Developing and testing a framework for analysing, comparing and visualizing content matter in and across mathematics textbooks

Jonas Bergman Ärlebäck & Peter Frejd
Department of Mathematics, Linköping University

In this methodological paper we discuss the development and testing of a framework for analysing and visualizing the overall structure of the content and tasks in mathematics textbooks. The aim of the framework is to provide a lens on mathematics textbooks at the secondary level that is mathematics topic- and content-area-unbiased to facilitate an as objective as possible comparisons between both (a) different contents, topics, and tasks within a textbook; and (b) contents, topics, and tasks across different series of text-books or within a textbook series. The paper explores the results of the framework applied to four mathematical textbook chapters visualized in terms of so-called hierarchy charts. Inter-rater coding reliability with respect to a developed coding manual, and future development and applications of the framework is also discussed.

Introduction and background

Textbooks are important and fill multiple purposes in the teaching and learning of mathematics (Fan et al., 2013; Schubring & Fan, 2018). Indeed, the influences of textbooks on the mathematics classroom are well documented both internationally and in the Swedish context (Valverde et al., 2002; Bergqvist et al., 2010a,b). In particular, much of the mathematics the students meet and work on in school is mediated through the mathematical tasks in terms of problems, exercises, and activities in the textbooks (Henningsen & Stein, 1997). Hence, textbooks, their content and use, have been an area of research in mathematics education going back at least to 300 years (Fan et la., 2013).

Fan et al. (2013) made a systematic literature review of 100 journal papers that all focused on analysing textbooks, and identified three main research foci roughly equally distributed: (1) *the role of textbooks* (non-empirical and more philosophically oriented papers about the role of the textbook in mathematics teaching and learning); (2) *textbook analysis and comparisons* (studies of textbook contents and topics as study-object in themselves as well as within and across different textbooks); and (3) *textbook use* (research on teachers' and/or students' use of textbooks). Fan et al. (2013) also used a fourth category, *other*, in their categorisation. Within the second research foci, textbook analysis and comparisons, which is the focus in this paper, Fan et al. (2013) further identified five sub-foci, namely research focusing on and analysing: (i) *mathematics contents and topics*; (ii) *cognition and pedagogy*; (iii) *gender, ethnicity, equity, culture* *and values*; (iv) *comparisons of different textbooks*; and (v) *conceptualization and methodological matters*.

From a Swedish perspective, and focusing on the secondary level, most research carried out to date is in line with foci (2) or (3). An example of research on textbook use (3) is Johansson (2006), who illustrated and discussed how three mathematics teachers use the textbook as an instrument in different ways to organize their mathematics lessons. However, research focusing on analysing and comparing various aspects of the actual textbook appears to be more common. There are multiple examples of Swedish studies that involve two or more of the sub-foci identified by Fan et al. (2013). Brändström (2005) for example investigated the cognitive demand of tasks in relation to differentiated tasks (ii) in the topic area of fractions (i) and compared textbooks from three different textbook series (iv). On the other hand, Jakobsson-Åhl (2006, 2008) studied how algebra (i) and algebra-oriented word problems (i) had been treated and changed in textbooks over a 40-year period (iv). As a last example, we have Brehmer et al. (2016) who looked at and characterized various aspects of problem-solving tasks (i) in three upper secondary mathematics textbook series (iv): type of reasoning required (ii), placement (i), level of difficulty (ii) and task context (i).

Form the literature on research on textbook analysis one can note two things. First, regarding the analytical tools and developed framework in general, and in particular the one used in the three latter examples of Swedish research, many tools contain at least one analytical dimension which require some type of value judgement (level of cognitive demand; phenomenographic- and hermeneutic interpretations, and; reasoning type and difficult level respectively). Secondly, the results of the textbook analyses are often presented in tabulated forms listing proportions or/and percentages, cross-tables and statistical comparisons of how the textbook were coded. Not seldom is it hard to get an overview of the coding and results.

In this paper, we want to develop a flexible framework that provides a general structural overview of the textbook and a rough categorisation of the tasks in the textbook that will facilitate

a) comparisons between different contents, topics and tasks *within* a textbook;
b) comparisons between contents, topics and tasks *across* different series of textbooks or within a textbook series;
c) comparisons in such a way that only more or less objective aspects of the textbooks (i.e. not including categories or classifications open to subjective interpretation such as cognitive demand, level of complexity etc.); and
d) a direct, intuitive and accessible way to provide an overview and visualization of the result of applying the framework, that point out major trends and patterns in the data as well as suggest what aspects of the analysis that might benefit from a closer scrutiny.

Motivation, aim, and rationale for the developed framework

As mentioned above, the aim of the framework presented in this paper is to provide a general description and overview of the structure of mathematics textbooks and the tasks found in them. In addition, we want this general description and overview to be possible to visualize in an intuitive and direct way. Our long-term motivation for developing the framework is to investigate to what extent different contents and content-specific tasks are presented and treated in similar ways. We are especially interested in characterizing how the content matters of *statistics* and *probability* are presented and treated compared to other mathematical contents at the secondary level. Hence, we are in particular interested in investigating the textbooks used for both the lower- (grades 7-9) and upper level (grades 10-12: "gymnasium"). However, in this paper we only discuss the framework itself and not the comparative study in which the framework is being used. In other word, using Fan et al. (2013) categorisation of textbook research foci, this paper qualifies in the category (v) *conceptualization and methodological matters*.

The methodological goal for the framework is three-fold: (1) to code all *content of the pages* of the textbook (i.e. all *area of the pages* should be coded); (2) to identify and code all *tasks* in the book; and (3) to explore so-called *Hierarchy Charts* to see if this type of representation and visualization of the coding provides a good overview and entry point to the analysis based on the framework. Hence, the framework has two types of codes with two different foci: One set of *structural codes* that categorise what types of content the pages of the textbook contain; and one set of *task codes* that categorise the tasks in the textbook in three rudimentary types of tasks depending on how they are presented in the textbook. The codes emerged collaboratively in an iterative processes inspired by previous frameworks used in research on textbooks (e.g. Brändström, 2006; Fan et al., 2013; Glasnovic Gracin, 2018;Valverde et al., 2002) and by going through multiple cycles of testing, evaluating and revising the codes as the framework was applied to empirical material in terms of mathematical textbooks used in secondary school.

Codes to characterise the overall structure of the textbooks

The codes that characterise the overall structure of the textbooks are intended to cover the total area of each page and can broadly be divided into two main types. The first type of code aims at capturing the overall structure of the textbook in terms of *passive content*. With passive content we here refer to content that not per se ask the students to actively perform any calculations or other mathematical work. There are nine such passive codes in the framework, such as for example *Central content/goal* (often a bullet list entailing the official leaning goals as specified in the national curricula documents) and *Instructional narrative* (a written presentation of the concepts, methods, and theory treaded in the chapter – on the use of the notion of *chapter* in this paper, se [1]).

The second type of code focuses on *active content* in the sense that they capture instances in the textbook where students are encouraged to actively engage in mathematical work. The framework has eight such active codes, such as for example *Exercises* (often coherent sections of the textbook with tasks where the students are offered opportunities to practice on the content and topic in the chapter) and *Activities* (suggested activities, either to be done in groups or individually, often involving some investigating, experimenting or other practical activity, aimed at motivating and getting the students interested in the content and topic at hand). For the definition of *task* adapted in the framework see below. Table 1 lists and briefly explains all the codes.

Codes to characterise the tasks in the textbooks

For the framework to have a mathematics topic- and content-area-influenced-free conceptualization of the notion of mathematical tasks, we define a *task* broadly to be a clearly designated self-contained set of instructions that asks the students to engage in some mathematical work and/or activity. This definition stands in contrast to many other conceptualisation of mathematical tasks used in research, such as the one used by Stein and Smith (1998) who define that a task is "…a segment of classroom activity that is devoted to the development of a particular mathematical idea. A task can involve several related problems or extended work, up to an entire class period, on a single complex problem" (p.269). The framework differentiates between three types of mathematical tasks (collectively abbreviated to *XIE*):

- **(X) eXercises:** Often briefly formulated tasks that normally only requires one or a few calculations and/or manipulations based on a formula or given method (c.f. Glasnovic Gracin (2018)).
- **(I) Intra-mathematical tasks:** Word problems that are completely formulated in a strictly mathematical context with no coupling to contexts or situations outside mathematics and the real word (c.f. Niss & Blum, 2020).
- **(E) Extra-mathematical tasks:** Word problems that are formulated in an extra-mathematical context that is more or less authentic, connected to the real world or set in an imaginary scenario (c.f. Niss & Blum, 2020).

The XIE codes work as sub-codes in the framework, and all tasks in seven of the eight active structural codes were assigned to exactly one of the above XIE codes. The XIE-codes were not applied to the sections of the textbook coded as *Activities*. Note that the *Diagnostic*, *Filling* and *Summary*, have additional (not XIE) sub-codes (see Table 1).

Although the framework tries to minimize the use of strongly subjective codes to increase inter-rater (inter-coding) reliability, some connections and parallels are possible to discerned with previous used research frameworks. For example, typically tasks that Jakobsson-Åhl (2006, 2008) considered to be word-problems are exclusively found coded as (I) or (E), as is the problem-solving tasks as described by Brehmer et al. (2016).

Applying and piloting the framework

To apply and pilot the framework we used the computer software NVivo12 (QSR, 2018) to code digitalized versions (high-resolution colour pdfs) of four mathematics textbook chapters: two from the Swedish lower secondary mathematics textbook *Prio Matematik 7* (Cederqvist et al., 2012); and one each from the Swedish upper secondary mathematics

Category *(Sub-categories)*	**Brief Explanation/example**
Central content/goal	(Bullet) list specifying the curricula content covered in the chapter and/or the students' learning goals.
List of content	A list presenting the mathematical content in the chapter.
Introductory text	Brief introductory and motivating text introducing a chapter – not an **Instructional narrative** (see below)
Activities	Suggested (group) activity (investigation, experiment or other practical) aimed at motivating and inspire the students.
List of concepts	List of concepts, potentially entailing brief explanations.
Introductory tasks *(XIE)*	Introductory and warm-up tasks ("starters").
History (H) & Social (S) *(XIE)*	A section presenting mathematics in a: (H) historical context, its development or historical significance; (S) social context and stressing the role and function of mathematics in society.
Instructional narrative	A written (introductory) text describing the content of the chapter discussing and explaining theory and concepts.
Solved tasks *(XIE)*	Fully solved tasks presented.
Exercises *(XIE)*	Tasks to practices the topics and contents of the chapter.
Extension-tasks *(XIE)*	Tasks that go beyond what could be considered to belong to the standard course covered in the chapter; extra tasks or material deepening or extending concepts , theory and/or methods.
Application of abilities *(XIE)*	Tasks explicitly stated to practice or test the mathematical abilities found in the mathematics curricula (such as communication, reasoning, and problem solving).
Summary *(List of concepts, Instructional narrative, Mindmap)*	Section summarising the chapter. *List of concepts:* list of concepts containing explanations and examples; *Instructional narrative:* a summary written in the style of an instructional narrative; *Mindmap:* a network representation summarising the important concepts, strategies and methods of the chapter.
Repetition tasks *(XIE)*	Tasks in the end of the chapter aimed at giving a repetition of the content covered in the chapter.
Diagnostics (*concept XIE, Proficiency XIE*)	Tasks that test the contents in focus of the chapter; either focusing on *concepts* or *proficiency*.
Filling (*Figures/decorative pattern, Header, Footer*)	*Figure/decorative pattern* such as a photo or confined area with decorative pattern (potential related to or illustrating a task, but not placed in an integrative way with respect to the task). *Header* and *Footer* are simply the pages' header and footer.
Other	Content in the chapter to which the above categories are not applicable, such as references to web-based material, pre-tests, lists of prerequisite knowledge for example.

Table 1. The codes and sub-codes of the developed framework with brief explanations.

textbooks *Matematik5000 1a* (Alfredsson et al., 2011a) and *Matematik5000 1c* (Alfredsson et al., 2011b). We iteratively developed a coding manual containing: (a) elaborated definitions of all the codes and illustrating examples taken from the iterative process that resulted in the framework; and (b) explicit instructions and guidelines about how to mark and select the sections in the textbooks that are to be assigned structural codes (using a region-/area-selection tool in NVivo) as well as how to mark and select the tasks to be assigned task codes (using a text-selection tool in NVivo). The first tentative coding manual was put together inspired from (a) non-value-laden codes from other frameworks (c.f. Brändström, 2006; Glasnovic Gracin, 2018; Jakobsson-Åhl, 2006; 2008); (b) literature on different types of mathematical tasks (c.f. Niss & Blum, 2020); and (c) by inspecting and discussing the content and tasks in assorted mathematics textbooks. The coding manual was then successively developed and refined iterative by four coders applying the manual to code randomly selected pages in the textbooks until the process resulted in a saturated framework (see Table 1).

Inter-rater reliability

To investigate the robustness of the framework two coders independently coded four whole mathematics textbooks chapters: from *Prio Matematik 7* the chapters on *Statistics* (30 pages) and *Fractions and percent* (48 pages), and from *Matematik5000 1a* the chapter on *Probability and Statistics* (48 pages) and from *Matematik5000 1c* the chapter *Graphs and functions* (44 pages). NVivo12 provides an automatic analytic comparison of two coders' coding with respect to both region-/area-coding using a pixel measure for the marked and selected section coded, as well as the coding of marked and selected text segments. In Nvivo, the agreement of the coders is measured using two methods: (a) *percentage agreement* showing the proportion of all coded units the coders agree on divided by the total number of unique units coded by the two coders; and (b) *Cohen's kappa coefficient*, κ, which provides a number between –1 and 1, where $\kappa \leq 0$ indicate no agreement and $\kappa=1$ indicates perfect agreement. Cohen's kappa is a statistical measure that factor in the chance of two coders assigning the same code to a unit by pure chance, and according to Fleiss et al. (2003), $0.75 \leq \kappa \leq 1$ is considered very good inter-rater reliability; $0.41 \leq \kappa \leq 0.75$ fair to good; and $\kappa \leq 0.4$ poor.

The analysis of the inter-rater reliability in the four coded chapters showed almost a 100% agreement for the coding of the tasks into the XIE-categories. Only in two cases did the coding differ, and in one of these cases one of the coders admitted to having accidentally picked the wrong code in the NVivo12 environment. In the other case one of the coders expressed that the coding manual allowed for some ambiguity with respect to the definition of what constituted an extra-mathematical tasks, rendering the divergent coding. The definition was discussed and the coding manual slightly revised, which resolved the ambiguity and divergent coding. With respect to the (sub-)codes XIE, we hence conclude the framework to reliably capture the three different (general) types of tasks: *exercises, intra-* and *extra-mathematical tasks*.

With respect to the overall structural codes the inter-rater reliability measures NVivo generates is not as straight forward to interpret as in the case with the XIE-codes.

This is due to the fact that part of the structural coding is to decide what area (in terms of a rectangle drawn in using the area selection-tool in NVivio) of the page it is that should be coded. In other words, there are two aspects to consider here: the (size of the) area that is coded and what code the designated area is assigned. Looking across the four chapters, the selected areas by the two coders overall agree to 96%. This relatively high level of agreement is partly due to that many of the areas of the pages in the textbooks that are coded in the framework are scaffolded by design features of the textbook. Some features, such as *heading*, *footer*, and (sets of) **Exercises** are clearly delimited in the textbooks with breaking lines, and the coloration in the books also provide structure that facilitate that the selection of the areas to code. The calculation Cohen's κ factors in and compares the size of the areas selected and what codes the two coders have assigned to these. Looking across the four chapter and all coded areas, κ on average is $\overline{\kappa} = .90$, with the lowest $\kappa_{min} = 0.497$, which indicates an overall very good inter-rater reliability (Fleiss et al., 2003). It is notable that both coders coded the exact same general areas (though slightly different in size) and exact the same tasks, indicating that the structure and procedure for the coding laid out and specified in the coding manual were adequate. Based on the above evaluation of the inter-rater reliability, we conclude that the framework works in a satisfactory and reliable way. Next, we explore the output of the framework applied to mathematics textbooks in terms of so-called *hierarchy charts*.

Results in terms of hierarchy charts

Figures 1 and 2 below show the *hierarchy charts* generated by NVivo for the coding of the two chapters from the lower secondary mathematics textbook *Prio Matematik 7*. The hierarchy charts depict the 17 structural codes in rectangles coloured in darker nuances, with the size of each rectangle of a given code (colour) proportional to the total area of the chapter assigned that code (see further remarks below). The lighter nuances coloured rectangles found within some of the darker coloured rectangles show the proportions of how the tasks, or other sub-codes, within the particular structural code were coded. For example, looking at the code **Diagnostics** (Diagnos) in the two *hierarchy charts*, we can see the that tasks are roughly proportionally distributed similarly with respect to the relative focus on *concepts* (begreppstest) and *proficiency* (färdighetstest).

With respect to **Exercises** (övningar), the Statistics chapter chart shows that there is an overwhelming majority of *extra-mathematical tasks* (utommatematiska övningar) compared to *intra-mathematical tasks* (inommatematiska övningar) and *exercises* (mängdträning). This is not surprising given that the content matter of statistics invites the use of real data, diagrams, and extra-mathematical situations to work on. Looking at **Exercises** in the chart for the Fraction and percent chapter however, shows a more balanced distribution of *extra-* and *intra mathematical tasks* and *exercises*. Here, *exercises* is the largest category of tasks, followed by first *extra-* and the then *intra-mathematical tasks*. This pattern of types of tasks found in **Exercises** in the Fraction and percent chapter is mirrored in for **Repetition tasks** for the same chapter, which is in stark contrast to the distribution of tasks in **Repetition tasks** for the Statistics chapter.

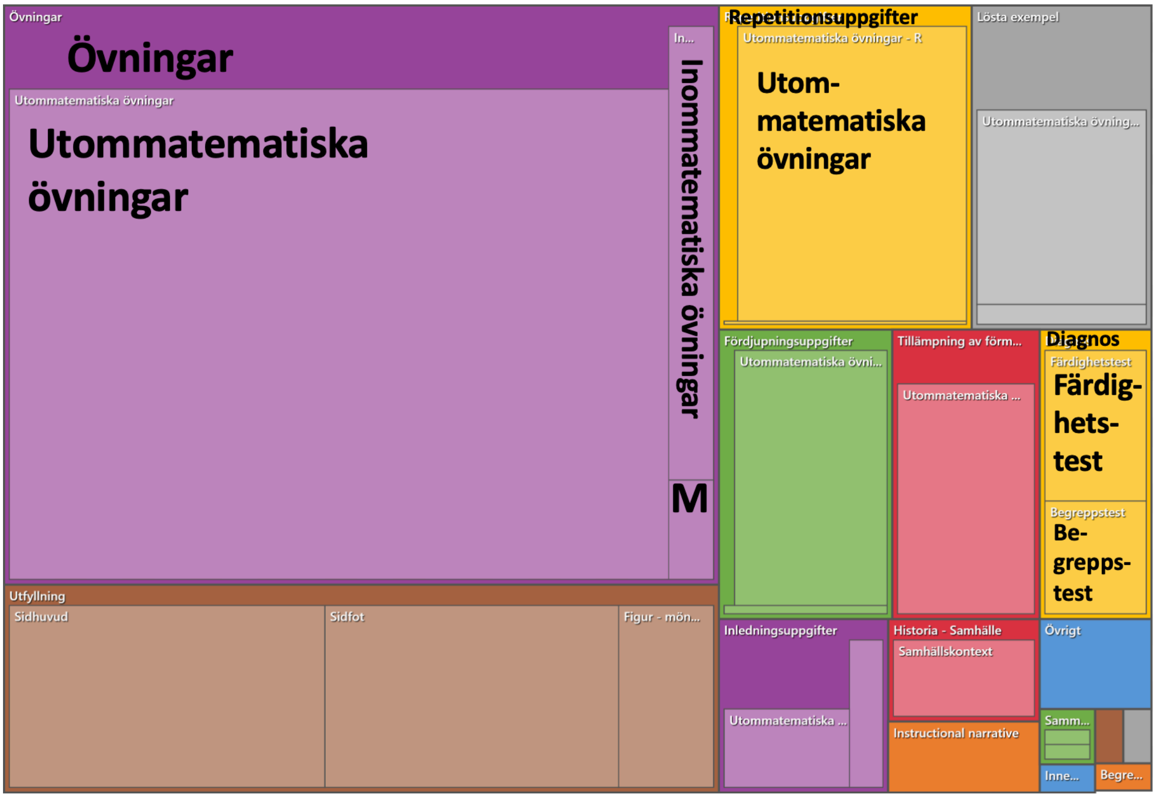

Figure 1. The hierarchy chart for the coding of the chapter *Statistics* (30 pages) in *Prio Matematik 7* (Cederqvist et al., 2012) [2]

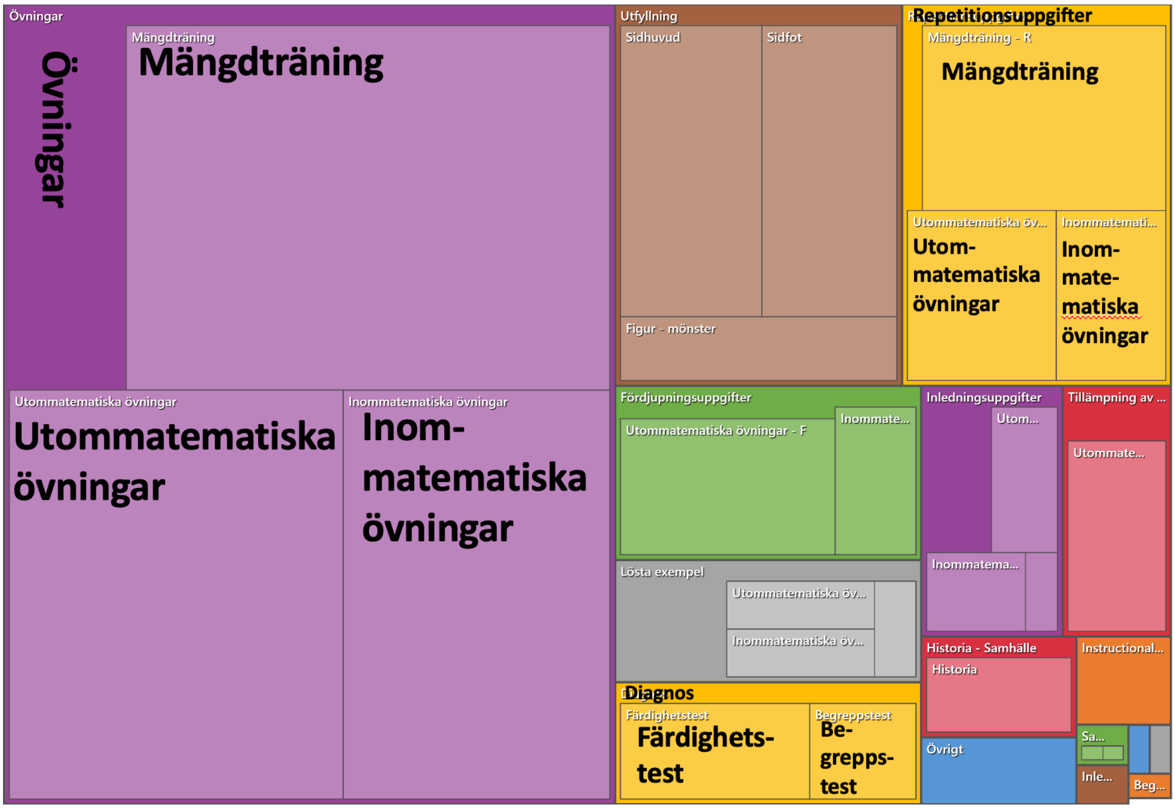

Figure 2. The hierarchy chart for the coding of the chapter *Fractions and percent* (48 pages) in *Prio Matematik 7* (Cederqvist et al., 2012) [2]

Remarks, conclusion, discussion, and future research

The framework we have presented in this paper have shown promising result with respect to its potential in providing a mathematics topic- and content-area-unbiased overview of the overall structure and tasks in mathematics textbooks. As such, the framework allows for a slightly new entry point when analysing mathematics textbooks in that it can function as the basis for making different types of comparisons across topics and contents within as well between textbooks. The analysis of our piloting of the framework showed very good inter-rater reliability, which is encouraging with respect to continue using and developing the framework.

The hierarchy charts proved to be useful to visualize the results of the coding and in identifying and recognizing both somewhat expected as well as more novel patterns, similarities, and differences in the chapters on *Statistics* and *Fractions and percent* in the Swedish lower secondary mathematics textbook *Prio Matematik 7*. Here, it should be noted that the hierarchy charts visualization produced in NVivo is dynamically scalable, meaning that printed in a different format, even the smallest regions will be more intelligible in terms of providing clearly readable codes for example. However, a much deeper and richer analysis would be possible if the hierarchy charts are complemented with the table explicitly showing the breakdown of the distribution of codes and sub-codes. This is one natural way to extend the analysis, in particular when it comes to gain more insight into the structural elements (codes) of the textbooks that are relatively sparely present in the analysed textbook (and hence are small and hard to make sense of in the hierarchy chart).

Another way to develop the framework in the context of comparing different textbooks and textbook series, is to bring in a topic-based and content-cantered framework to provide a more detailed content-specific analysis as a complement to the overall picture painted by the framework presented in this paper. With the teaching and learning of statistics as one of our more recent research interests, these two (add a break-down table of codes and a complementing topic-based and content-cantered analysis) will be our next endeavours – to further and complement the analysis presented here.

Notes

1. In defining the codes used in the framework presented in this paper, we use the notion of *chapter* to denote a selected continues text, such as a generic normal book chapter, section or sub-section of a textbook, that can be delimited by specifying a page-range (and if needed also on what line(s) on the first and/or last page the selection starts/ends).
2. The original labels provided by NVivo (in white font) here reinforced in black for clarity.

References

Alfredsson, L., Björk, L-E., Brolin, H., Ekstig, K., Erixon, P., & Heikne, H., & Ristamäki, A. (2011a). *Matematik 5000 1a*. Natur & Kultur.

Alfredsson, L., Björk, L-E., Brolin, H., Bråting, K., Erixon, P., & Heikne, H., & Ristamäki, A. (2011b). *Matematik 5000 1c*. Natur & Kultur.

Bergqvist, E., Bergqvist, T., Boesen, J., Helenius, O., Lithner, J., Palm, T., & Palmberg, B. (2010a). *Matematikutbildningens mål och undervisningens ändamålsenlighet. Grundskolan våren 2009*. Göteborgs universitet.

Bergqvist, E., Bergqvist, T., Boesen, J., Helenius, O., Lithner, J., Palm, T., & Palmberg, B. (2010b). *Matematikutbildningens mål och undervisningens ändamålsenlighet. Gymnasiet hösten 2009*. Göteborgs universitet.

Brehmer, D., Ryve, A., & Van Steenbrugge, H. (2016). Problem solving in Swedish mathematics textbooks for upper secondary school. *Scandinavian Journal of Educational Research, 60*(6), 577-593. https://doi.org/10.1080/00313831.2015.1066427

Brändström, A. (2005). *Differentiated tasks in mathematics textbooks – an analysis of the levels of difficulty* (Licentiate thesis). Luleå University of Technology.

Cederqvist, K., Gustafsson, P., Larsson, S., & Szabo, A. (2012). *Prio Matematik 7*, Grundbok. (1. uppl.) Sanoma utbildning.

Fan, L., Zhu, Y., & Miao, Z. (2013). Textbook research in mathematics education: Development status and directions. *ZDM Mathematics Education, 45*(5), 633–646. https://doi.org/10.1007/s11858-013-0539-x

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Wiley.

Glasnovic Gracin, D. (2018). Requirements in mathematics textbooks: a five-dimensional analysis of textbook exercises and examples. *International Journal of Mathematical Education in Science and Technology, 49(7)*, 1003-1024. https://doi.org/ 10.1080/0020739X.2018.1431849

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*(5), 524–549. https://doi.org/10.2307/749690

Jakobsson-Åhl, T. (2006). *Algebra in upper secondary mathematics* (Licentiate thesis). Luleå University of Technology.

Jakobsson-Åhl, T. (2008). Word problems in upper secondary algebra in Sweden over the years 1960–2000. *Nordic Studies in Mathematics Education*, *13*(1), 7–27.

Johansson, M. (2006). Textbooks as instruments. Three teachers' ways to organize their mathematics lessons. *Nordic Studies in Mathematics Education*, *11*(3), 5–30.

Niss, M., & Blum, W. (2020). *The learning and teaching of mathematical modelling.* Routledge.

Schubring, G., & Fan, L. (2018). Recent advances in mathematics textbook research and development: An overview. *ZDM Mathematics Education, 50*(3), 765–771. https://doi.org/10.1007/s11858-018-0979-4

Stein, M. K., & Smith, M. S. (1998). Mathematical task as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School, 3*(4), 268–275.

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Kluwer Academic Publishers.

QSR, International Pty Ltd. (2018) NVivo (Version 12) [Computer software], https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home