table 4
Frequency of reasoning-and-proving problems by textbook type.

Type of text materials	Total number of proof-related problems	Percentage of total possible problems (%)
NCTM Standards-based (<i>Everyday Mathematics</i> , <i>Investigations</i> , <i>Trailblazers</i> , <i>Math Expressions</i>)	402	3.5
Other (<i>Foresman Wesley</i> , <i>Silver Burdett Ginn</i> , <i>enVisionMATH</i>)	648	3.9
Total	1050	3.7

5.1. Frequency of reasoning-and-proving problems

Across the six textbooks, a chi-square test revealed no significant differences in the proportions of total problems that were related to RP (see Table 3). When clustering the results by type of curriculum program (NCTM Standards-based, Other), no significant differences arise (see Table 4).

5.2. Frequencies of reasoning-and-proving problems with different purposes

As illustrated in both Tables 3 and 4, the total number of RP problems in each of the textbooks was quite low. We also found vastly uneven distributions of problems with different purposes related to RP. Table 5 shows frequencies for each of the four different purposes coded within each textbook.

There are several key findings shown in Table 5. First, note that there are far more problems asking students to both make and justify claims across all of the seven textbooks analyzed than for the other three purposes coded. Table 2 provides examples of problems coded as simply justifying claims as well as making and justifying claims, and a key distinction between the two types of purposes is whether students are required to generate a claim. In terms of developing students' capacity to engage in the process of generating mathematical arguments, it may be preferable for students to experience opportunities to both create a claim and justify whether it is valid.

A second key finding is that *Everyday Mathematics* appears to achieve a better balance with RP problems involving different purposes than the other six textbooks. The textbook with the most skewed distribution of problems by different purpose is *Trailblazers* – approximately 84% of the RP problems involved making and justifying claims. This characteristic, however, may or may not be desirable. As stated earlier, it may be advantageous for students to have more opportunities to both generate and justify conjectures, however “their justifications might not always be sufficient to validate all the conjectures they are capable of identifying” (Carpenter & Levi, 2000, p. 16).

Another important finding from Table 5 suggesting how elementary students' opportunities to learn RP may be quite impoverished is the obvious lack of any problems involving evaluating claims. One possibility is that any such opportunities to evaluate claims were present in the lessons that were not sampled. It seems highly unlikely, however, to not encounter a single problem posing one or multiple possible justifications for students to evaluate in any of the sections reviewed across the seven textbooks. Another possibility is that the teacher support materials suggests opportunities for the justifications solicited from different students to be discussed as a whole class, thus designing moments during instruction where students

Table 5
Frequencies of problems by purpose of RP problem.

Purpose	Everyday Mathematics	Investi-gations	Trail-blazers	Math Expressions	Silver Burdett Ginn	Foresman-Wesley	enVision MATH	Total
Making claims	32 (42%)	17 (34%)	6 (8%)	33 (17%)	51 (35%)	77 (21%)	33 (23%)	249
Justifying claims	9 (12%)	2 (4%)	6 (8%)	32 (16%)	10 (7%)	59 (16%)	38 (15%)	156
Making and justifying claims	35 (46%)	31 (62%)	66 (84%)	133 (67%)	83 (58%)	227 (62%)	70 (50%)	645
Evaluating claims	0	0	0	0	0	0	0	0

can evaluate each other's justifications. Our review of the teacher's support materials with each text focused solely on expected answers to given problems, so we did not take into account any suggested strategies for implementing the problems in those materials. Thus, we acknowledge that such an analysis might yield more opportunities for students to engage in evaluating claims. Although, considering suggestions for implementation in textbook analysis assumes that teachers read and use such text support materials in planning and implementing instruction, and existing research shows there to be a wide variation in how teachers use support materials to guide instruction (Remillard & Bryans, 2004).

The absence of such problems in the written curriculum may indicate a key missed opportunity in curriculum development; Carpenter and Levi (2000) suggest that students may be able to recognize weaknesses in justifications at the elementary grades, while they are less likely to be able to produce arguments treating the general case. Therefore, providing more opportunities for students to evaluate justifications could be an important bridge from students' intuitive beliefs about convincing arguments to tasks that requires students to generate justifications.

5.3. Type of Argument Elicited in problems involving making and justifying claims

While making and justifying claims was the most prevalent type of RP problem across all of the textbooks, our coding of a problem's purpose allowed for problems that involved both proof and non-proof arguments. Although students in school mathematics are typically not expected to provide formal, proof-type arguments until middle or high school mathematics, we argue that distinguishing RP problems by either their explicit or intended proof potential reveals how a curriculum might prepare grade 5 students for more rigorous proof in secondary mathematics. Thus, analyses of the Type of Argument Elicited for problems asking students to make and justify claims allowed us to investigate whether such problems engage students in learning about valid modes of mathematical reasoning. The four codes: Empirical, Rationale, Generic Proof, and Demonstration (see Section 4.2.3. and Stylianides, 2009 for further description) represent a range of possible modes of mathematical reasoning, with Empirical being an insufficient mode of reasoning to verify the truth of a general claim and Demonstration being considered as the most general, valid form of reasoning. Table 6 shows the frequency of each of the possible types of arguments elicited in the problems involving both making and justifying claims.

Clearly, across all texts, the most prevalent Type of Argument Elicited in problems that involved making and justifying claims was empirical. In a few cases, explicit requests were made for students to provide examples to illustrate their reasoning, while in most cases the domain to which the problem applied was a single case or several cases. In these situations, an empirical or rationale-type justification was warranted (e.g., Making a claim about which of two given fractions is larger, and justifying the claim by discussing the difference between each given fraction and a benchmark fraction). These results show that, for the scant opportunities intended to engage students in making and justifying claims, students are rarely, if ever, asked to produce an argument that could apply to a general case. Existing research suggests that understanding and justifying a claim that applies to a general case may be out of the majority of elementary students' developmental capacity (Bell, 1976; Carpenter & Levi, 2000).

5.4. Dispersion of reasoning-and-proving problems

Time spent on task is a significant variable in determining opportunities to learn (Floden, 2002), thus we investigated the placement of RP problems within the units sampled as correlated to the likelihood that these problems would be implemented. Table 7 shows the dispersion of RP problems within textbook sections. Note that Table 4 only shows results of this analysis for problems with either the intended purpose to Justify Given Claims or to Make and Justify Claims.

Similar to Table 6 (see Section 5.3), this table shows how few proof-type arguments were elicited across any of the sections in the units sampled. Two new findings stand out in this table. First, a large proportion of problems in Foresman-Wesley *Grade 5 Math* (1999) appeared in the Extension sections as compared to the other six texts. We found the difference in the proportion of RP problems placed in Extension sections is statistically significant ($\chi^2(1, N = 311) = 7.429, p < 0.01$). The placement of these opportunities in Extension sections calls into question how often students will encounter these problems. If teachers stray from the way the lesson is outlined in the textbook, the first way that they will likely do so is by cutting out extensions. This scenario seems likely if the implementation of the narrative section takes longer than expected and the teacher is looking for a way to save time in completing the lesson.

Table 6
Type of Argument Elicited in making and justifying claims problems.

Purpose	Everyday Mathematics n = 35	Investigations n = 31	Trailblazers n = 66	Math Expressions n = 133	Silver Burdett Ginn n = 83	Foresman-Wesley n = 227	EnVision MATH n = 70
Empirical	22 (63%)	27 (87%)	47 (71%)	95 (71%)	66 (80%)	143 (63%)	45 (64%)
Rationale	12 (34%)	2 (6%)	18 (27%)	31 (23%)	13 (17%)	73 (32%)	24 (34%)
Generic Example	0 (0%)	1 (3%)	0 (0%)	4 (3%)	1 (3%)	2 (1%)	1 (1%)
Demonstration	1 (3%)	1 (3%)	1 (1%)	3 (2%)	3 (9%)	9 (4%)	0 (0%)

Table 7
Dispersion of problems.

Type of problem (location within text)	Type of Argument Elicited	Everyday Mathematics	Investi- gations	Trail- blazers	Math Expressions	Silver Burdett Ginn	Foresman- Wesley	EnVision MATH
Narrative	Proof	1	0	1	0	1	1	0
	Non-proof	23	11	30	80	20	30	11
Student Exercise	Proof	2	2	1	1	1	9	0
	Non-proof	16	18	23	48	40	89	67
Extension	Proof	0	0	0	1	1	1	1
	Non-proof	0	0	14	13	25	135	29
Assessment	Proof	0	0	0	1	1	0	0
	Non-proof	2	2	4	8	3	20	0

When collapsing the results shown in Table 7 by textbook type (see Section 4.1.1), another important finding emerges. In NCTM *Standards*-based texts, a greater proportion of problems were found in narratives and student exercises than in extensions sections. In fact, the difference in the proportion of RP problems coded as Narrative and Student Exercise for Type of Problem between *Standards*-based and non *Standards*-based texts is statistically significant ($\chi^2(1, N = 739) = 43.53, p < 0.01$). Houghton Mifflin Harcourt's *Math Expressions* (2009), in particular, stands out among the seven texts as there were 80 problems coded within the narrative section, far more than any other textbook. Although all of these problems involved justifying claims using a non-proof argument, the presence of these problems in the narrative may increase the likelihood that these questions would be posed during classroom implementation of the lesson. Likewise, UCSMP's *Everyday Mathematics* (2007) also had a slightly higher number of problems in the narrative section than in other sections.

6. Limitations

When undertaking a curriculum analysis of this scope, there are notable limitations that arise when choosing which analytic frameworks to apply, which methods to use to conduct the analysis, as well as how the texts will be coded and interpreted. The sample chosen for this study does not likely represent the full scope of students' opportunities to learn RP in elementary school, as only grade 5 texts were analyzed. Further, the choice to analyze only every other lesson within each unit of each textbook may have caused the research team to overlook rich opportunities in the text for students to engage in RP. Given that there were over 25,000 possible problems to analyze in the sample chosen, however, this decision greatly enhanced the feasibility of the study. An alternative approach could have been to analyze units or chapters in their entirety, but not sample from each chapter in the textbook. This approach, however, could have overlooked units or chapters with topics that lend themselves to more RP opportunities.

Another limitation of this study involved the choice to only apply content analyses to the student materials. As stated earlier, opportunities for the teacher to capitalize on students' responses and engage students in debating their justifications may have been present in the teacher materials but not recognized in our analyses. Additionally, a problem that looked to be, on the surface, not related to RP could be modified using recommendations from the teacher's guide to offer more RP potential. However, any analysis of teacher's guide also involves limitations; teachers may not use or understand the guidance provided in the teacher's guide.

Finally, an especially salient limitation of this work is the choice of analytic framework used to apply content analysis to the sample. This framework was minimally adapted from an existing framework to analyze RP opportunities in middle grades mathematics texts (Stylianides, 2009), but was not based on any developed theory of a trajectory of students' learning to reason and prove in elementary classrooms. It is likely that the kinds of experiences that can help prepare elementary students for understanding and producing more sophisticated mathematical arguments in future mathematics courses are not identical to those appropriate for secondary students but applied to elementary mathematics content. What this choice does allow, however, is for results that can be compared against other analyses of texts from middle school and high school mathematics that utilized a similar framework. It provides a picture of students' potential engagement with different types of tasks that can develop their abilities to comprehend and generate mathematical arguments across K-12 mathematics curriculum. As researchers continue to investigate the nature of opportunities that support students' understanding of RP across the grade levels, grade-level or grade-band-specific frameworks can emerge to articulate distinctions in the kinds of curricular opportunities that should be afforded to students at different points in their school mathematics careers.

7. Conclusion

The curriculum used in elementary mathematics classrooms may not be sufficient, as written, to provide students with meaningful opportunities to learn to generate and evaluate mathematical claims; the average percentage of tasks we sampled across seven elementary mathematics textbooks related to RP was only 3.7%. These findings suggest that the

usefulness of existing curricula for implementing the aims of the Standards for Mathematical Practices from the CCSS-M (CCSSI, 2010) may be limited. Our work raises a host of questions regarding how textbooks can support elementary students' learning to engage in RP: To what extent do elementary students need opportunities to consider the limitations of empirical arguments? Can, or should, elementary students justify properties or theorems that they will apply procedurally, or is it appropriate for students developmentally to convince themselves that properties and theorems are true by working several examples? Is it more likely for problems related to RP to be implemented if they are presented in Narrative sections? Another important consideration, although not addressed in this paper: How should properties and theorems be justified in exposition or Narrative sections? Should elementary grades texts present valid proofs to develop students' understanding of disciplinary norms? Answers to these questions can fundamentally shape the success of initiatives such as CCSS-M in making a core practice of disciplinary mathematics—proving – a regular and meaningful part of school mathematics.

References

- Balacheff, N. (1988). Aspects of proof in pupils' practice in school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp. 216–235).
- Balacheff, N. (2002). The researcher epistemology: A deadlock for educational research on proof. In F. L. Lin (Ed.), *In Proceedings of the international conference on mathematics* (pp. 23–44). Taipei, Taiwan: NSC and NTNU.
- Bell, A. W. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics*, 7, 23–40.
- Carpenter, T. P., & Levi, L. (2000). *Developing conceptions of algebraic reasoning in the primary grades*. NCISLA Research Report #00-2. Madison, WI: Wisconsin Center for Educational Research Retrieved from <http://ncisla.wceruw.org/publications/reports/RR-002.PDF>.
- Charles, R. I., Fennell, F., Caldwell, J., Cavanagh, M., Sammons, K., Schlielack, J., et al. (2009). *Scott Foresman-Addison Wesley enVision math: Grade 5*. Glenview, IL: Pearson Scott Foresman.
- Charles, R. I., Chancellor, D., Harcourt, D., Moore, D., Schlielack, J., Van de Walle, J., et al. (1999). *Scott Foresman-Addison Wesley math: Grade 5 (1 & 2)*. New York City, NY: Addison Wesley Longman.
- Chazan, D. (1993). High school students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, 24(4), 359–387.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 119–142.
- Common Core State Standards Initiative (CCSSI). (2010). Common Core State Standards for mathematics. Retrieved from http://corestandards.org/asserts/CCSSI_Math%20Standards.pdf.
- Cuoco, A., Goldenberg, E. P., & Mark, J. (1996). Habits of mind: An organizing principle for mathematics curricula. *Journal of Mathematical Behavior*, 15(4), 375–402.
- Dossey, J., Halvorsen, K., & McCrone, S. S. (2008). Mathematics education in the United States: A capsule summary fact book. Reston, VA: The Eleventh International Congress on Mathematics Education, Monterrey, Mexico. (N.C. o. T. o. Mathematics o. Document Number).
- Fennell, F. S., Ferrini-Mundy, J., Ginsburg, H., Greenes, C., Murphy, S., & Tate, W. (2001). *Silver Burdett Ginn mathematics: Grade 5*. Parsippany, NJ: Silver Burdett Ginn.
- Floden, R. E. (2002). The measurement of opportunity to learn. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 231–266). Washington, D.C. National Academies Press.
- Fuson, K. C. (2009). *Math expressions: Grade 5*. Houghton Mifflin Harcourt Publishers.
- Harel, G., & Sowder, L. (1998). Students' proof schemes. *Research in Collegiate Mathematics Education*, 3, 234–282.
- Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31, 396–428.
- Hoyles, C., & Küchemann, D. (2002). Students' understanding of logical implication. *Educational Studies in Mathematics*, 51, 193–223.
- Kilpatrick, J., Swafford, J., Findell, B., & U.S. National Research Council Mathematics Learning Study Committee. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academies Press.
- Knuth, E., Choppin, J., & Bieda, K. (2009). Proof: Examples and beyond. *Mathematics Teaching in the Middle School*, 15, 206–211.
- Lappan, G., Fey, J., Fitzgerald, W. M., Friel, S. N., & Phillips, E. D. (2003). *Connected mathematics*. Pearson Prentice Hall.
- Lannin, J. K. (2005). Generalization and justification: The challenge of introducing algebraic reasoning through patterning activities. *Mathematical Thinking and Learning*, 7, 231–258.
- Maher, C. A., & Martino, A. M. (1996). The development of the idea of mathematical proof: A 5-year case study. *Journal for Research in Mathematics Education*, 27(2), 194–214.
- National Council for Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: NCTM.
- National Council for Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Reid, D. A. (2002). Conjectures and refutations in grade 5 mathematics. *Journal for Research in Mathematics Education*, 33(1), 5–29.
- Remillard, J. T., & Bryans, M. B. (2004). Teacher's orientations toward mathematics curriculum materials: Implications for teacher learning. *Journal for Research in Mathematics Education*, 33(5), 352–388.
- Schmidt, W. H., & Prawat, R. S. (2006). Curriculum coherence and national control of education: Issue or non-issue? *Journal of Curriculum Studies*, 38(6), 641–658.
- Senk, S. L., & Thompson, D. R. (2003). *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Lawrence Erlbaum.
- Shapiro, G., & Markoff, J. (1997). A matter of definition. In C. W. Roberts (Ed.), *Text analysis in the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 9–31). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stylianides, A. J. (2007a). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, 38(3), 289–321.
- Stylianides, A. J. (2007b). The notion of proof in the context of elementary school mathematics. *Educational Studies in Mathematics*, 65(1), 1–20.
- Stylianides, G. J. (2009). Reasoning and proving in school mathematics textbooks. *Mathematical Thinking and Learning*, 11, 258–288.
- TERC. (2008). *Investigations in number, data, and space: Grade 5 (units 1–9)*. Glenview, IL: Pearson Education Inc.
- Thompson, D. R., Senk, S. L., & Johnson, G. J. (2012). Opportunities to learn reasoning and proof in high school mathematics textbooks. *Journal for Research in Mathematics Education*, 43(3), 253–295.
- University of Chicago School Mathematics Project. (2007). *Everyday mathematics: Grade 5 teacher's lesson guides (1 & 2)*. Chicago, IL: McGraw Hill Wright Group.
- University of Illinois-Chicago. (1998). *Math trailblazers: Grade 5*. Dubuque, IA: Kendall-Hunt Publishing Company.
- Zack, V. (1997). You have to prove us wrong: Proof at the elementary school level. In E. Pehkonen (Ed.), *In Proceedings of the twenty-first international conference for the psychology of mathematics education, Vol. 4* (pp. 291–298).
- Zack, V. (1999). Everyday and mathematical language in children's argumentation about proof. *Educational Review*, 51(2), 129–146.
- Zack, V., & Reid, D. A. (2003). Good-enough understanding: Theorising about the learning of complex ideas (part 1). *For the Learning of Mathematics*, 23(3), 43–50.
- Zack, V., & Reid, D. A. (2004). Good-enough understanding: Theorising about the learning of complex ideas (part 2). *For the Learning of Mathematics*, 24(1), 25–28.