Twenty years of research on technology in mathematics education at CERME

a literature review based on a data science approach

Herfort, Jonas Dreyøe; Tamborg, Andreas Lindenskov; Meier, Florian; Allsopp, Benjamin Brink; Misfeldt, Morten

Published in:
Educational Studies in Mathematics

DOI (link to publication from Publisher):
10.1007/s10649-022-10202-z

Publication date:
2023

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Herfort, J. D., Tamborg, A. L., Meier, F., Allsopp, B. B., & Misfeldt, M. (2023). Twenty years of research on technology in mathematics education at CERME: a literature review based on a data science approach. *Educational Studies in Mathematics*, *112*(2), 309-336. https://doi.org/10.1007/s10649-022-10202-z

Educational Studies in Mathematics manuscript No.
(will be inserted by the editor)

20 years of research on technology in mathematics education at CERME

a literature review based on a data science approach

Jonas Dreyøe Herfort · Andreas Lindenskov Tamborg · Florian Meier · Benjamin Brink Allsopp · Morten Misfeldt

Received: date / Accepted: date

Abstract In this article, we use data science to explore distinct topics in all the 336 papers that have been presented in Thematic Working Groups related to technology in the first eleven Congresses of the European Society for Research in Mathematics Education (CERME). We also explore how these topics are connected and have evolved. We apply topic modeling, and find 25 distinct topics that are grouped in four clusters; digital tools, teachers and their resources, technology experimentation, and a diverse cluster with a strong focus on student activity. Furthermore, we employed a Mann-Kendall test to investigate the temporality of the topics. We observed a tendency that technologies that are studied at a given time are primarily technologies that are new at that point in time. We also found an increased focus on teacher knowledge.

Keywords Topic modeling · Digital technology · Literature review · Educational data science

1 Introduction

Mathematics education has incorporated the use of digital technology since the 1980s (Papert, 1982), adopting first programming, then dynamic geometry systems

J. D. Herfort
University of Copenhagen
Tel.: +45 21 42 40 71
E-mail: jonasd@ind.ku.dk

A. L. Tamborg
University of Copenhagen

F. Meier
Aalborg University

B. B. Allsopp
Aalborg University

M. Misfeldt
University of Copenhagen

(DGS) (Laborde & Laborde, 1995), and computer algebra systems (CAS) (Johnstone, 1990) in mainstream mathematics teaching. The body of research-based knowledge about digital technology in mathematics education has since grown steadily to an extent where it is now difficult to overview all its sub-areas, how these are related, and how they have changed over time. Addressing these concerns could help identify gaps and inform future research (Bray & Tangney, 2015) and policy decisions (Kissane et al., 2015). However, conducting traditional reviews with their manual coding and analysis leads to two problems: first, it is difficult to conduct grounded coding completely transparently; and second, the work involved in manual coding becomes insurmountable for large corpora. Therefore, this paper investigates these questions through data science methods by applying them to 336 papers published in the proceedings of the Congress of the European Society for Research in Mathematics Education (CERME) in relation to the working groups concerned with technology and asks the following questions:

What are the distinct topics in the literature from the CERME working groups concerned with technology, how and to what extent are they connected, and how have they evolved from CERME 1 to 11?

We begin by surveying existing reviews in mathematics education and applications of topic modeling in mathematics education research. We then discuss how topic modeling can be used for content analysis, describing the application of the used data science tools, before presenting our findings and discussing the paper's contributions to the existing research in this field.

2 Related work

In the following section, we consider existing studies in mathematics education research that reviewed technology use and studies that have applied topic modeling as a review method.

2.1 Reviews concerning the use of technology in mathematics education

Existing reviews of technology use in mathematics education can roughly be categorized into two groups: systematic literature reviews (Borba et al., 2017; Ronau et al., 2014); and overview studies (Hoyles & Noss, 2003; Williams & Goos, 2012). Systematic literature reviews often have limited scope and build on specified methodologies. Such studies' specific focuses might include emerging trends in technology utilization (Borba et al., 2017), or how the scope and quality of Ph.D. dissertations regarding technology use in mathematics education have developed (Ronau et al., 2014). These reviews often synthesize the results or approaches in existing studies to achieve various purposes (Bray & Tangney, 2015; Kissane et al., 2015; Lagrange et al., 2003; Puentedura, 2006).

Systematic reviews may generate in-depth insights regarding the status or development of a field. However, as these studies are often highly focused, they do not investigate relationships between sub-fields within a broader landscape and are therefore unlikely to identify opportunities bridging these sub-fields (Zbiek et al., 2007).

Some studies have broader and more comprehensive aims (Hoyles & Noss, 2003; D. L. Roberts et al., 2012; Zbiek et al., 2007), however, as the amount of literature on technology use in mathematics education has increased tremendously over the last few decades (Hoyles & Noss, 2003), it is no longer feasible to meet such aims through the use of systematic reviews alone. We term such broader and more comprehensive reviews as overview studies. Overview studies are often conducted by highly experienced researchers who categorize the current state of development of the field by applying their experience and understanding (Hoyles & Noss, 2003; Zbiek et al., 2007). Overview studies typically provide theoretical conceptualizations of a broad research field's history or current state, allowing opportunities for suggesting new research directions, identifying promising variables, and promoting boundary-crossing across subfields that otherwise would not be combined (Zbiek et al., 2007).

Trgalová et al. (2018) is an example of such an overview study. This paper explores the thematic working groups' (TWG) from CERME that focus on technology from CERME 1 to 10 within the historical context of events outside the community. The study is not a systematic review but is conducted by experienced CERME researchers who provide accounts of the themes and discussions related to technology that has been utilized during CERME's history. Since their aim partly coincides with our study, we will summarize its main findings, paying particular attention to CERME's historical evolution.

CERME 1 discussions concentrated on the use of digital tools in the classroom, with a particular focus on students, their views, and the interpretation of dynamic tools, as well as a consideration of how digital tools could support student learning (Trgalová et al., 2018). While CERME 1 was driven by a fascination with the possibilities inherent in new tools, CERME 2 and 3 focused more extensively on theoretical reflections, comparing different tools and considering how they relate to one another. Additionally, more attention was paid to the teacher in technology-mediated activities to evaluate the emergence of technological design and use (Trgalová et al., 2018). In contrast, the discussions at CERME 4 moved to consider specific technologies such as CAS, DGS, and applets, as well as associated issues regarding the characteristics of the tools needed in mathematics teaching (Trgalová et al., 2018). Moreover, discussions at CERME 4 recognized the need for specialized frameworks that could analytically capture relationships between tools and mathematical content; this led to the emergence of the instrumental approach (Trgalová et al., 2018). At CERME 5, discussions regarding the application of this then-new framework took place, and questions were asked about tool appropriation and the integration and institutionalization of the framework. At CERME 6, the technology group broadened its list of welcomed contributions, moving from an exclusive focus on digital tools to incorporate a consideration of more traditional ones such as textbooks (Trgalová et al., 2018). This shift reflected the need to consider all the resources available to teachers to better understand the relationship between old and new resources. Since CERME 7, new theories that emphasize teachers' technological knowledge and practice have been developed (Trgalová et al., 2018). At CERME 8, the discussion focused on integrating technology from teachers' perspectives, emphasizing the need for a stronger focus on longitudinal studies that could capture the opinions of so-called "real teachers" (Trgalová et al., 2018). There were also requests for a more intensive focus on emerging themes in technology literature to determine how the then-new technology (e.g., Web 2.0

and mobile technologies) could be redesigned and used for mathematics education. Before CERME 9 and 10, many digital textbooks and e-books had emerged, and these discussions explored free online courses and their impact on mathematics education (Trgalová et al., 2018).

While systematic reviews "zoom in" in order to describe detailed aspects of a branch of study, overview papers like the study Trgalová et al. (2018) "zoom out" to give conceptual and broad (but less detailed) accounts of a field of study. In Section 6, we will discuss how overview studies and systematic literature reviews, and the types of findings in such studies, compared to the method and findings presented in this paper.

2.2 The application of topic modeling within mathematics education research

Topic modeling is an automatic method of content analysis. It is an unsupervised machine learning technique using statistical models to identify topics within extensive collections of textual data (Boyd-Graber et al., 2017).

In recent years, generating automatic literature reviews of extensive collections of scientific articles has become a popular application of topic modeling (Asmussen & Møller, 2019; Griffiths & Steyvers, 2004; Paul & Girju, 2009). In this context, a topic model not only provides insights into a field's dominant research areas but also identifies those areas that might otherwise be overlooked. Thus, a topic model can be used to give valuable suggestions for future research directions. Applications of topic models for this purpose can be found in almost any discipline. One of the first applications of topic models to understand scientific publications was Griffiths and Steyvers (2004), the study used a corpus of abstracts from papers published in *Proceedings in the National Academy of Sciences* from 1991 to 2001 to identify topics in scientific research. Similar undertakings have been performed in the study Blei and Lafferty (2007), where topic modeling was applied to a corpus of papers published in *Science*, in Mimno (2012) historiography of classical journals was created, and Bittermann and Fischer (2018) tried to identify *hot topics* in psychology research.

Topic modeling is already used in mathematics education research; Inglis and Foster (2018) review five decades of research using a corpus of all the articles published in the journals *Educational Studies in Mathematics* and the *Journal for Research in Mathematics Education*. The study identifies 28 topics in mathematics education research (see Table 1), of which only one was technology-specific (namely *dynamic geometry and visualization*). The models used in the study include *technology* as a standalone topic while identifying a decline in interest over the years. This is somewhat surprising, considering the increasing role technology plays in education; however, the study argue that technology has become more routinely embedded in mathematics teaching and that many other forums have emerged where studies discussing technologies can be published.

Chen et al. (2020) have investigated 3,963 papers published in the journal *Computers & Education* and have built a structured topic model that considers how these have evolved. This model includes 24 different topics (see Table 1), none of which appear to be related to any specific mathematical technologies (e.g., CAS, DGS, or spreadsheets).

Table 1: Table of topics from Inglis and Foster (2018) and Chen et al. (2020)

Inglis & Foster (2018)	Chen et al. (2020)
Addition and subtraction	Context and collaborative learning
Analysis	E-learning and policy
Constructivism	Experiments and methodologies
Curriculum (especially reform)	Human-computer interaction
Didactical theories	Social networks and communities
Discussions, reflections, and essays	Program and curriculum
Dynamic geometry and visualization	Demographic issues
Equity	Blended learning
Euclidean geometry	Data mining
Experimental designs	Online/web-based learning
Formal analyses	Multimedia and data-driven studies
Gender	Technology acceptance model
History and obituaries	Massive open online courses
Mathematics education around the world	Virtual reality
Multilingual learners	Programming language
Novel assessment	Mobile learning and early childhood education
Observations of classroom discussion	Game-based learning
Problem solving	Science education
Proof and argumentation	Teacher training
Quantitative assessment of reasoning	Language learning
Rational numbers	Special education
School algebra	Assessment
Semiotics and embodied cognition	Hardware
Sociocultural theory	Conceptual mapping
Spatial reasoning	
Statistics and probability	
Teachers' knowledge and beliefs	
Teaching approaches	

Another example of topic modeling in mathematics education research is given by Marks et al. (2020). The study synthesize 813 papers from *the Proceedings of the British Society for Research into Learning Mathematics* from 2003-2018. In addition, Marks et al. (2020) conducted a qualitative, thematic analysis and compared the findings to find convergence between the results generated by the two methods, both in which mathematical content and teachers were the two strongest keywords.

As exemplified above, topic modeling approaches have previously been applied to literature reviews that focus on mathematics education (Foster & Inglis, 2018; Marks et al., 2020) and technology in education (Chen et al., 2020). However, no previous studies have applied topic modeling with the specific aim of reviewing research on technology use in mathematics education.

3 Method: Topic Modeling

The most fundamental and popular type of topic model is Latent Dirichlet Allocation (LDA) (Blei & Lafferty, 2007; Blei et al., 2003). An LDA topic model combines two assumptions: (1) that each document is a distribution of all topics; and (2) a topic is a distribution over all the words of the collection (Blei, 2012). In other words, each topic consists of a weighted list of the most probable terms, and each document consists of a weighted list of the most probable topics. The assumption is that a document is created in a generative process with a set of underlying topics driving the writing process. To arrive at the final topics, LDA initializes by randomly assigning a topic to each word in each document. After this random initialization, the algorithm iterates through all words in all documents unassigning the assigned topic and then re-assigning a topic based on all the other topic assignments—considering both the probability of each topic in the document and the popularity of the word in each topic. LDA iterates through this process until it reaches a stable state. LDA will always converge toward the same topics (Blei et al., 2003).

An apparent reason for LDA's popularity is that many ready-to-use implementations of LDA are available in the form of programming packages (Graham et al., 2012; M. E. Roberts et al., 2014) or even software with a standard graphical user interface (Ramage et al., 2009). However, applying topic modeling is not always a straightforward process, and multiple aspects need to be considered as they all can influence the outcome and, thus, the quality and interpretability of a model. Among these aspects are: (1) The corpus characteristics (2) The applied text preprocessing steps (3) The hyperparameter configuration and evaluation metrics. Unfortunately, although topic modeling is a frequently applied technique, a systematic empirical exploration of what effect variations of these aspects have on topic model quality is missing. In practice, topic modeling is heavily built on anecdotal evidence, respectively successful observations from other researchers' use of the method under similar conditions. In the following, we will map out three aspects that have undergone detailed investigations and discuss minimum requirements and gold standards for topic modeling of scientific articles.

(1) Corpus characteristics: Corpus characteristics include all aspects related to the textual material itself, i.e., the heterogeneity of the articles' content and genres, the average length of the documents, the overall number of documents, and the number of unique words. The length of documents is usually not considered a problem as topic modeling has been applied successfully to documents like Tweets (Lim & Buntine, 2014). Thus, in principle, topic modeling could also be applied to a corpus of article abstracts. Regarding corpus size, the idea of 'the bigger the corpus, the better the model' seems omnipresent among topic modeling users. However, not many articles provide empirical validation for this statement.

The question of minimum corpus size, i.e., how many articles one needs for topic model experiments to become meaningful, is also lacking empirical investigations. However, one can argue that this is not only a question of the number of documents but usually also depends on the length of the documents, and thus the total number and the unique number of tokens. Maier et al. (2020) performed a systematic investigation on how much document sampling and vocabulary pruning influence the results of topic models. They showed that after words that occur in more than 99% or less than 0.5% of documents, a vocabulary size of 6,600 words, or more, is enough for LDA to create meaningful models.

(2) Preprocessing steps: Text preprocessing, such as removing punctuation and lowercasing, is an essential step in unsupervised machine learning (Denny & Spirling, 2018), and recommendations on which preprocessing steps to apply and in which order are widely available (Maier et al., 2018). Most if not all of these steps

are also well-grounded in empirical investigations. Schofield and Mimno (2016) studied the influence of stemming (reducing words to their stems), and Schofield, Magnusson, et al. (2017) investigated the influence of stopword removal on topic model quality. Later Schofield, Thompson, et al. (2017) quantified the effect of duplicate documents in topic modeling. Summarizing all this work, one can see that: (1) Stemming is unnecessary and can even be harmful as words can mean something entirely different when reduced to their base form outside their word class. (2) Stopwords do not harm the creation of the model and can easily be removed afterward when interpreting the topics. (3) Removing excessive numbers of duplicate documents is necessary. However, a small amount of duplication is tolerable. Finally, an essential part of the discussion about preprocessing steps is systematically and transparently documenting which steps one applied as this helps ensures reproducible results (Denny & Spirling, 2018). In this work, we follow the preprocessing recommendations provided by Maier et al. (2018) outlined in section 4.2.

(3) Hyperparameter configuration and evaluation metrics: The hyperparameters are α and β and the number of topics k. α represents the document-topic density, which means that the higher the value of α is, the more topics will be identified and vice versa for a lower number. β represents topic-word density, where a higher value makes topics contain more words and vice versa. In recent years, research in topic modeling presented several different evaluation metrics, e.g. topic distinctiveness, perplexity (Blei et al., 2003; Wallach et al., 2009), topic coherence (Blair et al., 2019; Lau et al., 2014)), and topic distinctiveness (Vega-Carrasco et al., 2020). The most common measure for evaluating topic models is topic coherence. Topic coherence is a measure of topic quality introduced by Mimno et al. (2011) as a reaction to the observation that not all topics a topic model creates are equally interpretable. This means that most words in most topics relate to a single coherent concept or theme, making a topic meaningful to a domain expert. In short, the topic coherence value is the average of the distances between words in the topic in the model. Up to 10% of topics are usually *flawed* by containing seemingly non-meaningful combinations of terms. Following Mimno et al. (2011), topics can be flawed in multiple ways: 1) *Chained:* In a topic with this flaw, every word is connected to every other word by some pairwise word chain. However, not all pairs make sense. 2) *Intruded:* A seemingly well-interpretable topic contains words that are unrelated to most words in the topic and give the impression of being *intruders*. 3) *Random:* The topic is nonsensical as the words are unrelated and seem to be assigned together randomly. 4) *Unbalanced:* a topic with this flaw contains words that are thematically connected, however, the topic combines very general and very specific terms, making the topic seem unbalanced.

We used the MAchine Learning for LanguagE Toolkit (MALLET) as a framework to build our model (McCallum, 2002). MALLET is an ideal tool for us as it contains easily applicable optimization solutions. Maier et al. (2018) mention that the choice of parameters heavily influences the validity of a topic model. We strive for higher validity by using a standard algorithm. MALLET optimizes α and β, and we also optimize for the number of topics k by multiple iterations and look for high topic coherence scores to minimize topic flaws.

However, automatic evaluation cannot stand alone, and a qualitative inspection of the top candidates is needed to choose the best-suited model (Maier et al., 2018). To be able to choose between the models, it is important to try out several configurations to find the model with the fewest topic flaws and thus with the most coherent topics.

4 Our process

This section describes the steps we took to provision and preprocess the corpus. We also describe the criteria used when choosing our model and our strategy for exploring the model, how we analyzed the temporal evolution, and the connections between the topics studied.

4.1 The corpus

The corpus used for this paper consists of contributions published in proceedings from the TWG on technology at CERME 1 to 11. CERME is the largest mathematics education conference in Europe; it has two TWGs that focus specifically on technology in mathematics education. Crucially, these groups only include papers that mathematics education scholars have reviewed. Thus, papers from these TWGs create a very relevant, accessible, and easily distinguishable corpus of data. While CERME contributions do not represent the field of mathematics education as such, CERME forms a large and significant community in the field of mathematics education research. Understanding how technology is studied within this community over time thus represents a valuable contribution to the field that has already been studied using manual methods (see Section 2). Additionally, topic models in the two fields (mathematics education and technology in education) that form the foundation of this hybrid field have already been published. It is, therefore, possible to compare the model presented in this paper with these models and thereby identify any research gaps.

When preparing our corpus, we chose not to distinguish between papers, posters, and introductions. One of the strengths of topic modeling is finding similarities between documents. If any set of these documents turns out to be discursively different, this should become apparent in the identified topics.

Table 2 summarizes the distribution of papers by conference year and TWG.

4.2 Preprocessing

We followed the ordering of preprocessing corpus data proposed by Maier et al. (2018). First, we cleaned the corpus by lowercasing all words and removing punctuation, spaces, quotes, new-line characters, hyphens, and other disturbing elements in the corpus. This gave us a long, clean text file containing only words for each paper. Second, we identified word classes (i.e., part-of-speech tagging), preserved nouns, verbs, and adverbs, and discarded any stopwords. Next, we identified single-entity terms consisting of two (bi-grams) or three words (tri-grams) (e.g., "instrumental genesis" and "computer algebra system"). We converted words

Table 2: Distribution of the corpus in CERME

CERME year and TWG	Number of papers
CERME 1 GROUP 2	9
CERME 2 GROUP 2	12
CERME 3 GROUP 9	21
CERME 4 GROUP 9	19
CERME 5 GROUP 9	22
CERME 6 GROUP 7	40
CERME 7 GROUP 15	24
CERME 8 GROUP 15	29
CERME 9 GROUP 15	22
CERME 9 GROUP 16	25
CERME 10 GROUP 15	26
CERME 10 GROUP 16	28
CERME 11 GROUP 15	20
CERME 11 GROUP 16	39

into their base form within their given word class while preserving their tense (e.g., computers → computer; computes → compute). Finally, we discarded words occurring in more than 99% of all papers or fewer than 0.5% to avoid both those words that occur in the vast majority of the texts and those that only occur rarely or in very few texts.

4.3 Choosing the model

This left us having to determine the number of topics (k). We created several topic models for 58 different k's from 3 to 60 topics and let MALLET optimize with an interval of 10. We performed this for 15 different random initializations, resulting in 870 iterations of a topic model with different k's and random seeds. For each iteration, we calculated the coherence value. We picked the coherence value as our evaluation metric as this value correlates well with the interpretability of a model by human readers.

Our step-wise approach produced two models with 25 and 29 topics, respectively. The model of 25 topics was the single iteration with the highest coherence value, whereas 29 topics scored the highest on average across the 15 run-throughs. However, the quality of the information generated depends on how well human researchers can interpret this model. As such, interpretability is the assurer of a model's validity (Maier et al., 2018). When making the choices mentioned above, we analyzed each model and used our field knowledge to choose the one (in this case, the 25-topic model) that seemed most interpretable. At a concrete level, the authors jointly analyzed both models by following the steps of interpretation described in 4.6 and looking for the most coherent model with the fewest topic flaws. Choosing a model is about finding the model with the fewest topic flaws that also seems coherent and interpretable for experts in the field. Our analysis of both models concluded that the model with 25 topics was the best.

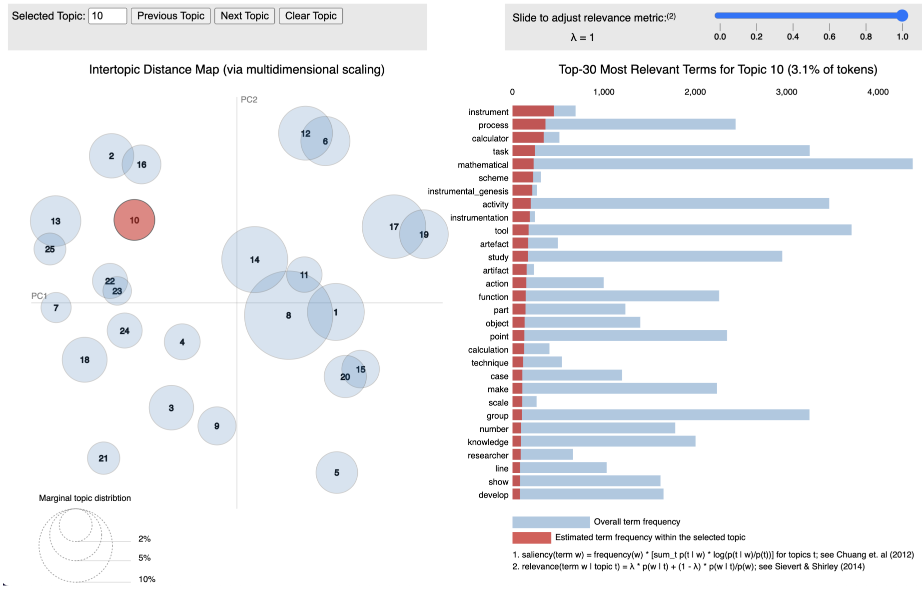

Fig. 1: Screenshot of pyLDAvis, a tool that we used to interactively explore our topic model. In this example, topic 10 is selected with $\lambda = 0.6$. The words instrument, calculator, and instrumental genesis are characteristic of that topic. This interactive visualisation can be found at https://zenodo.org/record/6319517.

4.4 Exploring the model

Once we had chosen our model, we created an interactive visualization with pyLDAvis (Sievert & Shirley, 2014). This visualization, illustrated in Figure 1, consists of two panels. In the left panel, the global topic view, one can see an intertopic distance map created via multidimensional scaling (MDS) applied to the topic distance matrix. The panel answers the following two questions: *How prevalent is each topic?* and *How do topics relate to each other?*. Topics with higher prevalence are visualized by the area of the circles. Moreover, circles that are closer together or even overlapping represent topics that appear together more frequently.

The right panel shows a bar chart of terms (with Topic 10 selected) that represent the individual terms that are most useful for interpreting the meaning of the selected topic. The blue bar represents the corpus-wide frequency, and the red bar represents the topic-specific frequency of the term. When hovering over a topic circle in the left panel, the right panel will show the most central terms for this topic. Hovering over a term on the right changes the left panel dynamically and shows in which topics the term occurs and with what frequency.

In our approach to interpret the topics' meanings, it is important to account both for terms that are exclusive to the topic and terms that appear very frequently in the topic. To account for this, pyLDAvis imposes a relevance metric, λ, which can scale both the importance of a term's topic exclusivity and its topic frequency (optimal λ stands at 0.6) (Sievert & Shirley, 2014).

4.5 Temporal evolution and relationships between topics

In order to investigate the evolution of topics, we analyzed whether or not a topic is trending by applying the Mann–Kendall test (Mann, 1945). Before running the test, we normalized the topic probability for each year, then ran the test. It is important to note that each paper in the corpus is considered to be a distribution of all the 25 topics (and, therefore, does not belong to only one topic). The trend analysis identifies evolutionary patterns for each topic (rather than individual papers) in the corpus.

We created a positive topic-correlation graph to investigate how the topics relate to each other. We constructed this graph by calculating the correlations between all topics and then removing all correlations below zero. The size of the nodes says something about how often and how much each node is correlated with other nodes, with the thickness of the edges indicating how strong the correlation is. The nodes are spread out in space by a force-directed algorithm (ForceAtlas2), whereby the strongly correlated nodes stay close to each other. The coloring results from a clustering algorithm called the modularity score, which identifies dense subclusters in a network. Based on the modularity score, four clusters emerge from this graph, depicted in Figure 2.

4.6 Steps of interpretation

We used the models described above for our analysis, following the steps described below:

1. We investigated each topic individually, using the right panel in pyLDAvis to ascertain which terms dominate the topic.
2. We inspected the five documents with the highest probability of belonging to the topic by examining the topic model and undertaking a qualitative manual inspection of the paper, going back and forth between the two approaches.
3. Steps 1 and 2 led us to find an appropriate name for each topic.
4. We investigated the clusters in the positive topic-correlation graph in order to understand the relationships between the different topics.
5. We used the Mann–Kendall tests to understand the trends' temporal evolution.

5 Findings

As mentioned in Section 4.6, we assigned names to the topics based on the most relevant terms ($\lambda = 0.6$) and the five papers with the highest topic probabilities. Most of these topics were easily identified from these resources; however, some were less obvious to interpret. We developed appropriate names for these less interpretable topics by interpreting similarities among the top contributing papers and the most relevant terms. We will describe in more detail how we named these specific topics. We have marked these less interpretable topics with a (*) against the topic number in Table 3.

Table 3: Topic names, terms, and trends.

Topic number	Topic name	Top-30 most relevant terms, sorted by relevance ($\lambda = 0.6$)	% of tokens	Trend ($p < 0.05$)
1*	Pupil-centered (primary and lower secondary)	Pupil, school, project, solution, work, task, year, give, tool, class, computer, write, teacher, make, method, good, mathematic, group, calculator, solve, question, problem, part, teach, competence, lesson, find, follow, easy, special	6%	Decreasing
2	Semiotics	Mathematical, meaning, sign, tool, object, artefact, digital, semiotic, discourse, process, perspective, task, refer, routine, analysis, affordance, construction, language, property, term, activity, product, concept, role, ball, perception, change, transition, world, element	3.6%	No trend
3	Algebra and CAS	Equation, expression, function, algebraic, algebra, window, sequence, solve, problem, numerical, variable, level, student, register, rule, system, suite, cas, environment, relation, difficulty, polynomial, symbolic, calculation, equivalence, recurrence_relation, maple, basic, conversion, method	3.7%	Decreasing
4	Problem solving	Problem, problem_solve, solve, solution, trigonometry, adult, problem_solving, competition, trigonometric, combinatorial, phase, reasoning, systemic_thinke, permutation, money, mathematical, system, represent, solver, pair, systematic, thinking, periodic, geogebra, triangle, fluency, process, relation, cosine, identify	2.4%	No trend
5	Effect studies	Test, score, group, result, post_test, control, low, high, school, effect, performance, frequency, item, grade, statistical, total, study, significantly, average, treatment, pre_test, table, correlation, ability, experimental, year, class, homework, achievement, difference	3.2%	No trend
6	Teacher practice in the classroom	Teacher, lesson, classroom, orchestration, discussion, work, practice, activity, episode, objective, observation, plan, technique, task, didactical, time, screen, instrumental_orchestration, teaching, format, case, didactic, set, room, goal, computer, teach, individual, pilot, interview	4.5%	No trend
7	Embodied interaction	Gesture, finger, touch, hand, screen, stretch, tap, pod, hold, movement, motion, action, body, touchscreen, manipulation, metaphor, ipad, move, multi_touch, count, tochcount, number, arzarello, place, dog, embody, pull, tangible, density, girl	1.7%	No trend
8*	Student activity	Student, study, question, learn, work, research, find, activity, follow, focus, give, provide, mathematic, time, problem, learning, relate, environment, answer, write, investigate, understand, result, mathematical, group, base, design, analysis, datum, present	14.4%	Increasing

Continued on next page

Table 3: continued from previous page

Topic number	Topic name	Top-30 most relevant terms, sorted by relevance ($\lambda = 0.6$)	% of tokens	Trend ($p < 0.05$)
9	Assessment and feedback	Task, feedback, correct, assessment, sample, step, strategy, test, diagnosis, error, solution, submit, automatically, incorrect, counter, number, item, response, tinkerplot, formula, memorization, calculate, intelligent, support, structure, goal, diagnostic, correctness, answer, produce	2.7%	Increasing
10	The instrumental approach	Instrument, calculator, instrumental_genesis, scheme, instrumentation, artifact, process, artefact, calculation, scale, pragmatic, signe, axis, task, action, rabardel, technique, page, chain, epistemic, operation, genesis, part, mathematical, mediate, activity, constitute, appropriation, hiccup, display	3.1%	No trend
11*	Programming and mobile learning	Programming, robot, mobile, activity, program, outdoor, trail, mathematic, tablet, gsp, step, computational, app, code, application, report, object, engagement, math_trail, eo, papert, engage, project, toontalk, task, concept, learning, outsource, math, mathematical	2.3%	Increasing
12*	Teacher education/ professional development	Teacher, trainee, session, practice, ict, situation, didactic, scenario, activity, didactical, instrumental, professional, evolution, dimension, group, propose, training, integration, educator, institutional, trainer, account, level, work, knowledge, analysis, math, organization, french, mathematique	5.4%	No trend
13	Dynamic geometry	Point, construction, triangle, geometry, cabri, geometrical, drag, geometric, figure, line, construct, area, software, dgs, segment, tool, circle, dynamic_geometry, property, prove, environment, proof, vertex, conjecture, dynamic, square, produce, diagonal, move, interpretation	4.7%	Decreasing
14*	Implementation of technology in the classroom	Learning, learn, environment, design, process, learner, system, base, model, simulation, support, theory, structure, principle, group, cognitive, representation, possibility, educational, software, complex, lead, content, work, concrete, experiment, describe, argumentation, create, basis	8.1%	No trend
15	Teaching materials	Resource, platform, quality, material, textbook, evaluation, exercise, document, search, author, user, criterion, metadata, dg, assistance, system, project, expert, content, usage, dynamic, review, high_quality, documentation, file, utilization, interactive, repository, book, choice	2.6%	Increasing

Continued on next page

Table 3: continued from previous page

Topic number	Topic name	Top-30 most relevant terms, sorted by relevance ($\lambda = 0.6$)	% of tokens	Trend ($p < 0.05$)
16*	Constructionism and creativity	Book, design, creativity, creative, idea, variable, turtle, constructionist, designer, subgroup, curvature, coi, share, logo, meaning_make, widget, team, digital, constructionism, cmt, windmill, microworld, coi_member, meaning_generation, riddle, angle, construction, line, book_unit, mathematical	2.8%	Increasing
17*	Overview papers	Technology, education, mathematic, research, teaching, paper, teacher, issue, digital, teach, impact, theme, tool, ict, mathematics, concern, classroom, approach, develop, belief, study, practice, integration, school, technological, et_al, design, development, twg, innovation	7.5%	No trend
18	Graphical representations of functions	Graph, function, representation, change, dynamic, slope, height, speed, coordinate, graphical, vertical, tool, horizontal, constant, water, graphical_representation, move, feature, worksheet, parameter, position, covariation, concept, conception, bottle, represent, graphs, qualitative, situation, task	3.7%	Increasing
19	Teacher knowledge when teaching with technology	Teacher, knowledge, geogebra, technology, teach, content, tpack, teaching, preservice, pedagogical, skill, participant, lesson, instructional, preparation, technological, integration, integrate, applet, stage, practice, experience, pre_service, prepare, study, program, mathematic, classroom, implementation, professional_development	4.4%	Increasing
20	E-/Blended-learning	Online, lecture, internet, video, peer, web, comment, flip, lecturer, forum, topic, face, post, site, motivation, community, learn, tutor, communication, week, share, website, seminar, virtual, response, interactive, argument, challenge, facebook, screencast	3.3%	No trend
21	Arithmetics	Fraction, child, number, balance, block, subtraction, representation, computation, division, addition, algebra_tile, represent, token, ten, count, decimal, model, pad, equivalent, bundle, multiplication, equivalence, quantity, denominator, place_value_chart, chart, operation, digit, tile, invent	1.9%	No trend
22	Spreadsheets	Pattern, width, spreadsheet, number, length, formula, child, microworld, symbolic, column, dependency, symbol, datum, generalisation, month, rectangle, generalization, rule, variable, tile, construct, transition, magnitude, cell, area, express, active_graphe, model, construction, expresser	2.3%	Decreasing

Continued on next page

Table 3: continued from previous page

Topic number	Topic name	Top-30 most relevant terms, sorted by relevance ($\lambda = 0.6$)	% of tokens	Trend ($p < 0.05$)
23	Game-based learning	Game, player, house, play, budget, sim, win, gameplay, child, visually_impaire, garden, scaffolding, chance, rule, gamification, probability, scaffold, goal, hintikka, shoot, event, understanding, slovak, probabilistic, family, furniture, puzzle, colour, hit, sight	1.5%	No trend
24	Calculus	Function, limit, calculus, concept, derivative, tall, integral, plot, tangent, interval, curve, inverse, approximation, discrete, definition, intuitive, discontinuity, area, conflict, upper, continuity, difference_quotient, point, process, sequence, rate, real, graph, infinite, approximate	2.3%	No trend
25	Linear algebra	Vector, transformation, matrix, space, plane, blue, movement, drag, mirror, eigenvector, machine, projection, overlap, move, dimensional, linear_transformation, boxing, linkage, geometric, direction, mark, notion, exhibition, slider, eigenvalue, scalar, image, prism, tsm, black	1.9%	No trend

Papers consisting of high proportions of words from the topic *pupil-centered (primary and lower secondary)* (Topic 1) focused on primary to lower-secondary pupils. This topic could be considered a sub-topic of Topic 8 (*student activity*). The main difference between Topics 1 and 8 is that Topic 1 includes the term "pupils," whereas Topic 8 includes the term "student." From reading the most dominant papers in Topic 1, it appears that Topic 1 explicitly focuses on primary and lower-secondary schools (Fuglestad 2005; 2007), whereas Topic 8 includes all grade levels (Biton et al., 2015; Kristinsdóttir et al., 2019). It is also possible that this points to a change of discourse (from pupil to student) rather than a change in the research object itself.

As expected, papers consisting of high proportions of words from *student activity* (Topic 8) focus on students and their activity. Unsurprisingly, this topic is the most represented in the papers. However, no individual paper dedicates more than 40.61% to this topic. This is unusually low for a big topic but explainable in terms of how it is difficult to consider a research paper in mathematics education that only considers student activity.

Papers with a high proportion of words in Topic 11, *programming and mobile learning*, focus on two aspects: programming in mathematics education; and mobile learning (via, for example, Math trail). Merging these two aspects results in this topic being an intruder topic, as the aspects do not necessarily have much in common. The topic could be connected by words like *app*, *application*, and *program*; for example, while Buteau et al. (2019) focus on student processes and strategies in programming for math investigations, Fabian (2019) evaluate student engagement in mobile learning activities.

Papers consisting of high proportions of words from *teacher education/professional development* (Topic 12) include descriptive studies that aim to investigate the practices and needs of inexperienced teachers (Assude, 2007; Emprin, 2007), studies that develop frameworks that inform the design of teacher-training resources (Emprin, 2007; 2010), and studies on the impact of specific teacher-training resources (Tapan, 2003).

The papers in Topic 14 *implementation of technology in the classroom* seek to support the systemic changes needed or initiated by implementing technology in classroom settings. An example of this is Wörler (2019), which focuses on developing a scheme to support teachers' selection of appropriate simulations for teaching and learning processes. Topic 14 also includes theoretical studies. Ulm (2010), for example, reports on the development of a framework that aims to introduce systemic innovations in educational systems. Thus, papers in Topic 14 include didactical design at a classroom level.

Papers consisting of high proportions of words from *constructionism and creativity* (Topic 16) focus on ways to foster mathematical creativity. We gave this topic its label with the knowledge that a common denominator in many papers is their constructionist approach; for example, Kolovou and Kynigos (2017) focus on designing e-books that incorporate dynamic constructionist widgets in order to foster creative mathematical thinking, while Papadopoulos et al. (2017) focus on meaning-making and mathematical creativity through problem-solving and constructionism.

The papers in Topic 17, *overview papers*, create an overview of an area of research on a given phenomenon. Topic 17 includes the introductions to the TWGs (Jones et al., 2002; Trgalova et al., 2011) and a paper that summarizes the use of graphics calculators in a specific geographical area (Routitsky & Tobin, 2001). The topic also includes papers that focus on overviewing technology integration as a broader concept (Lavicza et al., 2015).

5.1 The relationships between topics

As presented in Section 4.5, we constructed a positive topic-correlation graph to identify relationships between topics, this topic-correlation graph is represented in figure 2.

The nodes in the purple cluster are closely related, with the common themes being the types of software supporting cognitive processes and the specific parts and processes of mathematics that are closely linked to the utilization of this software. The topics *dynamic geometry*, and *spreadsheets*, purely concerned with this type of mathematical software. However, we also have *linear algebra*, and *problem-solving*, which concern specific parts of and processes in mathematics. Lastly, we have *algebra and CAS*, which are a mixture of these two focuses.

The common theme in the orange cluster appears to be technology utilization and experimentation with different mediation strategies and types of technology in primary and lower-secondary mathematics. For example, *pupil-centered*, and *arithmetics*, are closely related to primary and lower-secondary mathematics. The topic *pupil-centered* is concerned with this part of the educational system, while *arithmetics* is highly important in the years of primary and lower-secondary. In comparison, *game-based learning, implementation of technology in the classroom, semiotics*, and *embodied interaction* concern the utilization of or experimentation with technological mediation. *Game-based learning* and *implementation of technology in the classroom* are concerned with the use of technology in the classroom,

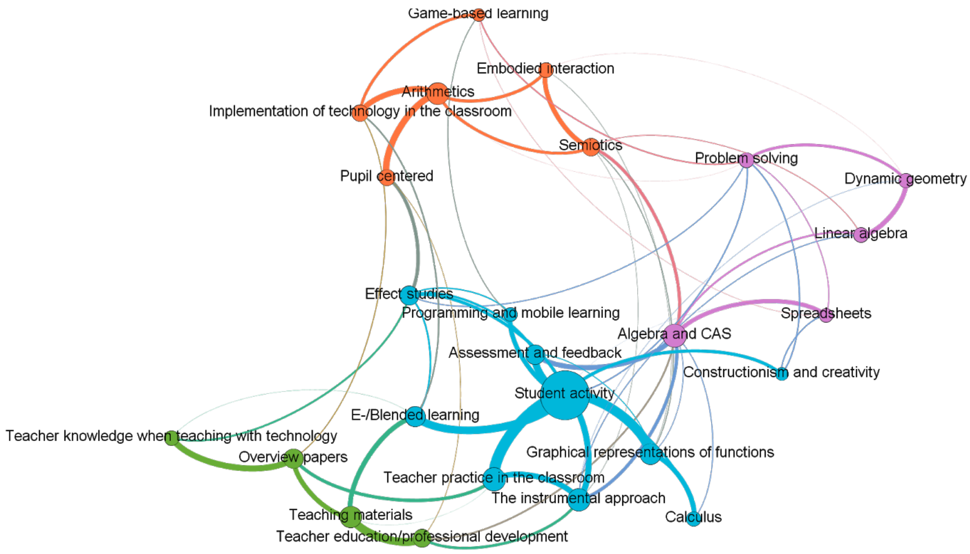

Fig. 2: Network visualisation of how the 25 topics (represented as nodes) are related. A link connects two topics if they are positively correlated. The bigger a node the more topics are positively correlated with this topic. Color represents topics that are thematically close.

while *semiotics* and *embodied interaction* is concerned with experimentation with technological mediation in mathematics.

The green cluster concerns teachers and teaching resources. *Teacher knowledge when teaching with technology, teacher education/professional development*, and *teaching materials* are all focused on this area of research. However, *overview papers* is a special topic, as it summarizes research conducted in the particular CERME year within the corresponding group for that year. It seems that much of the focus in this field is given to the teacher's role, resources, and classroom practice.

The blue cluster appears to be very diverse; it concerns activities in the classroom and theoretical/methodological approaches to mathematics education. In a way, this corresponds well with the thick edges in this cluster compared to those in the other clusters, as this could indicate that the role of this cluster is to connect the field.

5.2 The temporal evolution of topics

As noted earlier, we performed a Mann–Kendall test to identify whether certain topics are trending over the years that CERME has run. As indicated in Table 3, the topic *pupil-centered* has decreased, whereas the topic *student activity*, has increased. An essential difference between these topics is whether they refer to learners as "pupils" or as "students." This movement could be considered a mere change in wording. However, where papers in the *pupil-centered* topic explicitly focus on primary and lower-secondary schooling, papers in the *student activity* topic include studies at all levels of education. We could interpret this as a movement toward greater emphasis on the phenomenon studied rather than the educational level at which the phenomenon occurs.

Another trend at CERME involves increased focus on teacher knowledge when teaching with technology. This could be linked to a strong technological pedagogical content knowledge (TPACK) discourse that has become increasingly relevant as the role of technology in the mathematics classroom has grown (Chen et al., 2020). Increased attention is also being paid to technology's influence on assessment and feedback. As a result of this shift, we observe a greater focus on teaching materials, with a great deal of attention paid to teacher resources, especially within the instrumental approach.

Among the trending topics, we can observe studies on specific mathematical technologies. We see a decrease in papers dealing with *algebra and CAS, DGS* and *spreadsheets*, while papers focusing on *programming, mobile learning*, and *graphical representations of functions* are increasing. A likely explanation for the up-trending of *programming and mobile learning* is that programming has found its way into the mathematics curriculum in several countries. The graphical representations of functions may cause this, and a need to explore the overall potential of mathematical software in a way that is not dependent on specific technology is among the causes of this trend. Furthermore, while the above tools were previously considered new learning technologies, they are now viewed as a fundamental part of teaching and learning mathematics (Inglis & Foster, 2018). The decrease in these topics' prevalence could point to the fact that these technologies are now studied by specific TWGs that, for example, focus on algebra, statistics, and geometry. We also see a shift from algebra to *programming and mobile learning*, as well as to *constructionism and creativity*. In a sense, *programming and mobile learning* and *constructionism and creativity* are connected, as programming is the content and constructionism is the learning theory most closely associated with programming.

6 Discussion

The main goal of performing a literature review is to summarize and present an overview of an area and to identify research gaps or trending topics (Khoo et al., 2011).

6.1 The challenges with literature reviews

A constant increase in scientific output means that examining and reviewing every paper in an area in detail is highly challenging (Erren et al., 2009; Quinn et al., 2010). Hoyles and Noss (2003, p. 323) noted that "the vast corpus of study that now exists makes it no longer feasible to write a comprehensive review of the field [digital technology in mathematics education] in just one chapter". There are several methods to apply when conducting systematic reviews, such as critical reviews and meta-reviews (Grant & Booth, 2009). Despite different affordances and work processes associated with systematic review types, they are all centrally dependent on human processing. Topic models address this problem as they scale to larger corpora and help to overview amounts of text that otherwise would not be possible for human readers alone to process and organize into themes. When dealing with large datasets, topic models bring the advantage that they allow for a time-saving, structured, and transparent way of processing data that supports subsequent interpretation (Marks et al., 2020).

However, utilizing machine learning techniques in literature reviews also has its limitations. Even though we argue that this methodology is often more transparent than a systematic qualitative review (as every choice is denoted in the code), this transparency depends on understanding the approach's underlying technical mechanisms. There are also drawbacks regarding the ability to gain insight from specific sections of the paper (e.g., recommendations toward a practice or other research). In the topic modeling approach we adopted, all words are counted equal; as such, we do not consider that some sections (e.g., results sections) might be more important than others. Furthermore, topics modeled in LDA only consider the body text, ignoring figures or metadata (e.g., authors, date of publication, or source).

While researchers planning to use LDA should all consider these problems, many can be addressed. For example, it is possible (as is done in structural topic models (Chen et al., 2020)) to implement variables that include much of the metadata. However, as a simple rule when choosing machine learning models, a simpler model is usually more robust against overfitting (Srivastava et al., 2014). Because the corpus in this article is already highly pre-selected in virtue of being peer-reviewed papers specifically about technology use in mathematics education, we believe that implementing such variables would add too much complexity to justify any potential gains.

More generally, how we address the limitations of topic modeling depends on how we see them as an epistemological tool. While topic models can reveal new insights hard to obtain by (human) reading alone (Boyd-Graber et al., 2017) they do not replace other methods. They can be considered an algorithmic approach that complements existing practice when writing literature reviews. Since it supports (and does not replace) human interpretation, Marks et al., 2020 argues that topic modeling should be considered a mixed-methods approach rather than purely quantitative, despite being computationally driven. One could even argue that topic modeling should be seen as an algorithmic method of selecting what to read.

6.2 Similarities with other overviews

As mentioned in Section 2.1, the aim of the literature review study Trgalová et al. (2018) and this study partly coincide. It is important to note that the study also contribute similar findings. These similarities relate to the identified themes (which we refer to as "topics") such as teacher knowledge (which co-occurs in CERME 7, as reported in the chapter Trgalová et al. (2018), and Topic 19 in our model), mobile devices (CERME 8 and Topic 11), and the instrumental approach (CERME 4 and 5 and Topic 10).

However, although Trgalová et al. (2018) provides a valuable account of discussion themes of the technology TWGs at CERME, we can add new insights. Where Trgalová et al. (2018) independently identifies characteristics for individual CERME events, our study identifies topics across all events. While Trgalová et al.

(2018) identifies topic changes, our approach considers increases and decreases in topic prevalence throughout the entire history of the CERME technology TWGs.

We have shown a trend towards focusing on novel technology while giving less attention to more traditional ones, which is interesting when seen in the light of Trgalová et al.'s (2018) account of CERME 6, which brought forward the importance of considering the relationship between novel and traditional resources. Our approach is also different in that it allows us to conduct a positive correlation analysis, as illustrated in Figure 2. This analysis adds to Trgalová et al.'s (2018) finding in that it explores the relative distances between discourses in the textual corpus, thus allowing us to cluster the topics.

6.3 Identifying research gaps

By comparing our work with a model of mathematics education research (Inglis & Foster, 2018), and a model of research on technology in education (Chen et al., 2020), we can identify similarities with topics across the fields. Thus, we have created Table 4, in which we have noted topics from our model that share similarities with the topics found by Inglis and Foster (2018) and Chen et al. (2020). Some topics are marked with a (*), indicating that these topics are not one-to-one counterparts but are often more general or more specialized.

To compare the three models (our model, Inglis and Foster's (2018), and Chen et al.'s (2020)), we looked at both the terms that constitute the topics and the naming of the topics, which is why some topics might appear to be different if we consider them only based on their names. For example, as seen in Table 4, our topic *pupil-centered* is similar to the topic *mathematics education around the world* found by Inglis and Foster (2018). As we know the similarities between the models, we can now explore their differences and identify potential research gaps (e.g., topics that seem overlooked in the CERME technology context). We do so by identifying particular topics that our model does not include but are identified by either of the other models. That means looking at topics in Table 1 that do not relate to the similarities we identify in Table 4. We identified six groups of topics:

– ***Unsurprising focus differences***: this is a catch-all group for differences that are immediately understandable by virtue of differences between the journals' advertised focuses or format traditions. Inglis and Foster (2018) includes *discussions, reflections, and essay*, and *history and obituaries*, while Chen et al. (2020) contains *science education* and *language learning*. Science education encompasses mathematics education and more, while language learning is entirely different from mathematics education.
– ***Mathematics for all***: this includes anything about making mathematics available to and approachable for everyone. Inglis and Foster (2018) include *equity, gender, multilingual learners*, and *sociocultural theory*, while Chen et al. (2020) have *demographic issues* and *special education*. Our model has none of these; this may mean that these topics are not considered by research into the use of technology in mathematics education or that they have not received the same amount of attention, which could point to a research gap.
– ***Theories of learning***: this is represented in Inglis and Foster's (2018) model by topics on *constructivism*, and *didactical theories*, while Chen et al. (2020)

Table 4: All of the topics identified in our model were assessed for similarities with topics identified by Inglis and Foster (2018) and Chen et al. (2020).

Topic names for our model	Inglis and Foster (2018) (Topic model of Mathematics education research	Chen et al. (2020) (Topic model of Research on educational technology)
Pupil-centered (primary and lower secondary)	Mathematics education around the world*	
Semiotics	Semiotics and embodied cognition*	
Algebra and CAS	School algebra*	
Problem solving	Problem solving	
Effect studies	Experimental designs & Quantitative assessment of reasoning*	Experiments and methodologies
Teacher practice in the classroom	Observation of classroom discussion	
Embodied interaction	Semiotics and embodied cognition	Human computer interaction*
Student activity		
Assessment and feedback	Novel assessment*	Assessment
The instrumental approach		
Programming and mobile learning		Programming language & Mobile learning and early childhood education*
Teacher education/professional development	Teachers' knowledge and beliefs*	
Dynamic geometry	Dynamic geometry and visualization & Euclidean geometry*	Hardware*
Implementation of technology in the classroom		
Teaching materials		
Constructionism and creativity		
Overview papers		
Graphical representations of functions		
Teacher knowledge when teaching with technology	Teachers' knowledge and beliefs	Teacher training
E-/Blended learning	E-learning & policy	Blended learning
Arithmetics	Addition and subtraction	
Spreadsheets		
Game-based learning		Game-based learning
Calculus	Analysis	
Linear algebra		

do not consider any topics in this area. Our model includes *constructionism and creativity, instrumental approach, teacher knowledge when teaching with technology,* and *teacher practice in the classroom*. In this group, we can see different topics that address the same overall need for a theoretical foundation; this could be explained by the need for different theories to address specific areas, but it could also point to research gaps. Identifying such gaps would require particular attention, similar to the work undertaken within the networking of theories (Bikner-Ahsbahs & Prediger, 2014).

- ***Policy intervention research***: The topic model created by Inglis and Foster (2018) includes *curriculum (especially reform)*, while the model by Chen et al. (2020) includes *e-learning and policy* and *technology acceptance model*. Our model covers similar themes via *implementation of technology in the classroom*. The main difference here is that our model is more about adoption than acceptance and access.
- ***Mathematical content/competency areas***: Chen et al.'s (2020) model does (for obvious reasons) not cover topics included in this group, however among the topics found by Inglis and Foster (2018) themes like *formal analysis, proof and argumentation, rational numbers, spatial reasoning,* and *statistics and probability* are present. Our model has *algebra and CAS, problem-solving, programming and mobile learning, arithmetic, calculus*, and *linear algebra*. Interestingly, the mathematical content that is the focus of mathematics education in general and technology in mathematics education is quite different. Some of the topics included by Inglis and Foster (2018) that are not included in our model are easily explained; for instance, formal analysis is not included as there are no apparent advantages to conducting/writing formal analysis using technology. However, it is not as easy to explain the nonappearance of the other differences.
- ***Learning resources/environments***: Inglis and Foster (2018) did not find topics that are included in this group, while Chen et al. (2020) identified the topics *context & collaborative learning, social networks and communities, online and web-based learning, MOOCs*, and *hardware*. Our model contains *spreadsheets, dynamic geometry, programming, and mobile learning* and *algebra and CAS*. This difference of focus seems understandable, as it indicates a distinction between the technology used in education and that used in learning mathematics. However, there may still be mathematics education-specific issues relating to how technology is used in education that are under-explored.

When identifying topic discrepancies between the journals, another type of higher-level reflection emerged: situations where our model addresses a topic that can be seen as the integration or synthesis of a topics that emerged in the models created by Inglis and Foster (2018) and Chen et al. (2020). One example of this is that Inglis and Foster (2018) include the topic *statistics and probability* and Chen et al. (2020) include *hardware* (this topic includes software [and specifically spreadsheets] as terms), while the synthesis in our model seems to be *spreadsheets*, as this is a research area where the focus is on the learning potential and implications concerning the use of such software in the field of statistics and probability in mathematics education. Other examples include, for the same reasons, *Euclidean geometry* from Inglis and Foster's (2018) model and *hardware* from the model created by Chen et al. (2020), with the synthesis in our model being *dy-*

namic geometry, while *embodied interaction* in our model seems to be a synthesis of *semiotics and embodied cognition*, a topic found by Inglis and Foster (2018) and *human-computer interaction* from Chen et al.'s (2020) model. This pattern is expected, as the topics in our model innovate by addressing a topic in Inglis and Foster's (2018) model using technological topics in Chen et al. (2020) topic model. An interesting follow-on from this pattern is that it allows us to look for potential synthetic topics that do not yet exist. These would form another as-yet-unidentified research gap.

Earlier, we identified topics (or groups of topics) that exist in the topic models created by Inglis and Foster (2018) and Chen et al. (2020). We believe that the lack of one of these can be seen in the space between *virtual reality (VR)* that was identified by Chen et al. (2020) and *spatial reasoning* that was identified by Inglis and Foster (2018). Here we can imagine a topic for our model that focuses on using VR (or augmented reality (AR)) to develop spatial reasoning. The addition of this topic would not be that surprising, given that the application GeoGebra already employs AR technology for learning geometry. However, through a preliminary search on Google Scholar, we only found a single paper (by (Fowler et al., 2021)) focusing on this aspect of mathematics education.

7 Conclusion

In this paper, we present the application of LDA topic-modeling to the proceedings from CERME 1–11, focusing on the Thematic Working Groups related to technology, identifying 25 distinct topics. Based on this model, we create a graph of the positive correlations among the topics and calculate their modularity scores in order to identify clusters. Thus, we identify four clusters of topics: digital tools, teachers and their resources, technology experimentation, and a diverse cluster with a strong focus on student activity. The model shows that teachers and their resources are seldom associated with digital tools.

Further to this, we employed a Mann–Kendall test to understand how the research interest in our topics has decreased or increased over time, noting three main points: first, the technologies studied at a given time are predominantly those that are relatively new then; second, that there is an increased focus on the knowledge required by the teacher when teaching using technology; and third, that a shift has occurred from studying "pupil" activity to a move towards studying "student" activity.

We have related our research to three overviews of closely related fields: (Chen et al., 2020; Inglis & Foster, 2018; Trgalová et al., 2018). By doing so, we observed that some technologies had gained sufficient traction and are now also studied in general mathematics education. This is evident for DGS, which Inglis and Foster (2018) identifies as an independent topic in general mathematics education. We have also noted that not all technologies have reached this point. Spreadsheets are an example of technology that is not an independent topic studied in general mathematics research. This is remarkable given that it has seen mainstream use for many years and has become increasingly relevant due to its role as a data-processing tool. However, this observation has allowed us to identify specific characteristics of discourses that must be prominent enough to support independent topics in their own right. Such potential characteristics may include using theories (such as the instrumental approach and constructionism) and processes related to teaching with technology (such as implementing technology in the classroom and content-specific technologies (such as CAS and DGS).

This paper can serve as a starting point within the field of technology in mathematics education research, as it compares previous attempts to overview the field, identifies new topics, and shows trends in the prevalence of these topics.

8 Data availability

The material generated and analysed during the current study are available in the Zenodo repository, https://zenodo.org/record/6319517

9 Conflict of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Asmussen, C. B., & Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0255-7

Assude, T. (2007). Teachers' practices and degree of ict integration. In D. Pitta-Pantazi & C. Philippou (Eds.), *European Research in Mathematics Education V: Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education (CERME5, February 22–26, 2007)* (pp. 1339–1348). University of Cyprus; ERME.

Bikner-Ahsbahs, A., & Prediger, S. (2014). *Networking of theories as a research practice in mathematics education* (A. Bikner-Ahsbahs & S. Prediger, Eds.). Springer International Publishing. https://doi.org/10.1007/978-3-319-05389-9

Biton, Y., Hershkovitz, S., & Hoch, M. (2015). Peer learning in mathematics forum on facebook: A case study. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education (CERME 9, February 4 – 8, 2015)* (pp. 2473–2479).

Bittermann, A., & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, *226*(1), 3–13. https://doi.org/10.1027/2151-2604/a000318

Blair, S. J., Bi, Y., & Mulvenna, M. D. (2019). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, *50*(1), 138–156. https://doi.org/10.1007/s10489-019-01438-z

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, *1*(1), 17–35. https://doi.org/10.1214/07-aoas114

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Borba, M. C., Askar, P., Engelbrecht, J., Gadanidis, G., Llinares, S., & Aguilar, M. S. (2017). Digital technology in mathematics education: Research over the last decade. In G. Kaiser (Ed.), *Proceedings of the 13th international congress on mathematical education* (pp. 221–233). Springer International Publishing.

Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of topic models. *Foundations and Trends® in Information Retrieval*, *11*(2-3), 143–296. https://doi.org/10.1561/1500000030

Bray, A., & Tangney, B. (2015). Enhancing student engagement through the affordances of mobile technology: A 21st century learning perspective on realistic mathematics education. *Mathematics Education Research Journal*, *28*(1), 173–197. https://doi.org/10.1007/s13394-015-0158-7

Buteau, C., Muller, E., Dreise, K., Mgombelo, J., & Sacristán, A. I. (2019). Students' process and strategies as they program for mathematical investigations and applications. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME11, February 6–10, 2019)* (pp. 2796–2803).

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of computers & education. *Computers & Education*, *151*, 103855. https://doi.org/10.1016/j.compedu.2020.103855

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), 168–189. https://doi.org/10.1017/pan.2017.44

Emprin, F. (2007). Analysis of teacher education in mathematics and ict. In D. Pitta-Pantazi & C. Philippou (Eds.), *European Research in Mathematics Education V: Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education (CERME 5, February 22–26, 2007)* (pp. 1399–1408). University of Cyprus; ERME.

Erren, T. C., Cullen, P., & Erren, M. (2009). How to surf today's information tsunami: On the craft of effective reading. *Medical Hypotheses*, *73*(3), 278–279. https://doi.org/10.1016/j.mehy.2009.05.002

Fabian, K. (2019). Student engagement in mobile learning activities: Breakdowns and breakthroughs. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME11, February 6–10, 2019)* (pp. 2815–2825).

Foster, C., & Inglis, M. (2018). Mathematics teacher professional journals: What topics appear and how has this changed over time? *International Journal of Science and Mathematics Education*, *17*(8), 1627–1648. https://doi.org/10.1007/s10763-018-9937-4

Fowler, S., Cutting, C., Kennedy, J., Leonard, S. N., Gabriel, F., & Jaeschke, W. (2021). Technology enhanced learning environments and the potential for enhancing spatial reasoning: A mixed methods study. *Mathematics*

Education Research Journal, 1–24. https://doi.org/10.1007/s13394-021-00368-9

Graham, S., Weingart, S., & Milligan, I. (2012). Getting started with topic modeling and MALLET (A. Crymble, Ed.). *Programming Historian*, *1*(1). https://doi.org/10.46430/phen0017

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(suppl_1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Hoyles, C., & Noss, R. (2003). What can digital technologies take from and bring to research in mathematics education? In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 323–349). Springer Netherlands. https://doi.org/10.1007/978-94-010-0273-8_11

Inglis, M., & Foster, C. (2018). Five decades of mathematics education research. *Journal for Research in Mathematics Education*, *49*(4), 462–500. https://doi.org/10.5951/jresematheduc.49.4.0462

Johnstone, W. (1990). *The role of computer algebra systems in mathematics teaching at upper secondary school* (Doctoral dissertation). Edith Cowan University.

Jones, K., Lagrange, J.-b., & Lemut, E. (2002). Tools and technologies in mathematical didactics. In J. Novotná (Ed.), *European research in mathematics education ii: Proceedings of the second conference of the european society for research in mathematics education* (pp. 125–127). Charles University, Faculty of Education; ERME.

Khoo, C. S. G., Na, J.-C., & Jaidka, K. (2011). Analysis of the macro-level discourse structure of literature reviews. *Online Information Review*, *35*(2), 255–271. https://doi.org/10.1108/14684521111128032

Kissane, B., McConney, A., & Ho, K. F. (2015). *Review of the use of technology in mathematics education and the related use of cas calculators in external examinations and in post school tertiary education settings* (tech. rep.). Perth, WA: School Curriculum and Standards Authority.

Kolovou, A., & Kynigos, C. (2017). Teachers designing e-books to foster creative mathematical thinking: The case of curvature. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education (CERME10, February 1–5, 2017)* (pp. 2422–2429).

Kristinsdóttir, B., Hreinsdóttir, F., & Lavicza, Z. (2019). Silent video tasks: Towards a definition. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME11, February 6–10, 2019)* (pp. 2709–2710).

Laborde, C., & Laborde, J.-M. (1995). The case of cabri-géomètre: Learning geometry in a computer based environment. In D. Watson & D. Tinsley (Eds.), *Integrating information technology into education* (pp. 95–106). Springer US. https://doi.org/10.1007/978-0-387-34842-1_10

Lagrange, J.-B., Artigue, M., Laborde, C., & Trouche, L. (2003). Technology and mathematics education: A multidimensional study of the evolution of research and innovation. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 237–269). Springer Netherlands. https://doi.org/10.1007/978-94-010-0273-8_9

Lau, J. H., Newman, D., & Baldwin, T. (2014). Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In S. Wintner, S. Goldwater, & S. Riezler (Eds.), *Proceedings of the 14th conference of the european chapter of the association for computational linguistics* (pp. 530–539).

Lavicza, Z., Juhos, I., Koren, B., Fenyvesi, K., Csapodi, C., Kis, M., & Mantecón, J.-D. (2015). Integrating technology into primary and secondary school teaching to enhance mathematics education in hungary. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education (CERME 9, February 4–8, 2015)* (pp. 2430–2431).

Lim, K. W., & Buntine, W. (2014). Twitter opinion topic model. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 1319–1328. https://doi.org/10.1145/2661829.2662005

Maier, D., Niekler, A., Wiedemann, G., & Stoltenberg, D. (2020). How document sampling and vocabulary pruning affect the results of topic models. *Computational Communication Research*, *2*(2), 139–152. https://doi.org/10.5117/ccr2020.2.001.maie

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2-3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, *13*(3), 245. https://doi.org/10.2307/1907187

Marks, R., Foster, C., Barclay, N., Barnes, A., & Treacy, P. (2020). A comparative synthesis of UK mathematics education research: What are we talking about and do we align with international discourse? *Research in Mathematics Education*, *23*(1), 39–62. https://doi.org/10.1080/14794802.2020.1725612

McCallum, A. K. (2002). *MALLET: A machine learning for language toolkit* [http://www.cs.umass.edu/ mccallum/mallet].

Mimno, D. (2012). Computational historiography. *Journal on Computing and Cultural Heritage*, *5*(1), 1–19. https://doi.org/10.1145/2160165.2160168

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 262–272. https://aclanthology.org/D11-1024

Papadopoulos, I., Diamantidis, D., & Kynigos, C. (2017). Meaning-generation through an interplay between problem solving and constructionism in the c-book technology environment. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the Tenth Congress of the European Society for Research* *in Mathematics Education (CERME10, February 1–5, 2017)* (pp. 2619–2626).

Papert, S. (1982). *Mindstorms*. Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-5357-6

Paul, M., & Girju, R. (2009). Topic modeling of research fields: An interdisciplinary perspective. *Proceedings of the International Conference RANLP-2009*, 337–342. https://aclanthology.org/R09-1061

Puentedura, R. (2006). Transformation, technology, and education. http://hippasus.com/resources/tte/

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, *54*(1), 209–228. https://doi.org/10.1111/j.1540-5907.2009.00427.x

Ramage, D., Rosen, E., Chuang, J., Manning, C. D., & McFarland, D. A. (2009). Topic modeling for the social sciences. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems (nips): Workshop on applications for topic models: Text and beyond*.

Roberts, D. L., Leung, A. Y. L., & Lins, A. F. (2012). From the slate to the web: Technology in the mathematics curriculum. *Third international handbook of mathematics education* (pp. 525–547). Springer New York. https://doi.org/10.1007/978-1-4614-4684-2_17

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Ronau, R. N., Rakes, C. R., Bush, S. B., Driskell, S. O., Niess, M. L., & Pugalee, D. K. (2014). A survey of mathematics education technology dissertation scope and quality. *American Educational Research Journal*, *51*(5), 974–1006. https://doi.org/10.3102/0002831214531813

Routitsky, A., & Tobin, P. (2001). Graphics calculators use in mathematics victorian secondary schools. In J. Novotná (Ed.), *European research in mathematics education ii: Proceedings of the second conference of the european society for research in mathematics education* (pp. 227–240). Charles University, Faculty of Education; ERME.

Schofield, A., Magnusson, M., & Mimno, D. (2017). Pulling out the stops: Rethinking stopword removal for topic models. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 432–436. https://doi.org/10.18653/v1/e17-2069

Schofield, A., & Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics*, *4*(0), 287–300. https://doi.org/10.1162/tacl_a_00099

Schofield, A., Thompson, L., & Mimno, D. (2017). Quantifying the effects of text duplication on semantic models. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2737–2747. https://doi.org/10.18653/v1/d17-1290

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning,* *Visualization, and Interfaces*, 63–70. https://doi.org/10.3115/v1/w14-3110

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929–1958.

Tapan, S. (2003). Integration of ict in the teaching of mathematics in situations for treatment of difficulties in proving. In M. A. Mariotti (Ed.), *European Research in Mathematics Education III: Proceedings of the Third Conference of the European Society for Research in Mathematics Education (CERME3, February 28–March 3, 2003)* (pp. 1–8). University of Pisa; ERME.

Trgalova, J., Fuglestad, A. B., Maracci, M., & Weigand, H.-G. (2011). Introduction to the papers of wg15. In M. Pytlak, T. Rowland, & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education (CERME7, February 9–13, 2011)* (pp. 2144–2147). University of Rzeszów; ERME.

Trgalová, J., Clark-Wilson, A., & Weigand, H.-G. (2018, April). Developing research in mathematics education. In T. Dreyfus, M. Artigue, D. Potari, S. Prediger, & K. Ruthven (Eds.), *Developing research in mathematics education*. Routledge. https://doi.org/10.4324/9781315113562

Ulm, V. (2010). Systemic innovations of mathematics education with dynamic worksheets as catalysts. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education (CERME6, January 28–February 1, 2009)* (pp. 1280–1289). Institut National de Recherche Pédagogique; ERME.

Vega-Carrasco, M., O'sullivan, J., Prior, R., Manolopoulou, I., & Musolesi, M. (2020). Modelling grocery retail topic distributions: Evaluation, interpretability and stability. *arXiv preprint arXiv:2005.10125*. https://doi.org/10.48550/ARXIV.2005.10125

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, 1105–1112. https://doi.org/10.1145/1553374.1553515

Williams, J., & Goos, M. (2012). Modelling with mathematics and technologies. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 549–569). Springer New York. https://doi.org/10.1007/978-1-4614-4684-2_18

Wörler, J. F. (2019). How to distinguish simulations? development of a classification scheme for digital simulations for teaching and learning mathematics. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education (CERME11, February 6–10, 2019)* (pp. 2757–2764).

Zbiek, R. M., Heid, M. K., Blume, G. W., & Dick, T. P. (2007). Research on technology in mathematics education: A perspective of constructs. In K. Lester Frank (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics* (pp. 1169–1207). IAP.