Educational Research Review

Mapping the Landscape of Science Learning in Informal Environments from 2009 to 2024: A Review of the Literature Using Topic Modeling

--Manuscript Draft--

Manuscript Number:	EDUREV-D-24-00986
Article Type:	Review article
Keywords:	science learning; science education; informal environment; PRISMA; topic modeling
Abstract:	The present study reviewed the literature on science learning in informal environments from 2009 to 2024. Following the PRISMA guidelines, we screened 7,006 articles and included 552 for detailed analysis. Using structural topic modeling, we identified nine research topics and documented their evolution over time. These topics explored different aspects of science learning across various informal environments, such as learning in museums and science centers, online science learning, family science engagement, and the role of social media in science. The most prevalent topic involved how science was portrayed and communicated in the news media, while the least prevalent topic involved STEM motivation and gender stereotypes. Most topics remained relatively stable, but some appeared to be gaining increased interest, including the role of social media in science leaning, STEM motivation and gender stereotypes, and family science engagement. This study is the first comprehensive review of the informal science learning literature since the National Research Council's report in 2009. The findings were discussed in relation to the six strands of science learning and the ecological model of learning as outlined in that report. Overall, this study provides an overview of the extensive literature on science learning over the past 15 years and highlights important implications for theory, research, and practice.

Highlights

- This study reviewed 552 documents on informal science learning.
- This was the first review study on the topic since the NRC's 2009 report.
- Topic modeling was used to identify nine research themes in the literature.
- News media and science was the most prevalent topic.
- Rising interest in social media, gender stereotypes, and family engagement.

Abstract

The present study reviewed the literature on science learning in informal environments from 2009 to 2024. Following the PRISMA guidelines, we screened 7,006 articles and included 552 for detailed analysis. Using structural topic modeling, we identified nine research topics and documented their evolution over time. These topics explored different aspects of science learning across various informal environments, such as learning in museums and science centers, online science learning, family science engagement, and the role of social media in science. The most prevalent topic involved how science was portrayed and communicated in the news media, while the least prevalent topic involved STEM motivation and gender stereotypes. Most topics remained relatively stable, but some appeared to be gaining increased interest, including the role of social media in science leaning, STEM motivation and gender stereotypes, and family science engagement. This study is the first comprehensive review of the informal science learning literature since the National Research Council's report in 2009. The findings were discussed in relation to the six strands of science learning and the ecological model of learning as outlined in that report. Overall, this study provides an overview of the extensive literature on science learning over the past 15 years and highlights important implications for theory, research, and practice.

Keywords: science learning; science education; informal environment; PRISMA; topic modeling

Mapping the Landscape of Science Learning in Informal Environments from 2009 to 2024: A Review of the Literature Using Topic Modeling

Introduction

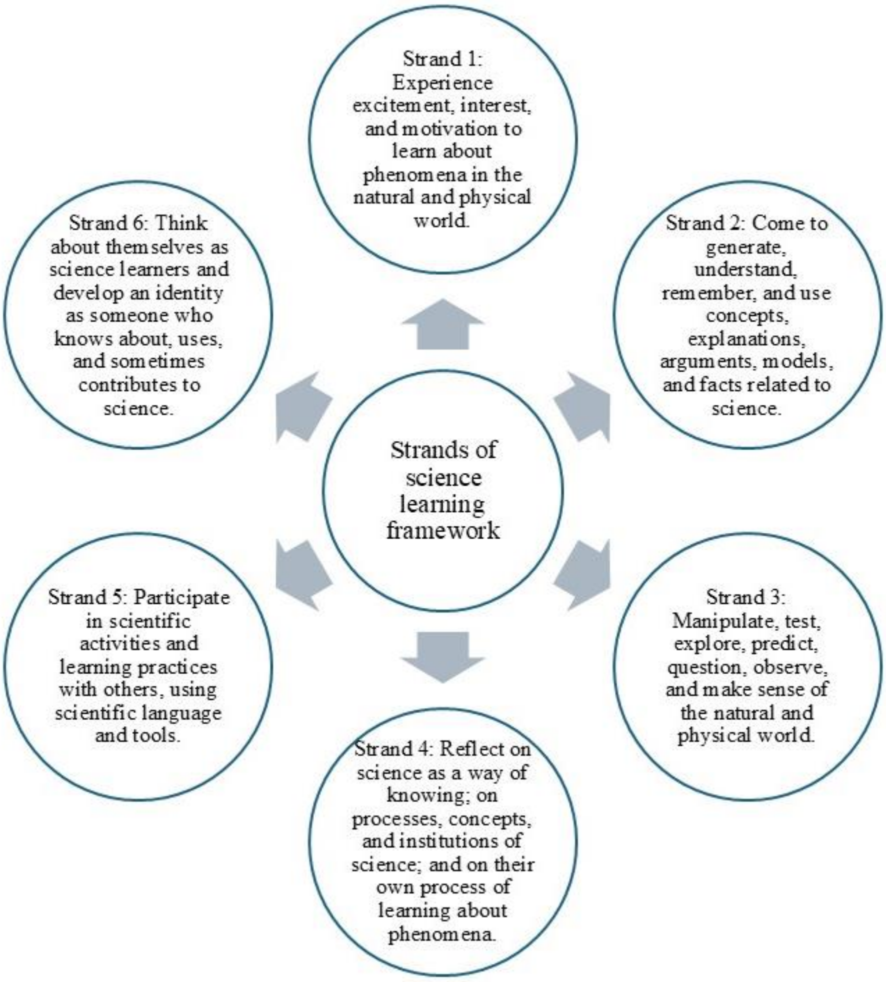
Rapid advances in science and technology are reshaping contemporary society in fundamental ways. Against this background, science competencies are increasingly needed for individuals to successfully navigate the 21st century (Klahr et al., 2011). Science not only plays an important role in academic achievement, but possessing science-related knowledge and skills can enhance one's career prospects, particularly given the growing demand for STEM (science, technology, engineering, and mathematics) majors in the job market (Habig, 2020). Moreover, understanding science is vital for personal decision-making about science-related matters, such as environmental protection and personal health (e.g., vaccination) (e.g., Hetherington et al., 2017). This importance is further amplified in the current climate of scientific misinformation and disinformation, which demands a scientifically literate citizenry who can evaluate the credibility of scientific information and make well-informed decisions (Osborne & Pimentel, 2023). Due to the widespread recognition of the importance of science, it has been increasingly integrated into the school curriculum throughout all levels of education, starting from preschool all the way through college and beyond (e.g., National Research Council, 2012).

Although formal school education serves as an important channel for science learning, evidence suggests that much if not most science is actually learnt outside of school settings (Falk & Dierking, 2010), with many science learning opportunities arising in the so-called informal environments. The National Research Council (2009) described three types of informal environments: everyday settings and family activities, designed settings, and science programs. Although the word "setting" is used for the first category, it does not denote any single setting but refers to a collection of everyday activities and routines where opportunities for science learning may occur. A hallmark characteristic of everyday and family science learning is that there is often no explicit goal of teaching or learning science. Rather, science-related elements are woven into the fabric of daily life and often encountered accidentally or opportunistically. For example, science may be encountered in dinner-table conversations, when pursing a hobby such as fishing or gardening, or even when caring for a sick family member. Designed settings refer to institutions such as museums, aquariums, and zoos as well as components within such settings (e.g., exhibitions and demonstrations). Unlike everyday settings, these settings are usually designed with the intentional purpose of science learning. Science learning also tends to be structured by participants and is experienced sporadically, but it also reflects the objectives of designers and educators. The final category represents science learning programs that are more structured and tend to have curricular goals. These programs are usually led by a professional rather than self-organized and can extend for weeks or months. To some extent, they mirror traditional classroom-based science instruction.

What constitutes science learning in these informal environments is also different from those typically considered in schools. While school education tends to prioritize disciplinary knowledge and skills, informal science learning encompasses a broader set of constructs. The National Research Council (2009) proposed the strands of science learning framework that clearly articulates "the science-specific capabilities supported by informal environments." (p. 3). This framework complements school science education and includes six strands (see Figure 1 for an overview). For example, Strands 2 and 3 focus on understanding scientific knowledge and developing scientific reasoning skills. Strand 1 underscores the importance of motivation in science. Strand 5 pertains to scientific practices, such as engaging in science activities or utilizing scientific tools. Meanwhile, Strand 4 emphasizes the epistemological aspects of science, portraying it as a distinct way of knowing. Lastly, Strand 6 highlights the development of identity in science. This model serves as an organizing framework for the areas and aspects of science learning that are characteristic of informal environments and has important implications for practice and research.

Figure 1

The Strands of Science Learning Framework Adapted from the National Research Council (2009)



The role of informal environments in science learning is informed by various contemporary theories, particularly cognitive and sociocultural theories of learning (Vygotsky, 1979). Among many others, key frameworks include the Contextual Model of Learning (Falk & Dierking, 2000), which emphasizes the sociocultural aspects of learning and the significance of the physical context, and the Multiple Identities Framework (Tate & Linn, 2005; Packard, 2003), which explores how participation in specific activities influences identity formation (see National Research Council, 2009, for a succinct summary). These theories provide a strong foundation for understanding the role that different informal environments play in science learning and how such environments should be designed. Integrating the diverse perspectives, the National Research Council (2009) proposed an integrative ecological model of learning. This framework acknowledges the ecological nature of human learning, characterizing it through three interrelated aspects: people, places, and cultures. Each aspect highlights specific factors relevant to the learning process. In a nutshell, they denote the observation that individual characteristics such as prior knowledge and metacognition affect learning, some learning takes place in certain settings, and that all learning is a culturally mediated process.

Over the past few decades, an extensive body of research has examined science learning in informal environments. For example, existing research has investigated such diverse informal settings as the home learning environment (Bae et al., 2023; Junge et al., 2021), museum exhibits/settings (Booth et al., 2020; Callanan et al., 2017), zoos and aquariums (Kelly et al., 2014), science programs (e.g., Schiefer et al., 2024), and social media (e.g., Greenhow et al., 2015; Maier et al., 2014). The target age of the participants covered almost any age group, including preschool children (Bae et al., 2023; Junge et al., 2021), primary and secondary school students (e.g., Maiorca et al., 2021), college students (e.g., Goff et al., 2020), as well as the broader public (e.g., Geiger et al., 2017). The science-related constructs examined in these informal science learning contexts include performance (e.g., Bae et al., 2023; Junge et al., 2021) but also include attitudes (e.g., Kelly et al., 2014), beliefs (e.g., Schiefer et al., 2022), interest (e.g., Maiorca et al., 2021), career aspirations (e.g., Maiorca et al., 2020), and interactions (e.g., Callanan et al., 2017), among other things. These studies showcase the vibrancy of the field and the immense interest in science learning in informal environments.

Despite a growing interest in this field, the existing literature remains fragmented, with insufficient synthesis to facilitate a systematic examination. There are only two published documents on this topic: the 2009 National Research Council report and the edited volume by Schweingruber and Fenichel (2010). The first report was a collaborative effort by 14 committee members commissioned by the National Research Council to survey the evidence of science learning across different venues, learner groups, and different spans of time. It also addressed issues such as the theoretical perspectives underpinning research on science learning, equity and diversity, and the role of science media. This report provided a solid evidence base for what scientists knew at the time, serving as a comprehensive framework for subsequent research and the design of informal science learning environments. The work by Schweingruber and Fenichel (2010) was not entirely distinct. In fact, it was grounded in the findings of the 2009 report and primarily aimed at communicating these findings to a diverse audience of practitioners, such as educators, museum professionals, and publishers, among others. While these documents were comprehensive and rigorous, enhancing our understanding of informal science learning in many ways, their expansive focus, substantial volume, and narrative style may have rendered them less accessible. Moreover, almost 15 years have passed since their publication, and the field has continued to evolve. There is a pressing need for a review study that uses systematic methods and adheres to established scientific protocols, to assess the current literature, evaluate the progress made, and identify potential directions for future exploration.

The goal of this investigation was to identify key research topics and trace the evolution of the field over time. Given that purpose and the voluminous literature that exists, we used topic modeling, a natural language processing technique, to uncover research topics or themes from textual data. This approach is well-suited for situations where there exists an extensive literature and the primary purpose is to uncover the general patterns or trends therein (Weston et al., 2023). Using topic modeling, we aimed to answer the following two research questions (RQs):

RQ1: What are the topics studied in the informal science learning literature from 2009 to 2024?

RQ2: How have the topics studied in the informal science learning literature evolved from 2009 to 2024?

Method

The present review was conducted with reference to the updated 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). The review was not preregistered, and due to the exploratory nature of our study, a review protocol was not prepared. The search code, extracted data, and analytic procedures are not publicly available but can be made available by the corresponding author upon reasonable request. We also reported whether and where each of the 27 items in the PRISMA checklist was addressed in the present study in the Supplementary Materials (see Table S1 for details).

Literature Search

We conducted a systematic literature search on August 14, 2024, using the Web of Science (WOS) Core Collection database. We compiled a list of terms and entered these terms in the title, abstract, keyword plus, and author keywords fields for our literature search using the advanced search function of WOS. The terms were informed by previous studies (e.g., Zhai et al., 2020) and are as follows: (TS=("scien* literacy" OR "scien* knowledge" OR "science learning" OR "science education" OR "science activit*" OR "scientific thinking" OR "scientific reasoning" OR "scien* inquiry" OR "scien* process" OR "scien* practice" OR "natural science*" OR "earth science*" OR "life science*" OR "physical science*" OR "space science*" OR astronomy OR biology OR chemistry OR geography OR physics)) AND TS=("informal learning" OR "non-formal learning" OR "informal learning environment*" OR "home learning" OR "home tutoring" OR "home learning environment*" OR "online environment*" OR "extracurricular learning" OR "out-of-school learning" OR "home-based" OR "after school" OR "science center" OR museum OR aquarium OR zoo OR "social media" OR "mass media" OR newspaper OR "TV program*" OR Internet). The asterisk (*) serves as a wildcard, allowing for variations of a given term (e.g., scien*: science, scientific). This search string yielded 14,568 hits.

Inclusion and Exclusion Criteria

We established clear inclusion and exclusion criteria to determine whether any of the studies should be included. A preliminary set of criteria were developed to ensure the quality and scope of our review. We included only articles indexed in the SCI and SSCI databases, which are recognized standards for research quality (Chou et al., 2013). This approach ensures that our dataset comprises high-quality research content, reducing the initial pool to 9,377 publications. Next, we excluded document types such as editorial materials, meeting abstracts, and letters, retaining only articles, early access articles, review articles, and book chapters, further reducing the pool to 9,148 documents. Finally, we restricted our time frame to articles published from 2009 onward. This threshold was chosen to document developments since the release of the comprehensive report on science learning in informal environments by the National Research Council in 2009. After applying these filters, our final dataset comprised 7,006 documents for further analysis.

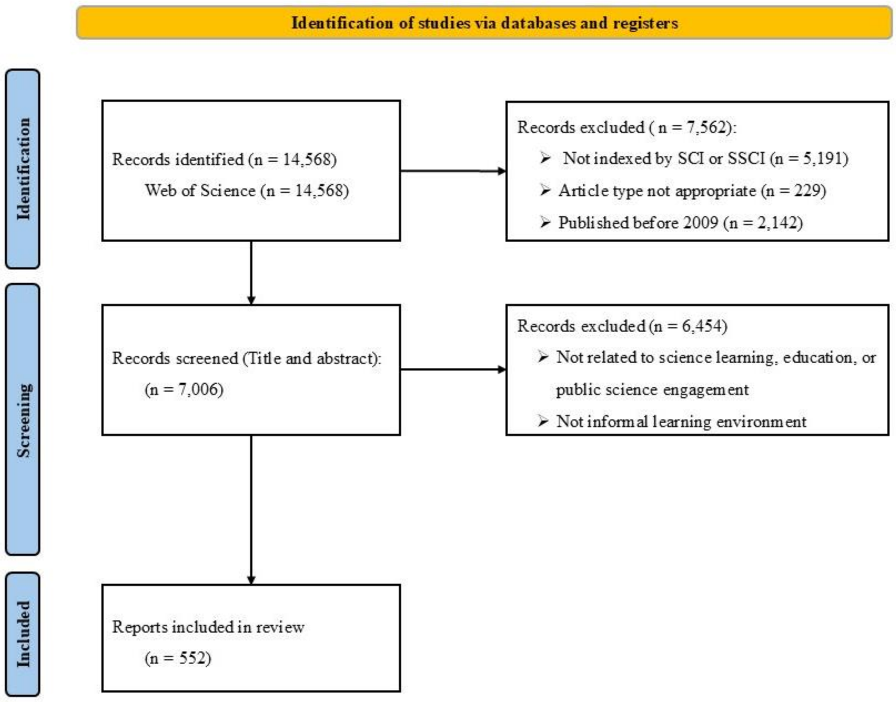
Using the refined dataset, we established two additional criteria to exclude irrelevant documents from our investigation: 1) the study must relate to science learning or teaching, science education, or public science engagement; and 2) the study must occur within an informal learning environment. Given the broad focus of this study, we adopted an expansive operational definition of science learning. This definition includes studies that involve the reception and transmission of science-related knowledge and information, encompassing conventional forms of learning such as meaning-making in collaborative science activities, as well as studies examining how scientific information is communicated to the public. The informal environments include designed settings such as museums, everyday family conversations, and online websites or social media platforms. Studies were excluded if they did not specify a particular context, even if they pertained to science learning. For example, although Cheng et al. (2021) clearly addressed the topic of science learning, it was a correlational study with no clear specification of context. Therefore, it was excluded. Similarly, a study was excluded if its main focus was not on informal science learning, even if some aspects of the study pertained to this topic. For instance, Ma et al. (2023) was excluded because although it did touch upon after-school science learning, the primary focus of the study was on the effect of study load on science achievement rather than informal science learning *per se*. Additionally, we included studies conducted in school settings if they related to science outreach programs or other after-school activities. However, studies were excluded if they involved formal classroom instruction, including those that integrated informal elements into the teaching process. An example was Wang et al. (2017). Although this study discussed the use of the social media platform WeChat in teaching biochemistry and cellular biology, the study's primary focus was on employing WeChat to support teachers' instructional goals within a formal course. Therefore, it was excluded from our dataset.

Title and Abstract Screening

Using the established criteria, we first randomly selected 500 articles from the pool of 7,006 documents. The first and second authors independently read the titles and abstracts of these articles to determine their inclusion or exclusion. Upon completing the screening, the two researchers compared their results and reached a high agreement of 98%, with any discrepancies resolved through discussion. This process resulted in the inclusion of 205 documents. Afterwards, the remaining 6,506 documents were roughly equally divided between the two researchers for separate screening. This process resulted in 347 additional articles meeting our inclusion criteria. Along with the 205 articles retained in our first round of screening, a total of 552 documents were retained in our final dataset. As the goal of our study was to extract underlying research topics or themes rather than code detailed information from the individual studies, full-text screening was deemed unnecessary and therefore not implemented. We provide a complete workflow documenting all the previous steps in Figure 2.

Figure 2

PRISMA Workflow



Data Analysis

For data analysis, we used structural topic modeling (STM) to uncover the recurring themes within the informal science learning literature. Below we describe in detail the conceptual foundations of STM, the analytic steps undertaken, and the software packages used for analysis.

A Conceptual Introduction to Structural Topic Modeling

Topic modeling is a natural language processing technique that uses unsupervised machine learning algorithms to extract patterns from textual content. Its popularity has surged in recent years across various fields, such as linguistics, organizational studies, education, and psychology (Brookes & McEnery, 2019; Mertala et al., 2024; Schmiedel et al., 2019; Weston et al., 2023). It can be regarded as a type of content analysis, yet it provides a more scientific approach than traditional methods due to its superior reproducibility and scalability. It is especially suited for "uncovering broad subject-matter-based themes" (Weston et al., 2023, p. 1). The underlying idea of topic models is that words co-occurring in the same context tend to have similar meanings. These clusters of co-occurring words reveal insights into recurring themes or topics present in a collection of texts. In topic modeling parlance, a document is considered to include a variety of topics, and each topic is considered to include a variety of words that are more or less likely to occur together. In a topic model, documents are statistically represented by a probability distribution over a finite number of topics, while each topic is characterized by a probability distribution over a finite set of words. This is typically illustrated through two matrices: the document-topic matrix and the topic-word matrix. In these matrices, each row corresponds to either a document or a topic, each column represents a topic or word, and each cell indicates the probability that a particular topic is associated with a specific document or that a specific word is associated with a given topic. STM is considered a type of probabilistic model, meaning that instead of assigning each topic or word to only one category, it can actually assign each topic or word to multiple documents or topics with varying probabilities. An advantage of STM over alternative topic models is that it allows for the incorporation of covariates or document metadata into the model, hence the name “structural”. This allows researchers to use the covariates to predict “the per-document topic distributions (topic prevalence) and per-topic word distributions (topic content)” (Schmiedel et al. 2019, p. 945). In this study, we used article abstracts as the input text for STM because the abstracts can conveniently convey the gist or the central idea of a document.

Data Preprocessing

Several preparatory steps were undertaken to ensure that our data were in an appropriate format that makes it amenable to STM analysis. We converted the raw text corpus into a document-feature matrix using the *bibliometrix* package (Aria & Cuccurullo, 2017) in R (R Core Team, 2023). We also obtained basic descriptive statistics of the documents, such as the number of publications, the number of authors, and the various outlets of publication. After this, we inspected the matrix and found that three articles had missing abstracts (Dawson, 2017; Hodeau, 2014; Stierand et al., 2012). We manually reviewed these studies and found that either the abstract was missing or a conventional abstract was not available. Therefore, we used the article titles as convenient substitutes, as both abstracts and article titles have been used in previous studies for STM analysis (e.g., Mertala et al., 2024).

Then the converted data was preprocessed using the *quanteda* package (Benoit et al., 2018). The text data were transformed to individual tokens (i.e., a single meaningful unit of data that may represent various elements such as words, punctuation marks, or numbers). Based on recommendations from the natural language processing literature (e.g., Grimmer et al., 2022), we performed a series of data cleaning steps, including converting all characters to lower case, eliminating numbers and punctuation marks, removing common English stop words (words that carry no substantive meaning and are primarily for functional purposes, such as articles and prepositions). The list of common English stop words was based on the Snowball project and can be freely accessed online (https://snowballstem.org/projects.html). Stemming was also performed to reduce different variations of words to their base or root forms (e.g., reducing "inquiries" to the root form "inquiri"). The stemming was implemented using Martin Porter's stemming algorithm for the English language and the C libstemmer library generated by Snowball (https://snowballstem.org/). Tokens were also tabulated by frequency, and low-frequency words were removed. In this study, a token must appear at least 10 times and in a minimum of two documents to be included. Eliminating low-frequency words helps simplify the complexity of the text data and is a common practice in natural language processing.

After completing these steps, we conducted a further inspection of the dataset. Upon closer examination, we discovered that some messiness remained. These included certain number words (e.g., "four"), punctuation marks (e.g., "="), additional function words (e.g., "whether"), and terms frequently used in scientific discourse that do not contribute to revealing underlying themes (e.g., "reveal" or "suggest"). We manually compiled a list of such words (see Supplementary Materials Table S2) and excluded them from our dataset. We chose to eliminate these words because they lacked substantive meaning and were not useful for uncovering themes or topics within the textual data. Additionally, eliminating them could further reduce the complexity of the text data. Once all preprocessing steps were completed, the final dataset was prepared for formal analysis.

Extracting Topics Using Structural Topic Modeling

A key step in STM is determining the number of topics to be extracted from the corpus. To identify an appropriate number of topics, we followed the recommendation of Weston et al. (2023) and relied on the following model fit statistics, including exclusivity, semantic coherence, residual, and variational lower bound. *Exclusivity* refers to the extent to which topic words are exclusive to a single topic rather than with multiple topics. *Semantic coherence* represents how frequently the most probable words in a given topic appear together and have been shown to correspond with human judgements of topic quality or logical consistency (Mimno et al., 2011). *Residual* is used to quantify the residual dispersion of a given topic model solution (Taddy, 2012). Variational lower bound is a statistical metric used to determine model convergence for a certain solution. Ideal topic solutions should have fewer residuals and higher values in exclusivity, semantic coherence, and variational lower bound (Weston et al., 2023).

However, more topics tend to increase fit but diminish coherence. Therefore, one must balance the tradeoff between statistical fit and interpretability in selecting models. In many cases, multiple model solutions could be plausible and no one definitive solution can be applied across all applications of a corpus. This is an important observation in text analysis and natural language processing, where there are often no "true" values to target (Grimmer et al., 2022). Instead, one must supplement text analysis with human judgement to make an informed decision.

In this study, we used the *stm* package (Roberts et al., 2019) to fit models ranging from five to 25 topics, following common heuristics used in previous research (e.g., Mertala et al., 2024). The four model fit statistics were calculated to choose candidate models for consideration. Once the candidate models were determined, we inspected the top 10 most probable topic words to determine whether the models made any substantive sense and whether the words could form any meaningful theme. In exploring the most probable words associated with each topic, we used the FREX metric (Airoldi & Bischof, 2016; Bischof & Airoldi, 2012). FREX weights topic words based on their overall frequency and exclusivity to a specific topic, thereby striking a good balance between these two factors.

To enhance objectivity in the human judgment process, we adapted and expanded the rating criteria established by Mertala et al. (2024) and developed a five-point scale. This scale evaluates various topic solutions across three dimensions: (1) topic coherence: the degree to which the topic keywords are homogeneous and the extracted topics are meaningful and interpretable; (2) topic distinctiveness: the extent to which different topics represent distinct themes with minimal overlap; and (3) theme coverage: whether the major themes have been adequately identified and if any significant themes are missing. A five-point rating scale was used, with a score of 1 representing the lowest coherence, distinctiveness, and coverage, and 5 the highest. The complete rating scale and its level descriptors can be found in Supplementary Materials Table S3. The first and second authors, who were familiar with the textual content of the corpus, independently rated the topic solutions. Then scores were averaged across raters and aggregated across dimensions to determine the final topic solution. The solution with the highest overall score was selected for final analysis. The FREX topic words from this model were subsequently interpreted, and a verbal label was assigned to facilitate understanding. After establishing this final topic solution, we also incorporated publication time as a covariate in the full STM model (akin to a regression model) and plotted the evolution of the topics over time. This allowed us to delve into the temporal dynamics of the extracted research topics.

Results

Description of the Dataset

Our dataset included 552 documents from 240 different sources involving a total of 1,824 authors. Overall, there was an upward trend in the number of publications per year from 2009 to 2024, with some slight fluctuations in between (Figure 3). The publication number was lowest at 13 in 2009 and peaked at 57 in 2022. Afterwards, the number decreased slightly in 2023 and 2024. The distribution of the publications was quite sparse, with the documents published in as many as 240 sources and each source accounting for less than 10% of the total number. *Journal of Chemical Education* was the most popular outlet with 37 publications in total, and many sources published only one article. We presented the top 10 outlets for easy reference in **Table 1**.

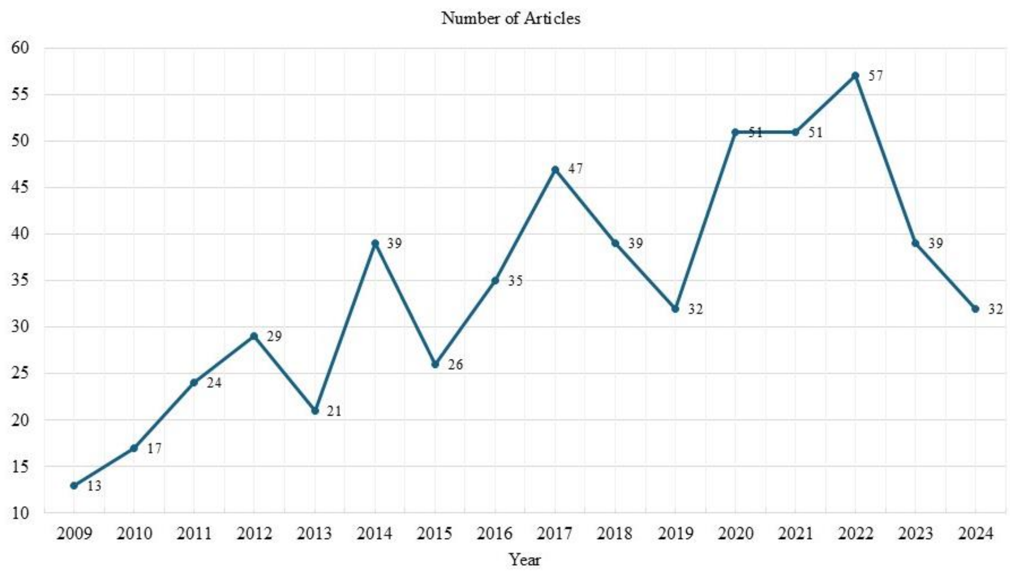
Table 1

Top Ten Outlets in Our Dataset

Rank	Sources	No. of Articles	Percent
1	*Journal of Chemical Education*	37	6.70%
2	*Journal of Research in Science Teaching*	28	5.07%
3	*Science Education*	23	4.17%
4	*Journal of Science Education and Technology*	20	3.62%
5	*Public Understanding of Science*	20	3.62%
6	*International Journal of Science Education*	17	3.08%
7	*Research in Science Education*	16	2.90%
8	*Frontiers in Psychology*	14	2.54%
9	*Plos One*	10	1.81%
10	*Sustainability*	10	1.81%

Figure 3

Number of Publications Per Year from 2009 to 2024



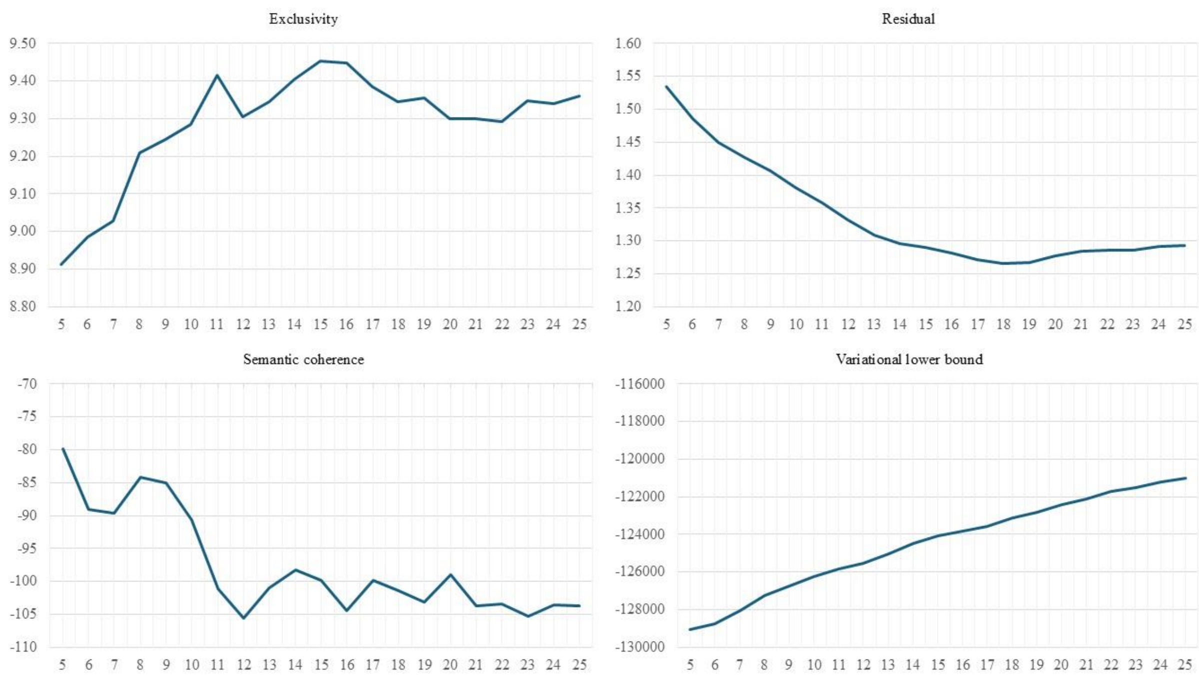
Model Fit Statistics for Different Topic Solutions

The fit statistics for models with five to 25 topics are plotted in Figure 4 (see

Supplementary Materials Table S4 for the exact statistics). As the number of topics increased, exclusivity and variational lower bound values increased, whereas semantic coherence and residual values decreased. Exclusivity increased steadily from five to 10 topics and remained relatively high between 10 and 25 topics. Semantic coherence values were highest between five and 10 topics and decreased sharply as the number of topics increased further. Residual and lower variational bound values decreased and increased systematically as the number of topics increased. Based on these statistics, it seemed reasonable to extract anywhere between 10 and 15 topics, as these solutions provided a good balance between solution quality and statistical fit. Furthermore, as the semantic coherence values (indicator of topic quality) were quite high at eight and nine topics, these were also selected as candidate models despite their relatively low statistical fit (i.e., residual and variational lower bound). Therefore, eight candidate models (eight to 15 topics) were preliminarily retained for further screening.

Figure 4

Model Fit Statistics as a Function of the Number of Topics



Human Ratings for Selected Topic Solutions

Table 2 presents the human rating results for the eight candidate models. Overall, the models with eight and nine topics received the highest ratings, followed closely by those with ten and thirteen topics. The remaining models garnered lower ratings. Since the eight- and nine-topic models were tied with a score of 14.5, the first and second researchers of this article reviewed the topic words for both models and discussed their viability. After deliberation, it was determined that the nine-topic solution should be retained, as the first eight topics in this model appeared to overlap with those in the eight-topic solution, while the additional ninth topic offered meaningful new insights. The top words for this nine-topic solution was discussed in the next section, while the topic words for the other topic solutions were also provided for reference in the Supplementary Materials Table S5.

Table 2

Average and Overall Rating Results of Selected Topic Solutions

Topic	Topic coherence	Topic distinctiveness	Theme coverage	Overall
8 topics	5	5	4.5	14.5
9 topics	4.5	5	5	14.5
10 topics	4	5	4.5	13.5
11 topics	4.5	4	4.5	13
12 topics	4.5	4.5	4.5	13.5
13 topics	4	4.5	4.5	13
14 topics	3.5	4	4	11.5
15 topics	3	4	4.5	11.5

Topics Extracted from the STM Analysis

Proceeding from the nine-topic solution, we tabulated the top 10 keywords assigned to each of the nine topics. Table 3 presents the FREX words for each topic. Inspection of the top words suggested that they were mostly homogeneous and could be easily interpreted to form a coherent theme. Based on the semantic content of the keywords, we assigned verbal labels to the nine topics to facilitate easy interpretation. As can be seen, a variety of prominent themes or topics were uncovered, including online science learning, family science engagement, chemistry education and science outreach, STEM motivation and stereotypes, to name just a few. We also calculated the overall topic prevalence or proportions. As can be seen, the most prevalent topic was news media and science, whereas the least prevalent topic was STEM motivation and gender stereotypes.

Table 3

Topic 10 Words for the Nine Topics Extracted from STM Analysis

Topic	Top words	Label	Proportion
1	news, newspap, blog, climat, expert, coverag, public, epistem, scientif, mass	News media and science	14.65%
2	parent, children, famili, youth, ident, adult, center, cultur, convers, talk	Family science engagement	12.33%
3	student, inquiri, motiv, out-of-school, mobil, lab, physic, control, school, experiment	After-school science	13.91%
4	search, video, internet, youtub, web, onlin, access, websit, tool, read	Online science learning	7.93%
5	museum, game, histori, pedagog, integr, concept, natur, interact, approach, model	Museum learning	11.81%
6	chemistri, program, laboratori, outreach, virtual, univers, teacher, non-form, educ, plan	Chemistry education and science outreach	13.86%
7	conserv, aquarium, plant, anim, zoo, garden, visitor, exhibit, biodivers, speci	Science centers	8.52%
8	twitter, covid-19, social, pandem, media, messag, health, vaccin, misinform, polici	Social media and science	9.76%
9	stem, interest, particip, report, femal, math, stereotyp, gender, question, competit	STEM motivation and gender stereotypes	7.24%

Note. The top words were reduced root forms of the original words extracted using the FREX algorithm. Proportion refers to the expected proportions of each topic extracted from the model results.

Evolution of the Extracted Topics over Time

To illustrate the evolution of the extracted topics, we present the statistical test results in Table 4 and plot the changes in proportions over time in Figure 5. It is important to note that the significance tests should be interpreted with caution due to the limited variability in publication time, which consists of only 16 discrete values from 2009 to 2024. This restricted variability may diminish the statistical power of the tests. Therefore, it is advisable to combine the results of the statistical tests with the visual representations of their evolution. Overall, it seems that most topics (e.g., topics 3, 4, 5, 6, 7) were quite stable as the topic prevalence remained relatively unchanged. However, some topics appeared to show evidence of increase or decrease. For example, both topic 8 (social media and science) (β = .006, p = .006) and topic 9 (STEM motivation and gender stereotypes) (β = .003, p = .033) showed statistically significant increase over time. Topic 2 (family science engagement) also seemed to be slightly increasing, although the increase was not statistically significant (β = .003, p = .234). In contrast, topic 1 (news media and science) seemed to be slightly losing momentum (β = -.005, p = .101).

Table 4

STM Topic Prevalence as a Function of Publication Time

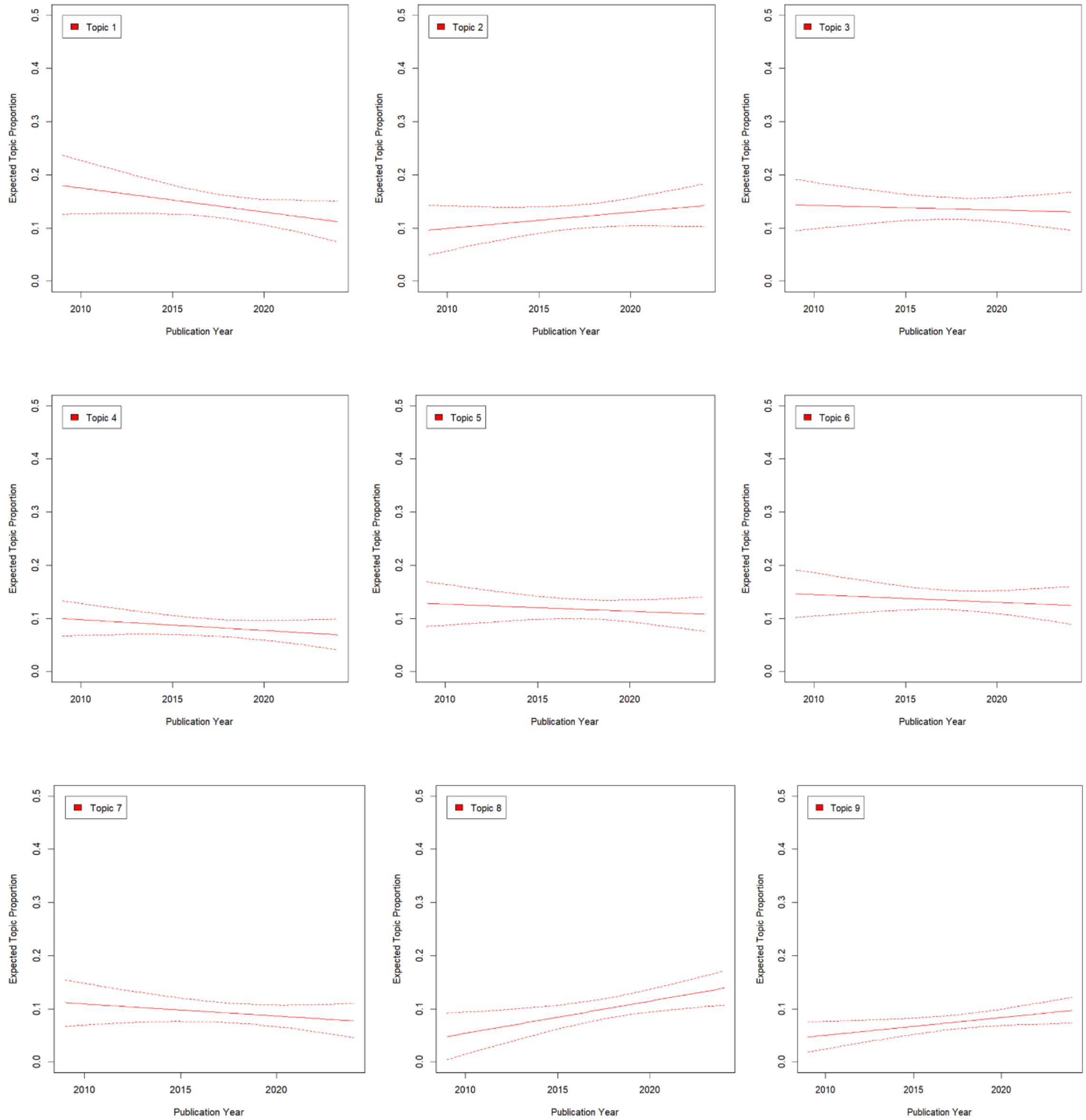
	β	Std. Error	t	p value
Topic 1	-.005	.003	-1.642	.101
Topic 2	.003	.003	1.190	.234
Topic 3	-.001	.002	-.333	.739
Topic 4	-.002	.002	-1.090	.276
Topic 5	-.001	.002	-.604	.546
Topic 6	-.002	.002	-.637	.525
Topic 7	-.002	.002	-1.012	.312

Topic 8	.006	.002	2.757	.006
Topic 9	.003	.002	2.133	.033

Note. In the STM model, the independent variable was publication time and the dependent variable was the expected topic proportion.

Figure 5

Evolution of the Extracted Topics Over Time



Note. In the figure, the horizontal axis represents the time of publication, while the vertical axis represents topic prevalence or proportion. Topic 1: news media and science; Topic 2:

family science engagement; Topic 3: after-school science; Topic 4: online science learning; Topic 5: museum learning; Topic 6: chemistry education and science outreach; Topic 7: science centers; Topic 8: social media and science; Topic 9: STEM motivation and gender stereotypes.

Discussion

The present study reviewed the literature on science learning in informal environments during 2009 and 2024 using structural topic modeling. The results revealed that the field has shown a continual increase over the years. We identified nine commonly studied topics and their evolution over time. Below, we discuss our findings in relation to the contexts of science learning and strands of science learning framework outlined by the National Research Council (2009). We also discuss how their proposed ecological model of learning bears on our findings and the implications they have for advancing science learning and education into the future.

Interpretations and Implications of the Findings

The present review identified nine research topics in the informal science literature from 2009 to 2024, encompassing a wide range of science-related aspects that align with the six strands of science learning framework (National Research Council, 2009). For instance, the topic of "STEM motivation and gender stereotypes" is closely linked to Strand 1, which highlights the motivational aspects of learning, as well as to Strand 6, as the topic addresses societal hurdles that may hinder the development of science-related identities. Strand 5, which highlights the collaborative and interactive nature of science learning and engagement in scientific practice is also evident in several topics, such as "news media and science", "family science engagement", and "social media and science". While the development of knowledge and scientific reasoning was not explicitly highlighted in these topics, they nonetheless serve as the foundational assumptions that guide research across many of the extracted themes.

A notable finding from the extracted topics is the great diversity of informal environments explored in the existing research. For example, several studies examined the roles of museums (topic 5), science centers (topic 7), family members (topic 2), and online platforms (topics 4 and 8) in facilitating science learning, while one topic specifically addressed students' after-school science learning (topic 3). Together, these studies suggested that informal science learning could occur in a variety of everyday contexts. Unlike formal education, which required dedicated classroom time, the topics highlighted how informal science could complement school education by utilizing diverse platforms. What is particularly evident is the heavy presence of online and social media platforms in science learning as well as the prominent role museums play. Indeed, of the nine identified research topics, the topic of "news media and science" constituted the largest proportion (14.65%), and the topic of "social media and science" evidenced the most pronounced increase. As argued by Maier et al. (2014), one of the benefits of media use for informal science learning is that it is neither bound to specific educational environments nor constrained to certain segments of society. In an increasingly digitized world, both social media and traditional media can make science accessible anytime and anywhere, reaching learners across physical and temporal boundaries. Similarly, museums have increasingly emerged as central institutions for science learning, catering to individuals of all ages and backgrounds. Unlike media, which is usually not designed to promote science learning, museums tend to be designed environments rich in scientific content and experiences, fostering a dynamic social-communicative context (Haden, 2010). Increasingly powered by modern technologies, exposure to science is integrated into immersive and participatory visiting experiences, whether through parent-child engagement (e.g., Callanan et al., 2017; Sobel, 2023) or as an extension of school-based learning in partnership with educators (e.g., Kisiel, 2014).

Together, these findings underscore the abundance of opportunities for informal science learning.

Of the nine topics identified, most appeared to be relatively stable over time; however, three topics appeared to be gaining traction and may warrant greater attention. The first was the role of social media in science learning. Judging by the top words in this topic, the surge of interest in this area was likely fueled by the COVID-19 pandemic and the concomitant political repercussions, which sparked intense discussions about the virus, vaccines, scientific misinformation, and other health-related issues (Loomba et al., 2021; Scheufele et al., 2021). The second topic was STEM motivation and gender stereotypes. Although this niche topic accounted for the smallest proportion among the nine, it showed signs of increasing momentum. Unlike science education in schools, which tends to follow a pre-established curriculum structure, informal science learning is often spontaneous and driven by individual interests in particular subject areas or topics (National Research Council, 2009). Therefore, developing the interest in and motivation to learn science is a prominent aspect of informal science learning. Regarding gender stereotypes, it is well-known that women have been historically underrepresented in the STEM workforce and continue to face significant barriers in academia and industry (Balducci et al., 2024; Moss-Racusin et al., 2021). This raises important questions of fairness and educational equity, with implications not only for nurturing the next generation of scientists but also for countries' future economic and technological development (Moss-Racusin et al., 2021). A crucial step moving forward is to reduce the barriers that prevent women and other minority groups from developing scientific knowledge and identities, as well as engaging in the scientific enterprise. The third topic that also appeared to be slightly increasing was family science engagement. Family is the nucleus of society, and people spend much time with family members outside of school. Research in the developmental sciences has long documented the importance of family influences on children's cognitive development (e.g., Bronfenbrenner, 1979; Lugo-Gil & Tamis-LeMonda, 2008). With the renewed recognition that even very young children are capable of scientific thinking (e.g., Koeber & Osterhaus, 2019), we expect more research on the role the family has on children's science learning. Several recent studies have already begun to address the question of how science learning is mediated through everyday parent-child interactions (e.g., Callanan et al., 2017; Sobel, 2023).

Another observation that emerged from our findings is that, among all the domains of science, chemistry education distinguished itself as a prominent field, while other areas of science did not stand out as independent fields of inquiry. This highlights the significant interest in chemistry compared to other scientific domains. Although our analysis did not allow us to explore the reasons for this disparity, we hypothesize that the specialized focus of the *Journal of Chemical Education* may partly explain this trend by providing a dedicated outlet for publishing research in chemistry. Nonetheless, this does not imply that other scientific domains are neglected. Furthermore, our findings also revealed evidence of cross-domain pollination, particularly in the topic of STEM motivation and gender stereotypes. Science functions both as an independent field of inquiry and as a crucial component of the broader STEM disciplines (Martín-Páez et al., 2019). There are significant connections between science and its neighboring fields, and STEM education often emphasizes the integration of these disciplines as a cohesive whole. Given this strong interrelationship, future research could explore synergies and points of connection between science and other STEM areas. Many issues, such as gender stereotypes, are not confined to science alone but are prevalent across all STEM disciplines (Balducci et al., 2024). Understanding how these stereotypes originate and how to counteract them throughout the STEM fields is an essential step forward for researchers and practitioners alike.

A notable aspect of the informal science literature is the significant absence of academic performance metrics, coupled with an emphasis on non-academic elements such as motivation. This observation is quite fitting, as informal science learning serves complementary roles to formal science education in schools (National Research Council, 2009). While formal education primarily focuses on imparting scientific knowledge and skills, informal science learning adopts a broader perspective. It nurtures interest and motivation to engage with science, fosters the development of scientific identities, and encourages participation in scientific practices, extending beyond the mere acquisition of disciplinary knowledge and skills (National Research Council, 2009). Given that much scientific understanding occurs outside of traditional classroom settings, it is essential that more research be dedicated to exploring aspects of science learning that are often overlooked or inadequately addressed in formal educational contexts. A key implication of this finding for researchers and practitioners is that assessments or evaluations of science learning in informal environments must accurately reflect the realities of these settings and the nature of science learning within them. Relying exclusively on academic performance metrics, such as standardized test scores, is inadequate. Instead, it is essential to develop context-appropriate assessments or evaluation methods or adapt existing ones to align with the specific objectives of each learning environment.

Throughout the various topics discussed, context emerged as a recurring theme, which was to be expected. However, the ecological model of learning emphasizes the importance of people and culture beyond mere locations (National Research Council, 2009). The references to personal characteristics were noticeably less pronounced in the extracted topics, while mentions of culture were even more subdued. Despite this, all three elements—context, people, and culture—must remain central themes in research on informal science learning. Individual characteristics have long been recognized as critical factors in science learning. These include prior knowledge, the ability to differentiate between hypotheses and evidence, the use of metacognitive strategies, and an epistemological understanding of science, among others (e.g., Zimmerman, 2007; Kuhn et al., 2008). Future studies should adequately address these issues when investigating science learning in informal contexts. Additionally, since scientific practices are culturally embedded, the activities and approaches learners adopt for science sense-making are unlikely to be universal across different cultural settings (Medin & Bang, 2014). Therefore, researchers must study science learning *in situ* and always bear in mind the importance of cultural specificity and appropriateness.

Limitations and Future Directions

Although we believe this review makes an important and timely contribution to the extant literature, we also note some limitations. First, we relied on only one database: the Web of Science database. This means that some documents were inevitably left out in the search process, potentially biasing our findings and conclusions. Nonetheless, we believe this is unlikely to detract from our findings in fundamental ways. In making the intentional decision to rely on the WOS core database, specifically the SCI- and SSCI-indexed databases, we wanted to ensure all our documents included were of high quality. On the other hand, topic modeling as practiced in this study primarily relied on word co-occurrences to derive topics. Given the huge corpus we used for analysis, the topics we uncovered are likely quite stable and may not be easily influenced by the addition or omission of a few individual studies.

Second, although topic modeling provides an informative means to uncover broad themes in the literature, it cannot provide the nuanced insights that are only possible through careful human coding. Therefore, we could not provide specifics about important questions such as the frequency with which different contexts or different aspects of science were addressed. Future work is needed to supplement this review with a traditional systematic review to uncover such specific information.

Additionally, we acknowledge the exploratory nature of our analytic approach and the inherent subjectivity that accompanies it. As briefly mentioned, the topic modeling process involves several steps that are often open to debate, such as the number of topics to extract and their interpretation. Ongoing discussions surround many of these issues, and in many cases, researchers have little choice but to rely on their best judgment, making definitive conclusions impossible. To address this, we have meticulously detailed our analytic steps and provided alternative topic solutions in the Supplementary Materials. Additionally, the rating scale we developed to complement model fit statistics reflects our commitment to transparency and reproducibility, allowing readers to assess the validity of our solutions and the interpretability of our results. The nine-topic solution represents our best assessment of the themes within the informal science literature and appeared meaningful in capturing the studied themes. While we believe in its utility and significance, we acknowledge that other solutions could also serve as plausible models. By openly acknowledging these limitations, we emphasize that our findings should not be regarded as the sole or definitive interpretation, and we welcome further debate and discussion. In summary, we hope our efforts will serve as valuable navigational aids for both junior and senior researchers seeking to understand and advance research in this field.

Conclusion

The present study systematically reviewed the literature on science learning in informal environments during 2009 and 2024. Using a topic modeling approach, we identified nine major research topics in the literature and documented their evolution over time. These topics highlighted a range of informal contexts related to various aspects of science learning. Among them, the most prevalent topic was about how science was portrayed or communicated in the news, whereas the least prevalent topic pertained to STEM motivation and gender stereotypes. Overall, most topics seemed to be relatively stable over time, but some topics showed an upward trend, including the role of social media in science, STEM motivation and gender stereotypes, and family science engagement. The study documents important developments in the informal science literature since 2009. By taking stock of the extant research, the findings not only serve as important navigational aids for researchers but also have significant implications for designing research programs based on underexplored or emerging topics in the field.

References

Airoldi, E. M., & Bischof, J. M. (2016). Improving and evaluating topic models and other models of text. *Journal of the American Statistical Association*, *111*(516), 1381–1403. https://doi.org/10.1080/01621459.2015.1051182

Aria, M., & Cuccurullo, C. (2017). *bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975, https://doi.org/10.1016/j.joi.2017.08.007

Bae, J., Shavlik, M., Shatrowsky, C. E., Haden, C. A., & Booth, A. E. (2023). Predicting grade school scientific literacy from aspects of the early home science environment. *Frontiers in Psychology*, *14*, Article 1113196. https://doi.org/10.3389/fpsyg.2023.1113196

Balducci, M., Larose, M. P., Stoet, G., & Geary, D. C. (2024). The gender-equality paradox in intraindividual academic strengths: A cross-temporal analysis. *Psychological Science*, Article 09567976241271330. https://doi.org/10.1177/09567976241271330

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). *quanteda*: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774–774. https://doi.org/10.21105/joss.00774

Bischof, J., & Airoldi, E. M. (2012). Summarizing topical content with word frequency and exclusivity. In *Proceedings of the 29th international conference on machine learning (icml-12)* (pp. 201-208). Omnipress, New York.

Booth, A. E., Shavlik, M., & Haden, C. A. (2020). Parents' causal talk: Links to children's causal stance and emerging scientific literacy. *Developmental Psychology, 56*(11), 2055–2064. https://doi.org/10.1037/dev0001108

Brookes, G., & McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, *21*(1), 3–21.

https://doi.org/10.1177/1461445618814032

Callanan, M. A., Castañeda, C. L., Luce, M. R., & Martin, J. L. (2017). Family science talk in museums: Predicting children's engagement from variations in talk and activity. *Child Development*, *88*(5), 1492–1504. https://doi.org/10.1111/cdev.12886

Cheng, C. H., Bråten, I., Yang, F. Y., & Brandmo, C. (2021). Investigating structural relationships among upper-secondary school students' beliefs about knowledge, justification for knowing, and Internet-specific justification in the domain of science. *Journal of Research in Science Teaching*, *58*(7), 980–1009. https://doi.org/10.1002/tea.21689

Chou, C. P., Lin, H. F., & Chiu, Y. J. (2013). The impact of SSCI and SCI on Taiwan's academy: An outcry for fair play. *Asia Pacific Education Review*, *14*, 23–31. https://doi.org/10.1007/s12564-013-9245-1

Dawson, E. (2017). Social justice and out-of-school science learning: Exploring equity in science television, science clubs and maker spaces. *Science Education*, *101*(4), Article 539. https://doi.org/10.1002/sce.21288

Geiger, N., Swim, J. K., Fraser, J., & Flinner, K. (2017). Catalyzing public engagement with climate change through informal science learning centers. *Science Communication*, *39*(2), 221–249. https://doi.org/10.1177/1075547017697980

Goff, E. E., Mulvey, K. L., Irvin, M. J., & Hartstone-Rose, A. (2020). The effects of prior informal science and math experiences on undergraduate STEM identity. *Research in Science & Technological Education*, *38*(3), 272–288. https://doi.org/10.1080/02635143.2019.1627307

Greenhow, C., Gibbins, T., & Menzer, M. M. (2015). Re-thinking scientific literacy out-of-school: Arguing science issues in a niche Facebook application. *Computers in Human Behavior*, *53*, 593–604. https://doi.org/10.1016/j.chb.2015.06.031

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Habig, B., Gupta, P., Levine, B., & Adams, J. (2020). An informal science education program's impact on STEM major and STEM career outcomes. *Research in Science Education*, *50*(3), 1051–1074. https://doi.org/10.1007/s11165-018-9722-y

Haden, C. A. (2010). Talking about science in museums. *Child Development Perspectives*, *4*(1), 62–67. https://doi.org/10.1111/j.1750-8606.2009.00119.x

Hetherington, E., Eggers, M., Wamoyi, J., Hatfield, J., Manyama, M., Kutz, S., & Bastien, S. (2017). Participatory science and innovation for improved sanitation and hygiene: Process and outcome evaluation of project SHINE, a school-based intervention in Rural Tanzania. *BMC Public Health*, *17*, 1–15. https://doi.org/10.1186/s12889-017-4100-7

Hodeau, J. L. (2014, August). The Experience of "Voyage dans le cristal," A travelling museum exhibition. In *IUCr-XXIII: International Congress of Crystallography*. http://dx.doi.org/10.1107/s2053273314099720

Junge, K., Schmerse, D., Lankes, E. M., Carstensen, C. H., & Steffensky, M. (2021). How the home learning environment contributes to children's early science knowledge—associations with parental characteristics and science-related activities. *Early Childhood Research Quarterly*, *56*, 294–305. https://doi.org/10.1016/j.ecresq.2021.04.004

Kelly, L. A. D., Luebke, J. F., Clayton, S., Saunders, C. D., Matiasek, J., & Grajal, A. (2014). Climate change attitudes of zoo and aquarium visitors: Implications for climate literacy education. *Journal of Geoscience Education*, *62*(3), 502–510. https://doi.org/10.5408/13-078.1

Kisiel, J. F. (2014). Clarifying the complexities of school–museum interactions: Perspectives from two communities. *Journal of Research in Science Teaching*, *51*(3), 342–367. https://doi.org/10.1002/tea.21129

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, *333*(6045), 971–975. https://doi.org/10.1126/science.1204528

Koerber, S., & Osterhaus, C. (2019). Individual differences in early scientific thinking: Assessment, cognitive influences, and their relevance for science learning. *Journal of Cognition and Development*, *20*(4), 510–533. https://doi.org/10.1080/15248372.2019.1620232

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking?. *Cognitive Development*, *23*(4), 435–451. https://doi.org/10.1016/j.cogdev.2008.09.006

Loomba, S., De Figueiredo, A., Piatek, S. J., De Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, *5*(3), 337–348. https://doi.org/10.1038/s41562-021-01056-1

Lugo-Gil, J., & Tamis-LeMonda, C. S. (2008). Family resources and parenting quality: Links to children's cognitive development across the first 3 years. *Child Development*, *79*(4), 1065–1085. https://doi.org/10.1111/j.1467-8624.2008.01176.x

Maier, M., Rothmund, T., Retzbach, A., Otto, L., & Besley, J. C. (2014). Informal learning through science media usage. *Educational Psychologist*, *49*(2), 86–103. https://doi.org/10.1080/00461520.2014.916215

Maiorca, C., Roberts, T., Jackson, C., Bush, S., Delaney, A., Mohr-Schroeder, M. J., & Soledad, S. Y. (2021). Informal learning environments and impact on interest in STEM careers. *International Journal of Science and Mathematics Education*, *19*, 45–

64. https://doi.org/10.1007/s10763-019-10038-9

Martín-Páez, T., Aguilera, D., Perales-Palacios, F. J., & Vílchez-González, J. M. (2019). What are we talking about when we talk about STEM education? A review of literature. *Science Education*, *103*(4), 799–822. https://doi.org/10.1002/sce.21522

Medin, D. L., & Bang, M. (2014). *Who's asking?: Native science, western science, and science education*. MIT Press.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262–272). Association for Computational Linguistics.

Moss-Racusin, C. A., Pietri, E. S., van der Toorn, J., & Ashburn-Nardo, L. (2021). Boosting the sustainable representation of women in STEM with evidence-based policy initiatives. *Policy Insights from the Behavioral and Brain Sciences*, *8*(1), 50–58. https://doi.org/10.1177/2372732220980092

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. Washington DC: National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington DC: National Academies Press.

Osborne, J., & Pimentel, D. (2023). Science education in an age of misinformation. *Science Education*, *107*(3), 553–571. https://doi.org/10.1002/sce.21790

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/bmj.n71

R Core Team. (2023). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available online at: https:// www.R-project.org/.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, *91*, 1–40. https://doi.org/10.18637/jss.v091.i02

Scheufele, D. A., Hoffman, A. J., Neeley, L., & Reid, C. M. (2021). Misinformation about science in the public sphere. *Proceedings of the National Academy of Sciences*, *118*(15), Article e2104068118. https://doi.org/10.1073/pnas.2104068118

Schiefer, J., Caspari, J., Moscoso, J. A., Catarino, A. I., Miranda Afonso, P., Golle, J., & Rebuschat, P. (2024). Science and Heritage Language Integrated Learning (SHLIL): Evidence of the effectiveness of an innovative science outreach program for migrant students. *Science Education*, *108*(4), 983–1014. https://doi.org/10.1002/sce.21860

Schiefer, J., Edelsbrunner, P. A., Bernholt, A., Kampa, N., & Nehring, A. (2022). Epistemic beliefs in science—a systematic integration of evidence from multiple studies. *Educational Psychology Review*, *34*(3), 1541–1575. https://doi.org/10.1007/s10648-022-09661-w

Schmiedel, T., Müller, O., & Vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, *22*(4), 941–968. https://doi.org/10.1177/1094428118773858

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

Schweingruber, H. A., & Fenichel, M. (Eds.). (2010). *Surrounded by science: Learning science in informal environments*. National Academies Press.

Sobel, D. M. (2023). Science, Technology, Engineering, and Mathematics (STEM) engagement from parent-child interaction in informal learning environments. *Current Directions in Psychological Science*, *32*(6), 454–461. https://doi.org/10.1177/09637214231190632

Stierand, K., Harder, T., Marek, T., Hilbig, M., Lemmen, C., & Rarey, M. (2012). The Internet as scientific knowledge base: Navigating the Chem-Bio space. *Molecular Informatics*, *31*(8), Article 543. https://doi.org/10.1002/minf.201200037

Taddy, M. (2012). On estimation and selection for topic models. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, PMLR 22*, 1184–1193.

Wang, J., Gao, F., Li, J., Zhang, J., Li, S., Xu, G. T., ... & Lu, L. (2017). The usability of WeChat as a mobile and interactive medium in student-centered medical teaching. *Biochemistry and Molecular Biology Education*, *45*(5), 421–425. https://doi.org/10.1002/bmb.21065

Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the number and labels of topics in topic modeling: A tutorial. *Advances in Methods and Practices in Psychological Science*, *6*(2), Article 25152459231160105. https://doi.org/10.1177/25152459231160105

Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: A systematic review. *Studies in Science Education*, *56*(1), 111–151. https://doi.org/10.1080/03057267.2020.1735757

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001