Analyzing the Reasoning Types potentially Required by Tasks in an Open-Source Introductory Statistics Textbook.

Joash M. Geteregechi
Ithaca College

This study examined the tasks in the exercise sections of a popular open-source statistics textbook in order to characterize the types of reasoning the tasks are likely to elicit from readers. To achieve this, I examined the narrative sections of the book alongside the solved examples to determine the extent to which the kind of guidance provided was relevant in solving the tasks in the exercise sections of the book. Findings show that although most tasks could be solved by closely following the procedures and methods in the book (imitative tasks) there were other tasks that could elicit more nuanced forms of reasoning (creative mathematical reasoning tasks). These tasks were more prominent on the tail ends of the exercises section than on the beginning parts. Implications of these findings are discussed.

Keywords: Creative mathematical reasoning, Task analysis, Mathematics textbooks, Imitative reasoning.

Mathematics textbooks are among the most important resources for the teaching and learning of mathematics at the college level (Geteregechi & Waswa, 2020). Both instructors and students rely on mathematics textbooks to meet various learning needs such as self-study, practice, among others (Van Steenbrugge et al., 2013). Although it is difficult to exclusively characterize the role of textbooks in students learning of mathematics (Tarr et al. 2008), researchers (e.g., Stylianides, 2009) have reported that textbooks have a significant influence on how students make sense of and learn mathematics concepts.

One of the most important aspects of student learning is the kinds of reasoning that they engage in. Instructors often want to understand students' ways of thinking so they can better support their learning. However, what is often ignored is the sources that may inform students' thinking. In this paper, I argue that textbooks play a significant role in informing students' ways of thinking and thus there is need to examine the opportunities they present. Although there are many studies on the kinds of opportunities that mathematics textbooks provide for their readers (Thompson et al., 2012), not many such studies focus on how the narrative section of the texts influence the kinds of reasoning that students use for future task solving including tasks found in the exercise sections of the texts. I define the narrative section as the parts of a textbook where the readers learn the intended concepts and ideas. For most textbooks, the narrative section comprises of descriptions, illustrations, solved examples, among others. Exercise sections, on the other hand, comprise of the tasks intended to be completed by the readers of the text.

Most studies examining mathematics textbooks focus on calculus and geometry textbooks hence leaving a dearth of research in other important areas such as statistics. Furthermore, not many studies examine tasks from the perspective of the kinds of reasoning they are likely to elicit from readers. The term reasoning is often used without a clear definition of what it is and how to characterize it which means many studies on textbook analysis do not provide ways to characterize the potential of tasks to elicit specific kinds of reasoning. In this study, therefore, I sought to answer the following research questions:

1. What kinds of reasoning can be expected from tasks in a popular introductory open-source statistics textbook?
2. How are these tasks distributed within the exercises section of the book?

Related Literature

Most studies analyzing opportunities for reasoning in mathematics describe general characteristics associated with higher order thinking. Dolev and Even (2015), for example, examined the extent to which tasks in mathematics textbooks required students to provide justifications for claims made. The study compared six textbooks with a focus on geometry and algebra strands. Their findings indicated that geometry tasks tended to have more opportunities for justifying than algebra tasks across all textbooks. Similarly, Otten et al. (2014) examined the reasoning-and-proving opportunities available in secondary geometry textbooks and reported that although there were numerous reasoning and proving opportunities in both the narrative section and the exercises section of the books, these opportunities were not an object of focus in these books. While these studies provide insights into the kinds of tasks in textbooks, they rarely classify tasks themselves based on the reasoning types they are likely to elicit. This is due, in part, to the fact that the studies did not approach the analysis with a specified definition of reasoning. Thus, this study uses Lithner's well defined framework on reasoning to characterize tasks according to the kinds of reasoning they are likely to elicit. There are, however, some studies that have used specified definitions of reasoning in conducting their analysis. For example, Sidenvalli et al. (2015) analyzed the reasoning types required by secondary mathematics textbooks and compared them to the types of reasoning actually used. The study reported that most tasks (80%) could be solved by imitating the procedures given in the book. Studies such as this are rare for college-level textbooks especially ones that focus on statistics concepts.

Frameworks

Lithner (2008) defined the term reasoning as "the line of thought adopted to produce assertions and reach conclusions in task solving" (p. 257). Argumentation is the part of reasoning that one uses to ground the truthfulness or otherwise of assertions. Argumentation could be based on intrinsic properties (relevant and correctly applied mathematical features in a given situation) or non-intrinsic ones. If one is trying to tell which fraction between 7/8 and 8/9 is bigger, for example, an intrinsic property would be the quotient while a non-intrinsic one would be a focus on the magnitudes of the numbers themselves (e.g., 9 is bigger than 8 so 8/9 is bigger). Lithner (2008) categorized mathematical reasoning into two broad categories namely imitative reasoning (IR) and creatively founded mathematical reasoning (CMR). IR is analogous to rote learning and is further categorized as memorized imitative reasoning (MIR) or algorithmic imitative reasoning (AIR). MIR involves complete recall of a solution to a problem and is common especially in tasks requiring proofs of mathematical claims and theorems. AIR, on the other hand, involves following certain solution procedures/algorithms to a certain class of problems without a deeper understanding of the process.

Reasoning is classified as CMR, if it meets the following conditions:

1. Novelty. A new (to the reasoner) reasoning sequence is created or a forgotten one is recreated.
2. Plausibility. There are arguments supporting the strategy choice and/or strategy implementation motivating why the conclusions are true or plausible.

3. Mathematical foundation. The arguments are anchored in intrinsic mathematical properties of the components involved in the reasoning. (Lithner, 2008, p. 266)

CMR is further classified as local or global. Local CMR can be thought of as an improved form of AIR. Local CMR happens when someone makes non-trivial modifications to parts of an already known solution strategy or algorithm in solving a new task. Global CMR happens when a solver has no known solution strategy or algorithm and thus engages in CMR throughout the task. Using these classifications of reasoning, I define tasks accordingly. A task that requires imitative reasoning is called an IR task while a task that requires CMR is labelled a CMR task. CMR tasks are further classified as global or local while IR tasks can be algorithmic (AIR) or memorized (MIR).

In making the above classification of tasks, I recognize the fact that different solvers may experience the same task differently depending on several extraneous factors. A task may be a local CMR task for one person while being a global CMR task for another. Since the focus of this study was to examine the tasks themselves, the study only considered the guidance provided in the narrative section of the book and the extent to which such guidance would likely elicit different kinds of reasoning among readers of the book.

Method

Data and Setting

The text analyzed in this study is the fourth edition of *OpenIntro Statistics* by David Diez, Mine Centikaya-Rundel and Christopher Barr. I chose this book because it is one of the most popular open-source books being used as the main text or as an additional reference "at community colleges to the Ivy League" (OpenIntro, n.d). The book is divided into nine chapters each with several sections. Each section has one exercise set with several exercise tasks for readers to complete. I randomly sampled four tasks from selected sections within chapters five, six, seven, and eight. I chose these chapters because their primary focus was on statistical inference, which according to previous studies, is one of the areas that many students struggle in (Makar & Rubin, 2018). Chapter five had three sections with a total of 26 tasks while chapter six had five sections with a total of 38 tasks. Chapter seven on the other hand had five sections and a total of 46 tasks. For chapter eight, only one section focused on inference (i.e., inference for linear regression) and had six tasks. Previous studies (e.g., Author, 2020) have shown that for many textbooks, more cognitively demanding tasks tend to appear towards the end of the exercises. Since I were not sure whether this is the case for the book that I analyzed, I wanted to ensure our sample had at least one task from all parts of the exercises (i.e., initial, middle and last). To achieve this, I used stratified sampling where I split the questions into the first 25%, middle 50% and upper 25%. I then used simple random sampling to select one task from the first stratum, two from the second, and one from the last. Some selected tasks had several parts (like a, b, c etc.) that were related and built on one another. Whenever such a task was selected in the random sampling, I purposefully chose the last one in the series since it was likely to incorporate all the others in some way. In the end, I had 54 tasks in the sample, representing 47% of all tasks within the target sections of the book.

Within the narrative of the text, I focused mainly on the solved examples, guided practice items, and the solved case studies. I also brushed through the other parts of the narrative section (e.g., introduction and motivation) for descriptions of procedures and methods of solving certain kinds of tasks. This information was useful in characterizing the reasoning forms that the tasks in the exercise section of the book might elicit.

Analytic Procedures

To analyze the data described above, I used a modified version of Lithner's framework as provided in Mac an Bhaird et al. (2017). As described earlier, argumentation is an important component in any attempts to characterize reasoning. However, since the aim in this study was to characterize exercise tasks according to the kinds of reasoning they are likely to elicit from their readers and not the reasoning of specific students, I thought that examining the model solutions alongside the argumentations given in the book would be more meaningful. This is because the solutions in the book provide a model that readers may use with varying degrees of modification when solving the exercise tasks. While I recognize that a reader can use knowledge gained elsewhere to solve problems in the book, I believe that for most readers, the methods described in the book would play a significant role in their task solving process. Indeed, Mac an Bhaird et al.'s (2017) noted that the solved examples provided in math textbooks and lectures play a key role in students' task solving processes. I then solved the sampled tasks from the exercise sections by following the methods given in the book whenever possible while keeping track of the number and nature of any required modifications on the book methods. I defined modifications as any alterations on the book methods without which a successful solution would be impossible. Trivial alterations such as a different problem context and numerical values were not considered modifications. Another expert also worked on the problems separately then we met to discuss. The interrater agreement on the modifications was 94%. I labelled tasks that required no modification of procedure in the book as IR tasks and those that required at least one modification or that did not have a procedure in the book as CMR tasks. I classified CMR tasks as local CMR if they required only one modification and global CMR tasks if they required at least two modifications or if there was no algorithm or procedure for solution in the book.

Results

To concretize the analysis described above, I start by using a task (see Figure 1) drawn from a section called "Confidence Intervals for a Proportion" under the chapter titled "Foundations for Inference".

> **5.14 Coupons driving visits.** A store randomly samples 603 shoppers over the course of a year and finds that 142 of them made their visit because of a coupon they'd received in the mail. Construct a 95% confidence interval for the fraction of all shoppers during the year whose visit was because of a coupon they'd received in the mail.

Figure 1. A task from Inference for Single Proportion

We were able to solve this task by closely following the solution model to one of the guided exercises (see Figure 2) in the book. Although the problem does not provide a percentage value for the success rate (in this case percentage of shoppers who made a visit because of a coupon), this is considered trivial given that most readers of the book would be college or high school students. The rest of the solution requires using different numbers and substituting them in the formulas appropriately. The critical value associated with the new confidence interval can be confusing to some readers but the process of finding it was provided elsewhere in the book and could be applied in this new situation.

GUIDED PRACTICE 5.15

In the Pew Research poll about solar energy, they also inquired about other forms of energy, and 84.8% of the 1000 respondents supported expanding the use of wind turbines.[9]

(a) Is it reasonable to model the proportion of US adults who support expanding wind turbines using a normal distribution?

(b) Create a 99% confidence interval for the level of American support for expanding the use of wind turbines for power generation.

Solution:

[9](a) The survey was a random sample and counts are both ≥ 10 ($1000 \times 0.848 = 848$ and $1000 \times 0.152 = 152$), so independence and the success-failure condition are satisfied, and $\hat{p} = 0.848$ can be modeled using a normal distribution.
(b) Guided Practice 5.15 confirmed that $\hat{p}$ closely follows a normal distribution, so we can use the C.I. formula:

$$\text{point estimate} \pm z^\star \times SE$$

In this case, the point estimate is $\hat{p} = 0.848$. For a 99% confidence interval, $z^\star = 2.58$. Computing the standard error: $SE_{\hat{p}} = \sqrt{\frac{0.848(1-0.848)}{1000}} = 0.0114$. Finally, we compute the interval as $0.848 \pm 2.58 \times 0.0114 \rightarrow (0.8186, 0.8774)$. It is also important to *always* provide an interpretation for the interval: we are 99% confident the proportion of American adults that support expanding the use of wind turbines in 2018 is between 81.9% and 87.7%.

[10] No, a confidence interval only provides a range of plausible values for a parameter, not future point estimates.

Figure 2. Sample guided practice exercise and solution

A reader solving the task in Figure 1 can provide a similar sequence of argumentation as given in the guided practice and other parts within the text. For this reason, I categorized the task as an AIR task. Next (Figure 3) I consider an example of a CMR task.

5.11 Waiting at an ER, Part I. A hospital administrator hoping to improve wait times decides to estimate the average emergency room waiting time at her hospital. She collects a simple random sample of 64 patients and determines the time (in minutes) between when they checked in to the ER until they were first seen by a doctor. A 95% confidence interval based on this sample is (128 minutes, 147 minutes), which is based on the normal model for the mean. Determine whether the following statements are true or false, and explain your reasoning.

(a) We are 95% confident that the average waiting time of these 64 emergency room patients is between 128 and 147 minutes.

(b) We are 95% confident that the average waiting time of all patients at this hospital's emergency room is between 128 and 147 minutes.

(c) 95% of random samples have a sample mean between 128 and 147 minutes.

(d) A 99% confidence interval would be narrower than the 95% confidence interval since we need to be more sure of our estimate.

(e) The margin of error is 9.5 and the sample mean is 137.5.

(f) In order to decrease the margin of error of a 95% confidence interval to half of what it is now, we would need to double the sample size. (Hint: the margin of error for a mean scales in the same way with sample size as the margin of error for a proportion.)

Figure 3. A *sample CMR task*

Although there were several solved examples and guided exercises in the narrative section of the text, none of them was presented in a way that can be directly used to answer the task in Figure 3 (part d). The examples provided step by step procedures for creating confidence intervals but did not provide ways of assessing the effect of sample size on the margin of error. As a result of this, I concluded that a solver would have to reuse the procedure for creating confidence intervals in a different way that goes beyond a mere plugging in of numbers. Specifically, a solver will need to think about the procedure in a reverse manner since the confidence interval is already provided. These two modifications meant that the task was further classified as a global CMR. In an analogous manner, I examined all sampled tasks within the chapters focusing on statistical inference. I summarized the results in Table 1.

Table 1. Classification of Task Types by Chapter and Distribution

Chapter	IR Tasks			CMR Tasks					
				Local CMR			Global CMR		
	First 25%	Mid. 50%	Upper 25%	First 25%	Mid. 50%	Upper 25%	First 25%	Mid. 50%	Upper 25%
Foundations of inference	5	3	1	0	1	1	0	0	1
Inference for Categorical data	4	6	2	0	1	1	0	1	1
Inference for numerical data	7	6	2	1	2	1	0	0	1
Linear Regression	2	2	1	0	0	1	0	0	0
	18	17	6	1	4	4	0	1	3

The general trend prevalent in Table 1 is that the text has more IR tasks than CMR tasks. Out of the 54 sampled tasks, 41 (76%) were found to be IR tasks. These tasks tended to appear more at or below the 50th percentile. Indeed, out of the 41 IR tasks, 35 (85%) appeared at or below the 50th percentile. This trend appears to be reversed for CMR tasks, especially for global CMR ones. Three out of 4 global CMR tasks were all found at or above the 75th percentile of tasks. Similarly, for local CMR tasks, 8 out of 9 tasks were at or above the 50th percentile.

Discussion

This study sought to explore the kinds of exercise tasks in a popular open-source statistics textbook based on the kinds of reasoning the tasks are likely to require for successful solution. As noted earlier, most studies on textbook analysis are based on calculus and geometry strands and do not use well specified definitions of reasoning. Thus, this study addresses this gap by using a well-specified framework for mathematical reasoning to characterize mathematical tasks based on the kinds of reasoning they are likely to elicit from readers. The study targeted one of the most challenging parts of statistics (inference) within a popular open-source textbook used in several colleges for introductory statistics courses.

In general, the study found that the processes described and/or implemented within the narrative sections of the book could be followed with little to no modifications for successful solution of most tasks within the exercise sections. This means that opportunities for engaging in deeper thinking were limited at least in the exercise sections. This finding should not be taken to mean that there are no opportunities for deeper thinking in the textbook. In fact, the narrative section did have examples that provide opportunities for deeper thinking but only of the reader attempts them before looking at the solution provided. Nevertheless, there were several tasks that provided opportunities for deeper thinking in the sense that solvers would have to engage in creative reasoning for successful solution. These tasks require major modifications to the procedures described in the book and a reader would need to think deeply about the concepts for successful solution. This finding is similar to those reported by Sidenvalli et al. (2015) which indicated that mathematics textbooks tended to have more imitative reasoning tasks than creative reasoning tasks.

In regard to the distribution of the tasks within the exercises section, I found a general trend in which imitative reasoning tasks appeared at or below the 25th percentile while creative reasoning tasks appeared at the tail end (above the 75th percentile). A possible explanation for this trend could be that the authors intended readers to practice with basic concepts and factual tasks before they could engage with more cognitively engaging ones.

One possible implication of this study is for instructors using the text as a resource in their classrooms. When assigning exercises, it would be helpful to have students complete at least one task from the initial, middle and the last parts of the exercise sections. Depending on the goal of assigning the exercises, it might also be helpful for instructors to modify the tasks in some way before assigning them to the students.

One limitation of this study is the lack of student voices in characterizing the reasoning likely to be elicited by the tasks. Nevertheless, the fact that the study examines the tasks in relation to the narrative section still provides useful information because the arguments provided in the narrative section are likely to be repeated by students solving the tasks.

References

Dolev, S., & Even, R. (2015). Justifications and explanations in Israeli 7th grade math textbooks. *International Journal of Science and Mathematics Education*, *13*(2), 309-327.

Geteregechi, J.M., & Waswa A. (2020). Opportunities for reasoning-and-proving in a University precalculus textbook. *Conference on Research in Undergraduate Mathematics Education, the* 23rd *meeting of the Special Interest Group of the Mathematical Association of America (SIGMAA) on Research in Under- graduate Mathematics Education*. Boston, Massachusetts, USA.

Lithner, J. (2008). A research framework for creative and imitative reasoning. Educational Studies in Mathematics, 67(3), 255-276.

Mac an Bhaird, C., Nolan, B. C., O'Shea, A., & Pfeiffer, K. (2017). A study of creative reasoning opportunities in assessments in undergraduate calculus courses. *Research in Mathematics Education*, *19*(2), 147-162.

Makar, K., & Rubin, A. (2018). Learning about statistical inference. *International handbook of research in statistics education*, 261-294.

Otten, S., Gilbertson, N. J., Males, L. M., & Clark, D. L. (2014). The mathematical nature of reasoning-and-proving opportunities in geometry textbooks. *Mathematical Thinking and Learning*, *16*(1), 51-79.

OpenIntro. (n.d.). OpenIntro Statistics. https://www.openintro.org/book/os/

Sidenvall, J., Lithner, J., & Jäder, J. (2015). Students' reasoning in mathematics textbook task-solving. *International journal of mathematical education in science and technology*, *46*(4), 533-552.

Stylianides, G. J. (2009). Reasoning-and-proving in school mathematics textbooks. *Mathematical thinking and learning*, *11*(4), 258-288.

Tarr, J. E., Reys, R. E., Reys, B. J., Chávez, Ó., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for research in mathematics education*, *39*(3), 247-280.

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.
Van Steenbrugge, H., Valcke, M., & Desoete, A. (2013). Teachers' views of mathematics textbook series in Flanders: Does it (not) matter which mathematics textbook series schools choose?. *Journal of Curriculum Studies*, *45*(3), 322-353.