We investigate how mathematical content in a textbook is taken up by instructors, as they plan and teach their lessons, and by students, who participate in those lessons. We first identify mathematical competencies afforded by the textbook applying manual and automatic coding of the textbook’s raw content that uses natural language processing techniques. Next, we analyse textbook usage reported by users collected via periodic surveys and automatic real-time textbook viewing data with the goal of determining whether textbook features lead to the activation of competencies by the students.

INTRODUCTION

Within the array of resources for teaching and learning, the textbook continues to be the most prevalent one for instructors and students. With new technological developments, textbook formats have been changing from paper to digital and open source formats, including sophisticated tools such as computing cells, annotation tools, and powerful search engines, easing mass access at relatively low cost. Importantly, open source textbooks never expire or go out-of-print and can be distributed at no cost to students, making them practically fully accessible. In countries in which post-secondary education costs are high wide accessibility contributes to eliminating financial barriers to education. However, the full potential of these textbooks can only be understood through empirical studies of how students and instructors actually use these enhanced textbooks. The study we report here is part of the large project that seeks to describe how instructors and students use open-source, technologically enhanced university mathematics textbooks. We specifically want to see the extent to which the competencies espoused in the textbooks are talked about by the students using those textbooks which may speak indirectly about how instructors enforce those competencies.

THEORETICAL BACKGROUND

The didactical tetrahedron (Rezat & Strässer, 2012) helps in understanding the pivotal role of resource use in teaching (Figure 1). In the base of the tetrahedron are elements of the instructional triangle, a definition of instruction as the interaction among the instructor, the students, and the content. In this model, resources are an interdependent element that modifies such interaction. In addition, the model allows researchers to attend simultaneously to students’ use of the resources to learn mathematics (Students-Mathematics-Resources) and to the

---

interdependent way in which instructors and students interact with the resources. Although initial conceptualizations of use have been proposed about textbook readers, ours is a purposeful attempt to investigate the use of textbooks for teaching and learning in real time.

Niss and Højgaard (2011) defined mathematical competency as “a well-informed readiness to act appropriately in situations involving a certain type of mathematical challenge” (p. 49). Their framework can be thought of an intersection of various sets of competencies believed to be necessary for students who seek a STEM degree. For this study, we constructed a mathematical competence framework inspired by the Kompetencer Og Matematiklæring (KOM, Competencies and mathematics learning) model (Niss, 2003), by Maaß’s work (2006), and by the joint position on mathematical skills for STEM set by the National Council of Supervisors of Mathematics and National Council of Teachers of Mathematics (2018). We focus on the following eight competencies: (1) Mathematical Thinking, (2) Posing and Solving Mathematical Problems, (3) Mathematical Modelling, (4) Mathematical Reasoning, (5) Representing Mathematical Entities, (6) Handling Mathematical Symbols and Formalisms, (7) Communicating in, with, and about Mathematics, and (8) Using of Aids and Tools. Several elements are used to describe each competency as a whole. For example, the Mathematical Thinking and Acting competency includes: i) posing questions that have a mathematical characteristic; ii) understanding and handling the scope and limitations of a given concept; iii) extending the scope of a concept; and iv) distinguishing between different kinds of mathematical statements (conditioned assertions, ‘if-then’, assumptions, definitions, theorems, conjectures, and cases).

CONTEXT AND METHODS

The data we analyzed for this investigation have been drawn from a more extensive study that investigates students’ and instructors’ use of open-source dynamic textbooks in calculus, linear algebra, and abstract algebra courses. Linear Algebra (Beezer, 2019), is an open source, dynamic textbook written using PreTeXt, an authoring markup language designed specifically by our development team to produce interactive online textbooks. This textbook includes the following features: theorems, proofs, definitions, exercises (and solutions), examples, Sage cells (interactive computational cells), and Reading Questions. In particular, Reading Questions, are interactive areas that invite students to write answers directly into the textbook. As students respond to the questions, the instructor receives them, in real time. PreTeXt captures the structure of textbooks to ease conversion to multiple other formats (https://pretextbook.org/).

To analyze the textbook features, we applied manual and automatic coding, using 25% of the textbook as training data, with nine sections randomly chosen out of 39. The manual coding constituted the first phase of analysis and was based on the competency framework. We coded the textbook by identifying which competencies were being activated within each textbook feature. The competencies are not mutually exclusive;
concurrent activation is possible as are multiple activations of the same competency within the same textbook feature: e.g., the reasoning competency was activated in several parts of the same proof. However, we record whether activation of the competency occurs or not, and not the frequency or quality of its activation. To establish intercoder agreement, three individuals coded two sections in the Dimension' subsection of the Vector Spaces section. This subsection included 44 items and were coded by textbook features. The inter-rater agreement was 91% (agreement in 40 out of 44 items). Discrepancies resulted from similarity in definitions for the competencies; those were clarified.

Once coding reached a sufficient amount of textbook content, we used the data as the training dataset for the text classification models. The dataset needed to contain several hundred of labeled text documents to ensure a reasonable degree of accuracy for the model. The second phase of the analysis included statistical cross-validation methods to train the model iteratively using automatic coding using natural language processing techniques that derived a set of competencies for the rest of the textbook. Once the model was mature enough, we run the model for the rest of the textbook to tag feature section content with different mathematical competencies. The model itself continuously evolved and improved reaching higher accuracy.

To complement the study of the textbook, we analyzed survey responses filled out by over 100 students who were taking linear algebra courses with the HTML version of the textbook. The students were in six different courses taught by instructors located in different states; the data were collected during the spring of 2018. We used topic modeling to analyze these responses in a scalable and efficient manner. GENSIM library was selected to conduct Latent Dirichlet Allocation (LDA, Blei, Ng, & Jordan, 2003) topic modeling on each survey question. The LDA algorithm aims to map the documents to a predefined number of topics; each topic is mapped to a number to keywords that help describe it. Our script repeatedly applied the topic modeling algorithm based on a range of topic number we specified beforehand. We picked the model with the highest evaluation score as the final model. For each question, we generated three tables: the number of documents for each topic, representative documents for each topic, and the dominant topics for each document. The outputs were generally very interpretable, with meaningful differences between topics. One particularly useful feature of LDA topic modeling is that it identifies the most representative student responses for each question. With these results, we identified the sections of the textbook that students have used most heavily as well as the specific ways they interacted with the textbook. These results were further compared with the actual usage data generated from HTML textbook server log. Critical pieces of information to uncover from the server log include the frequencies of access by students for each section of the textbook during different timeframes, the number of students accessing each chapter, as well as the general usage trends exhibited by the

1 https://books.aimath.org/fcla/section-D.html
class throughout the semester. While, the amount of coded textbook samples and student responses was relatively small, the findings provide sufficient confirmation that the processes for analyzing the textbooks and the student responses can yield important patterns of use.

**FINDINGS**

We present first the findings regarding the competencies that appear to be afforded by the textbook features followed by the textbook features that seem to be most used by the students, according to the LDA topic modelling analysis. The frequency of coded features by competencies are presented in Table 1.

<table>
<thead>
<tr>
<th>Features (ft)</th>
<th>Mathematical Thinking</th>
<th>Posing &amp; Solving Math Problems</th>
<th>Mathematical Modelling</th>
<th>Mathematical Reasoning</th>
<th>Representing Math Entities</th>
<th>Handling Symbols &amp; Formalism</th>
<th>Communicating in, with, about Math</th>
<th>Using Aids and Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition (n=34)</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>28</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Example (n=65)</td>
<td>58</td>
<td>3</td>
<td>1</td>
<td>25</td>
<td>50</td>
<td>48</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Sage Cell (n=25)</td>
<td>3</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>Theorem (n=78)</td>
<td>77</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>70</td>
<td>38</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Reading Questions (n=30)</td>
<td>21</td>
<td>22</td>
<td>0</td>
<td>14</td>
<td>11</td>
<td>15</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Exercise (n=147)</td>
<td>72</td>
<td>123</td>
<td>2</td>
<td>42</td>
<td>87</td>
<td>107</td>
<td>72</td>
<td>0</td>
</tr>
<tr>
<td>Proof (n=77)</td>
<td>68</td>
<td>0</td>
<td>0</td>
<td>66</td>
<td>68</td>
<td>14</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>325</td>
<td>170</td>
<td>3</td>
<td>148</td>
<td>316</td>
<td>254</td>
<td>188</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 1. Frequencies of competencies coded by feature in the textbook.

There are primarily two ways to interpret Table 1. First (feature-oriented), we focus on a competency and check how it is enacted across different textbook features. By calculating the proportion for each feature (the active level) given the fixed competency\(^2\). Active level can be seen as a compromise between the proportion of any particular textbook feature within our sample and the proportion of the fixed competency within the given textbook feature. These two proportions determine active level. If a feature is not highly present in the textbook but a particular competency is highly reflected in such feature, such combination will not likely be deemed as prominent. Second (competency-oriented), we focus on one textbook feature and check how different competencies are reflected within that feature. For example, there are 34 definitions in total and 32 of them reflect mathematical thinking, the percentage we assign is 94%. Any percentage above 50% is deemed “very high,” between 12.5% and 50% is deemed “high,” and under 12.5% is deemed low.

**Competencies afforded in textbook features**

From a feature-oriented perspective, we found that the feature *Exercises* addressed a wide variety of mathematical competencies: seven out of eight have their highest active level in the exercises. *Examples* closely followed and achieved prominence (see Table 1) in six out of eight competencies. Overall, *Exercises* and *Examples*, addressed

\(^2\) We deem a feature prominent if the proportion is above 15%: 100 percent out of 7 features ~15%.
a high variety of mathematical competencies. In addition, five out of eight competencies were found in at least three to four textbook features, which suggests that the textbook balances and presents different competencies in a variety of features. We found that the modelling competency rarely appeared in the textbook. Two features, Definitions and Sage Cells, seem to address a narrower spectrum of competencies; the former, achieved virtually no prominence across all eight competencies; the latter focuses exclusively on the eighth competency (Making Use of Aids and Tools). These frequencies might not be surprising, as we could expect that different features be relevant for different competencies. As we document how students interact with particular textbook features, we will be able to anticipate which competencies are they more likely to activate. From a competency-oriented perspective, Mathematical Thinking, Representing Mathematical Entities, and Handling Mathematical Symbols and Formalisms, are prevalent across many textbook features, each dominating in at least four to five out of seven features. Mathematical Reasoning and Communicating in, with, and about Mathematics are moderately prevalent, each achieved high active level (between 13% and 50%) in four to five out of seven textbook features. We also found that the competencies Mathematical Modeling, Making Use of Aids and Tools, and Posing and Solving Mathematical Problems competencies are relatively rare to find across textbook features, with less than three textbook features emphasizing the development of these competencies.

**Textbook features most used by students**

One survey question asked the students to describe what they did while using the textbook. Among the representative responses using Examples was frequent. Students seem to use the Examples as a starting point while studying a section: “*While reading the textbook I will mostly ignore large blocks of words or explanations and move straight to the examples.*” Examples were also the main source for creating their personal documents for the course: “*I will do the same example multiple times until it’s very known.*” The textbook’s homework exercises, including the embedded solutions, were also identified as a main textbook feature, used while preparing for a particular section. The LDA topic modeling method revealed multiples references to the use of exercises and solutions by the students: “*The solutions to the exercises at the end of the sections were the most useful for me.*” Therefore, based on the findings of the previous section, we can assume that students will activate a wide variety of mathematical competencies even though they seem to engage mostly with only two textbook features.

**DISCUSSION**

We sought to explore what competencies are afforded by the textbook features and to what degree the students interacted with the textbook features. The results suggest an association between the mathematical content offered by the textbook—in the form of mathematical competencies—and the competencies that the students will likely be asked to bring into action and possibly improve on while interacting with particular
textbook features. If students are interacting primarily with the *Examples* and *Exercises*, then they will likely have to activate and advance various competencies: mathematical thinking skills, use of mathematical entities and symbols, and posing and solving mathematical problems. We also found competencies that were less prominent and some that were not very common in the textbook: this textbook does not offer many opportunities for students to advance their mathematical modelling skills (e.g., features that would ask students to interpret mathematical results in an extra-mathematical context). Likewise, *Making Use of Aids and Tools* was rarely found in the textbook (only afforded by one textbook feature, Sage). Finally, the *Posing and Solving Mathematical Problems* competency was highly prominent within *Exercises*, but it was relatively rare in the features of the textbook.

The LDA topic modelling we used to analyze students’ responses in our survey was facilitated by incorporating a summary of features the students emphasize most. The LDA method identified the most representative responses. However, to interpret, derive conclusions, and create the summary, we needed to read and understand what these selected responses were highlighting regarding the use of the textbook features. A future step towards the analysis of the textbooks is to expand the competency framework to include finer categories, in order to obtain more nuanced patterns.

**References**


