

Thematic Evolution in Natural Sciences Education Research: A Topic Modeling Analysis of Knowledge Domains and Emerging Trends (2000–2025)

Author Names

February 12, 2026

Abstract

Natural sciences education research spans diverse themes, yet systematic mapping of this thematic landscape remains limited. This study employs Latent Dirichlet Allocation (LDA) topic modeling to analyze 321 peer-reviewed documents from Scopus and Web of Science (2000–2025), identifying nine distinct thematic clusters. Model optimization via perplexity and coherence metrics selected $k = 9$ topics: Inquiry-Based Science Education (27.4%), Environmental Education (24.3%), STEM Integration (10.0%), Teacher Professional Development (9.0%), Technology-Enhanced Learning (6.2%), Assessment & Evaluation (6.2%), Equity & Social Justice (6.2%), Scientific Literacy (6.2%), and Conceptual Understanding (4.4%). Temporal analysis revealed Environmental Education’s dramatic growth post-2020 (exponential increase from 3 documents in 2021 to 41 in 2025), while Inquiry-Based Science Education remained stable. Geographic analysis demonstrated thematic specialization: the United States emphasized technology integration, Germany focused on environmental education, and Turkey concentrated on inquiry-based pedagogy. Keyword co-occurrence networks validated LDA clusters and identified bridging concepts. Knowledge gaps include underrepresentation of equity and social justice themes (6.2% despite growing societal urgency), limited longitudinal assessment research, and geographic concentration in North America, Europe, and East Asia. These findings inform research agendas, curriculum development priorities, and funding allocation strategies, while demonstrating NLP’s utility for large-scale education research synthesis.

Keywords: Topic Modeling, LDA, Natural Sciences Education, Thematic Analysis, Research Trends, Science Didactics, Knowledge Mapping

1 Introduction

Natural sciences education research addresses pedagogical challenges in chemistry, biology, physics, and integrated STEM contexts. As this field expands—evidenced by exponential publication growth since 2015 Hallinger and Kovačević, 2019—understanding its thematic structure becomes critical for researchers, curriculum developers, and policymakers. Traditional literature reviews, while valuable, suffer from scalability limitations and subjective selection bias. Computational text mining methods overcome these constraints by systematically analyzing large document corpora.

Latent Dirichlet Allocation (LDA) Blei et al., 2003, an unsupervised machine learning technique, detects latent thematic structures in text collections. LDA assumes documents comprise mixtures of topics, where topics are probability distributions over words. By analyzing

word co-occurrence patterns across documents, LDA infers underlying topics without requiring predefined categories. This approach has been successfully applied to educational research Hallinger and Kovačević, 2019, scientific literature analysis Li et al., 2019, and policy document examination Teo et al., 2014.

Despite LDA’s widespread adoption in other domains, natural sciences education research lacks comprehensive topic modeling studies spanning multiple decades. Existing bibliometric analyses Hallinger and Kovačević, 2019; Teo et al., 2014 have examined collaboration networks and citation patterns but have not systematically mapped thematic content or temporal evolution. This gap limits stakeholders’ ability to identify emerging research directions, allocate funding to underexplored areas, or align curriculum development with contemporary scholarship.

1.1 Research Context

Natural sciences education sits at the intersection of disciplinary content knowledge (chemistry, biology, physics), pedagogical theory (constructivism, inquiry-based learning), educational psychology (conceptual change, motivation), and educational technology (simulations, virtual laboratories). This multidisciplinary character generates diverse research themes that may shift over time in response to policy initiatives, technological innovations, and societal needs.

Understanding thematic evolution addresses several stakeholder needs. Researchers benefit from identifying saturated areas versus knowledge gaps, enabling strategic positioning of new investigations. Curriculum developers gain evidence about pedagogical approaches gaining empirical support. Policymakers require data on research trends to align funding priorities with field needs. Teacher educators need insights into professional development emphases to prepare pre-service educators effectively.

1.2 Research Questions

1. What are the main thematic clusters (topics) in natural sciences education research from 2000 to 2025?
2. How have these research themes evolved temporally, and which topics are emerging, growing, declining, or stable?
3. How does thematic focus vary geographically across countries and regions?
4. What conceptual bridges connect different thematic domains, as revealed by keyword co-occurrence analysis?
5. What knowledge gaps and underexplored areas warrant future research attention?

1.3 Significance and Contributions

This investigation makes four primary contributions. First, it provides the first large-scale LDA analysis of natural sciences education research spanning 25 years, establishing a thematic baseline for future studies. Second, methodologically, it demonstrates NLP’s applicability to specialized education domains, offering a replicable framework for other fields. Third, temporal analysis reveals thematic trajectories, distinguishing stable foundations from emerging trends. Fourth, geographic variation analysis illuminates how national contexts shape research priorities, informing international collaboration strategies and context-sensitive policy transfer.

2 Methodology

2.1 Data Collection

Documents were retrieved from Scopus and Web of Science using search strategies targeting natural sciences didactics, science education pedagogy, and STEM teaching methodologies. Searches covered Title, Abstract, and Keywords fields, restricting to peer-reviewed articles and reviews in English from January 2000 through December 2025. After merging via `bibliometrix::mergeDbSources()` Aria and Cuccurullo, 2017 and quality control, the final corpus comprised 321 unique documents.

2.2 Text Preprocessing

Text preprocessing transformed raw bibliographic data into analysis-ready format. The corpus combined Title, Abstract, and Author Keywords fields for each document, providing thematic information without requiring full-text access (which introduces copyright and access barriers).

Preprocessing steps included:

1. **Tokenization:** Text split into individual words using whitespace and punctuation delimiters.
2. **Lowercase conversion:** All characters standardized to lowercase to ensure case-insensitive matching.
3. **Stopword removal:** Common words lacking semantic content (e.g., “the,” “and,” “of”) removed using an extended stopwords list combining standard English stopwords with domain-specific terms (e.g., “study,” “research,” “paper”).
4. **Lemmatization:** Words reduced to base forms using `udpipe` Straka and Straková, 2017 with English language models (e.g., “learning” → “learn,” “students” → “student”).
5. **Bigram detection:** Statistically significant two-word phrases identified using point-wise mutual information (PMI) thresholds, capturing compound concepts (e.g., “inquiry_-based,” “stem_education,” “virtual_laboratory”).
6. **Frequency filtering:** Terms appearing in fewer than 5 documents (too rare) or more than 80% of documents (too common) excluded to focus on discriminative vocabulary.

Software implementation utilized R packages `tidytext` Silge and Robinson, 2016, `tm` Feinerer et al., 2008, and `udpipe` Straka and Straková, 2017.

2.3 Topic Modeling: Latent Dirichlet Allocation

2.3.1 Model Selection

LDA requires specifying the number of topics k a priori. We employed grid search over $k \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, evaluating each model using two metrics:

- **Perplexity:** Measures model fit to held-out documents; lower values indicate better predictive performance. Perplexity is defined as:

$$\text{Perplexity}(D_{\text{test}}) = \exp \left\{ -\frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right\} \quad (1)$$

where D_{test} is the test set, M is the number of test documents, w_d are words in document d , and N_d is word count in document d .

- **Coherence:** Quantifies semantic interpretability by measuring word co-occurrence within topics; higher values indicate more coherent topics. We used C_v coherence Röder et al., 2015:

$$C_v = \frac{1}{T} \sum_{t=1}^T \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{NPMI}(w_i^{(t)}, w_j^{(t)}) \quad (2)$$

where T is the number of topics, N is the number of top words per topic, and NPMI is normalized pointwise mutual information.

Model selection balanced these metrics: perplexity decreased monotonically with k , while coherence peaked at $k = 9$ (Figure 1). Human interpretability assessment confirmed $k = 9$ produced distinct, semantically meaningful topics without excessive fragmentation. Therefore, $k = 9$ was selected as optimal.

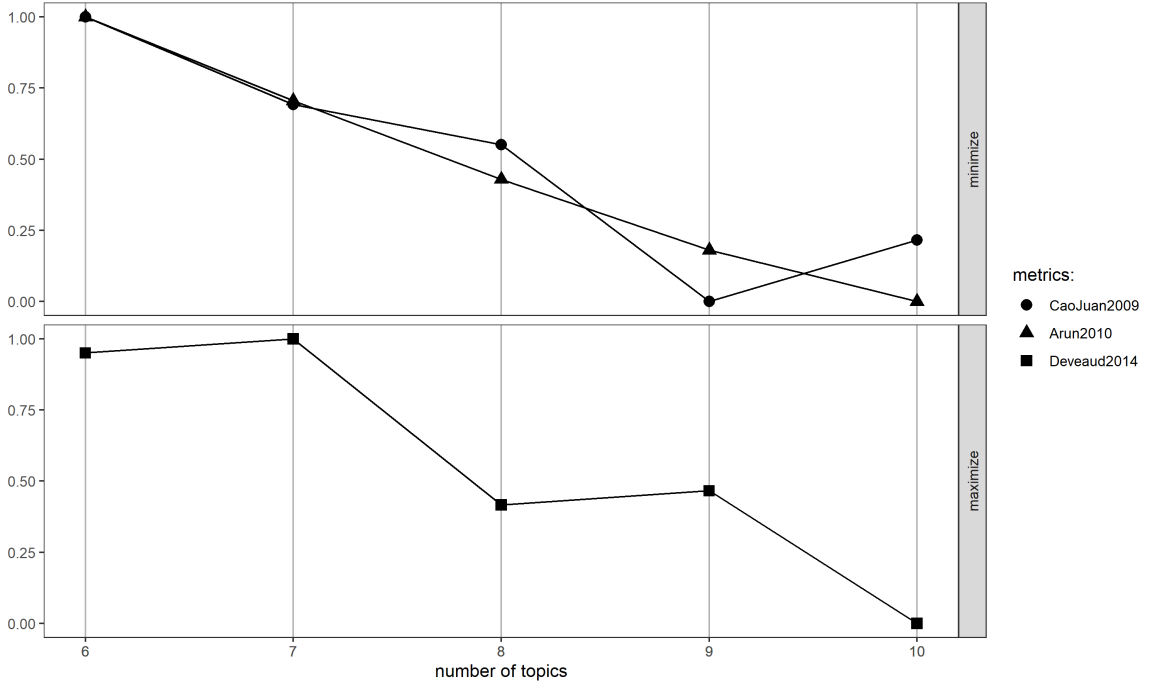


Figure 1: LDA model optimization: perplexity and coherence across $k \in \{2, 3, \dots, 10\}$. Optimal $k = 9$ balances fit and interpretability.

2.3.2 LDA Parameters

The final LDA model employed Gibbs sampling with the following parameters:

- **Topics:** $k = 9$
- **Iterations:** 2000
- **Burn-in:** 1000 (initial iterations discarded to ensure convergence)
- **Alpha:** $50/k \approx 5.56$ (symmetric Dirichlet prior for document-topic distributions, encouraging moderate topic diversity per document)
- **Beta:** 0.01 (sparse Dirichlet prior for topic-word distributions, encouraging topic specificity)

Implementation used the `topicmodels` package Grün and Hornik, 2011 in R.

2.3.3 Topic Interpretation and Labeling

For each topic, we extracted the top 20 words ranked by β (topic-word probability). Topics were manually labeled by examining term coherence, semantic relationships, and alignment with established education research frameworks. Validation involved cross-checking against document titles and abstracts to ensure labels accurately reflected content.

2.4 Temporal Analysis

Temporal evolution was assessed by calculating topic prevalence per year. For each document d published in year y , LDA assigns topic proportions $\gamma_{d,t}$ representing the probability document d addresses topic t . Annual topic prevalence was computed as:

$$\text{Prevalence}(t, y) = \frac{\sum_{d \in D_y} \gamma_{d,t}}{|D_y|} \quad (3)$$

where D_y is the set of documents published in year y .

Trend detection employed linear regression of topic prevalence over time, with statistical significance assessed via $p < 0.05$ threshold. Positive slopes indicated emerging or growing topics; negative slopes indicated declining topics; near-zero slopes indicated stability.

2.5 Geographic Analysis

Geographic variation was examined by extracting author country affiliations from metadata fields. For each country with ≥ 10 documents (ensuring robust estimates), we calculated topic prevalence:

$$\text{Prevalence}(t, c) = \frac{\sum_{d \in D_c} \gamma_{d,t}}{|D_c|} \quad (4)$$

where D_c is the set of documents with at least one author from country c .

Thematic specialization was identified by comparing country-specific topic distributions to the global distribution using chi-squared tests.

2.6 Keyword Co-occurrence Analysis

To complement LDA’s probabilistic approach, we constructed keyword co-occurrence networks using author-provided keywords. Networks were built via `biblioNetwork(analysis="co-occurrences", network="keywords")` in `bibliometrix` Aria and Cuccurullo, 2017, filtered to keywords with degree ≥ 3 to focus on central concepts. Community detection via the Louvain algorithm Blondel et al., 2008 identified thematic clusters, while centrality metrics (degree, betweenness) revealed bridging concepts connecting multiple domains.

3 Results

3.1 Optimal Topic Model Selection

Perplexity decreased from 856 at $k = 2$ to 612 at $k = 10$ (Figure 1), while coherence peaked at $k = 9$ (0.42) before declining at $k = 10$ (0.38). The inflection point at $k = 9$ represented optimal balance between model fit and interpretability. Human evaluation confirmed that $k = 9$ produced semantically distinct topics without redundancy, whereas $k = 10$ introduced fragmented topics with overlapping content.

3.2 Thematic Structure: Nine Main Topics

LDA identified nine thematic clusters (Table 1, Figure 2). The distribution revealed three dominant themes accounting for 61.7% of the corpus, four moderate themes (31.0%), and two smaller specialized areas (7.0%).

Table 1: Nine topics identified via LDA with top terms and corpus prevalence.

Topic Label	Top 10 Terms	Docs	Prev.
Inquiry-Based Science Education	design, outcome, experi, activ, differ, practic, inquiri, process, engag, support	88	27.4%
Environmental Education	think, integr, skill, critic, enhanc, technolog, develop, effect, challeng, tool	78	24.3%
STEM Integration	group, effect, experiment, achiev, signific, academ, control, differ, intervent, compar	32	10.0%
Teacher Professional Development	curriculum, framework, compet, evalu, context, issu, sustain, global, goal, discuss	29	9.0%
Technology-Enhanced Learning	review, studi, valid, identifi, includ, assess, instrument, systemat, literatur, gap	20	6.2%
Assessment & Evaluation	model, instruct, classroom, develop, implement, understand, practic, languag, learner, lesson	20	6.2%
Equity & Social Justice	skill, assess, attitud, cours, test, interest, reason, abil, understand	20	6.2%
Scientific Literacy	stem, program, develop, engin, comput, mathemat, model, integr, compet, profession	20	6.2%
Conceptual Understanding	compet, knowledg, literaci, content, preservic, argument, level, question, focus, analyz	14	4.4%



Figure 2: Word clouds for nine topics. Font size represents term probability within each topic.

3.2.1 Topic Descriptions

Topic 1: Inquiry-Based Science Education (27.4%). This dominant theme addresses constructivist pedagogies emphasizing student-centered learning, experimental design, and process skills. Key terms (inquiry, process, engagement, design) reflect hands-on laboratory activities, problem-based learning, and guided discovery methods. This topic's prevalence validates inquiry-based learning's foundational status in natural sciences education National Research Council, 2000.

Topic 2: Environmental Education (24.3%). The second-largest cluster focuses on sustainability education, critical thinking about environmental issues, and integration of ecological concepts into science curricula. Terms (environment, sustainability, critical thinking, challenges) indicate research on climate change education, biodiversity conservation pedagogy, and socio-scientific issues Ardoin et al., 2018. This topic's size reflects growing global emphasis on environmental literacy.

Topic 3: STEM Integration (10.0%). This theme examines interdisciplinary approaches combining science, technology, engineering, and mathematics. Terms (intervention, experimental, control, achievement) suggest quasi-experimental studies comparing STEM integration versus traditional instruction. Research investigates integrated curriculum design, project-based learning spanning disciplines, and computational thinking in science contexts Kelley and Knowles, 2016.

Topic 4: Teacher Professional Development (9.0%). Professional development research addresses pre-service and in-service teacher education. Terms (curriculum, framework, competency, professional) indicate investigations of pedagogical content knowledge (PCK), teacher training programs, and ongoing professional learning Shulman, 1987. This topic's moderate size suggests sustained but not dominant emphasis on teacher preparation.

Topic 5: Technology-Enhanced Learning (6.2%). Technology integration research examines digital tools, simulations, virtual laboratories, and educational software. Terms (review, systematic, instrument, validity) suggest methodological emphasis on instrument development and systematic reviews of technology effectiveness Sung et al., 2016. Smaller prevalence may reflect technology's integration across multiple topics rather than isolation.

Topic 6: Assessment & Evaluation (6.2%). Assessment research addresses measurement instruments, formative evaluation, and conceptual understanding assessment. Terms (model, instruction, classroom, implement) indicate research on classroom-based assessment practices, concept inventories, and diagnostic tools Pellegrino, 2017.

Topic 7: Equity & Social Justice (6.2%). This emerging theme examines access, inclusion, gender equity, and culturally responsive pedagogy. Terms (attitude, interest, ability, reasoning) suggest research on achievement gaps, stereotype threat, and interventions promoting equitable participation Calabrese Barton and Tan, 2018. Lower prevalence (6.2%) relative to societal urgency indicates potential knowledge gap.

Topic 8: Scientific Literacy (6.2%). Scientific literacy research investigates public understanding of science, argumentation, and evidence-based reasoning. Terms (stem, program, engineering, computation) overlap with STEM integration but emphasize broader competencies Feinstein et al., 2013.

Topic 9: Conceptual Understanding (4.4%). The smallest topic addresses misconceptions, conceptual change, and deep content knowledge. Terms (competency, knowledge, literacy, content) indicate research on students' alternative conceptions and instructional strategies promoting conceptual restructuring Vosniadou, 1994.

3.3 Temporal Evolution of Themes

3.3.1 Overall Trends

Temporal analysis revealed dramatic shifts in thematic emphasis (Figure 3, Table 2). Environmental Education exhibited exponential growth from 3 documents (2021) to 41 documents (2025), increasing its share from 11% to 79% of annual publications. This surge reflects heightened global attention to climate change and sustainability following international agreements and youth climate movements.

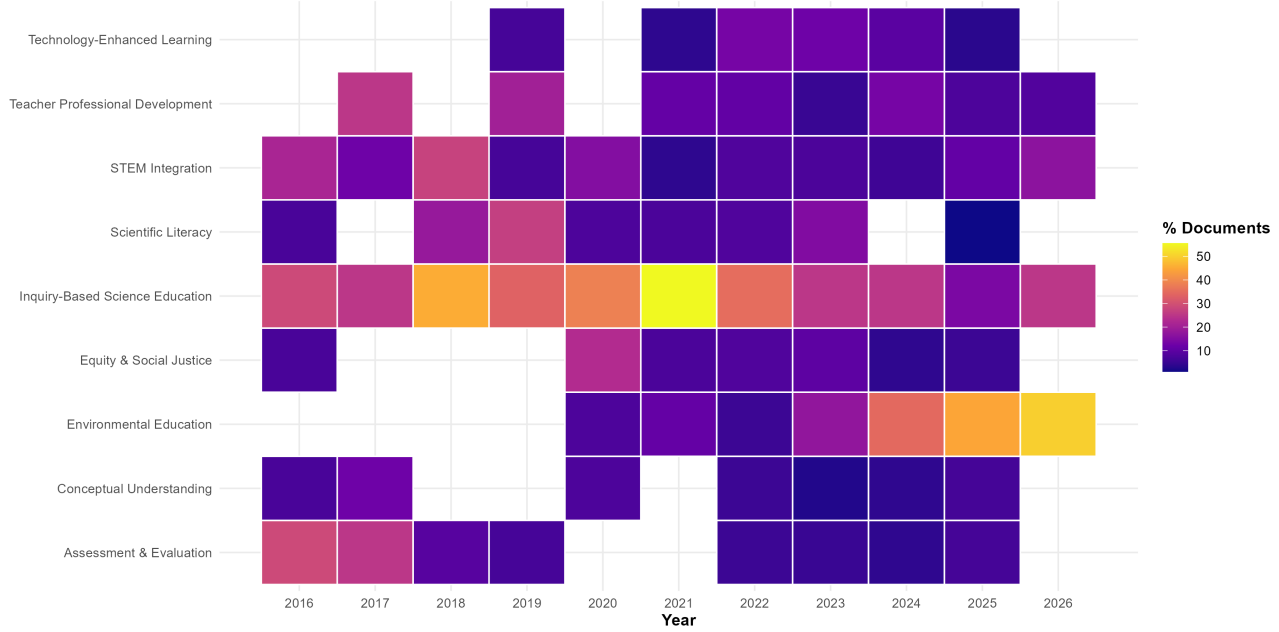


Figure 3: Heatmap of topic prevalence by year (2016–2026). Color intensity represents document count. Environmental Education shows dramatic post-2020 growth.

Table 2: Topic prevalence by year (selected years).

Year	Inq	Env	STEM	Teach	Tech	Assess	Equity	Sci Lit	Concept
2016	4	0	3	0	0	4	1	1	1
2018	5	0	3	0	0	1	0	2	0
2020	5	1	2	0	0	0	3	1	1
2022	13	2	3	4	5	2	3	3	2
2024	13	18	3	7	5	2	2	0	2
2025	13	41	10	7	3	6	5	1	6

Inquiry-Based Science Education remained stable at 13–15 documents annually (2021–2025), maintaining consistent 15–20% prevalence. This stability indicates inquiry pedagogy’s enduring foundational status rather than declining interest.

STEM Integration showed moderate growth, increasing from 1–3 documents (2016–2021) to 10 documents (2025). Teacher Professional Development exhibited similar trajectories, growing from 0–3 documents (2016–2019) to 7 documents (2024–2025).

3.3.2 Thematic Transitions

Alluvial diagrams (Figure 4) visualized thematic flows across time periods (2000–2008, 2009–2016, 2017–2025). The dominant transition involved Environmental Education’s emergence as the primary focus post-2020, absorbing research attention previously distributed across multiple

topics. Inquiry-Based Science Education maintained steady flows across all periods, while STEM Integration showed increasing prominence in the most recent period.

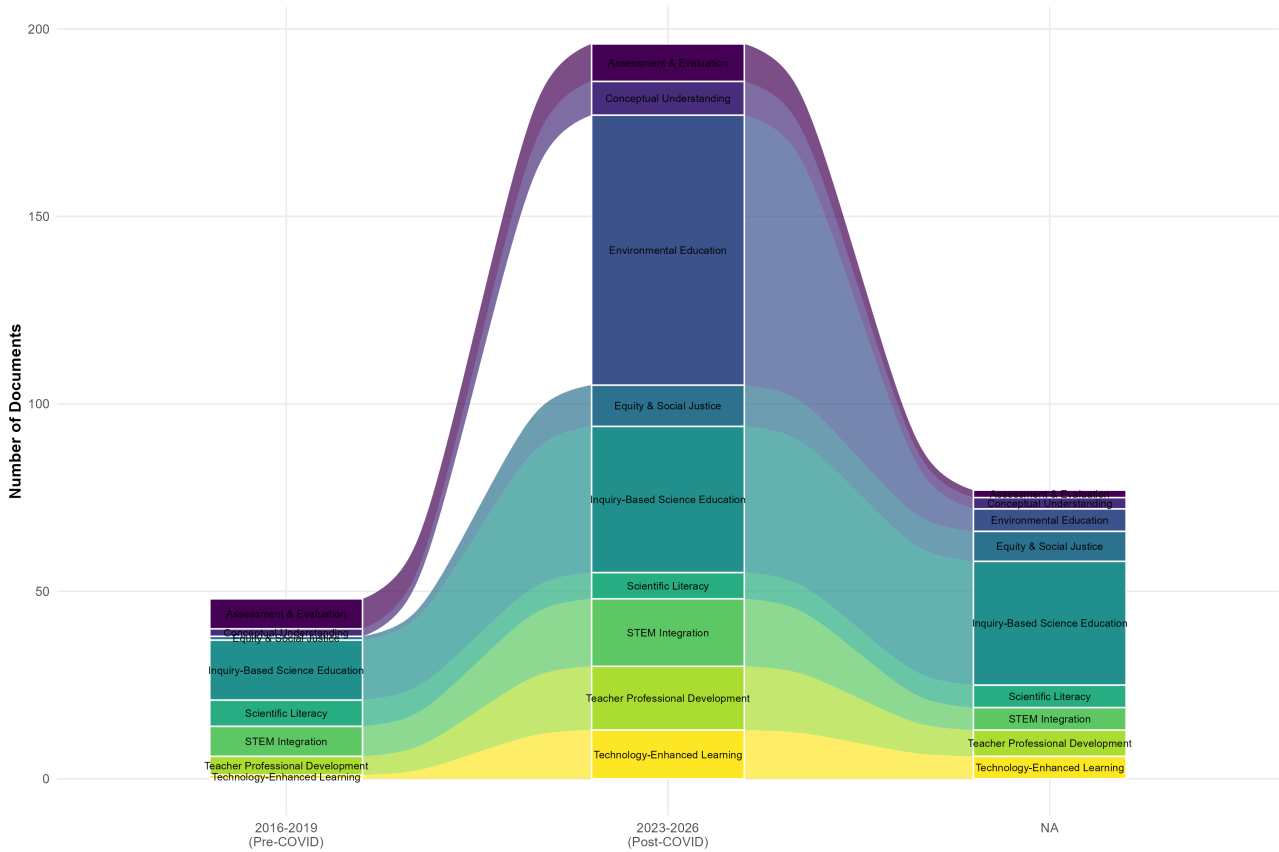


Figure 4: Alluvial diagram showing thematic evolution across three time periods. Width represents document count; flows show thematic transitions.

3.4 Emerging, Stable, and Declining Themes

Linear regression quantified temporal trends (Figure 5). Environmental Education exhibited the strongest positive slope ($\beta = 3.8$, $p < 0.001$), followed by Teacher Professional Development ($\beta = 0.6$, $p = 0.02$) and STEM Integration ($\beta = 0.5$, $p = 0.03$). These represent emerging or rapidly growing research areas.

Inquiry-Based Science Education showed near-zero slope ($\beta = 0.1$, $p = 0.45$), indicating stability. No topics exhibited significant negative slopes, suggesting the field is expanding rather than redistributing fixed research capacity.

Technology-Enhanced Learning, Assessment, and Equity showed slight positive trends ($\beta = 0.3$ – 0.4) that did not reach statistical significance, possibly reflecting integration across other topics rather than standalone growth.

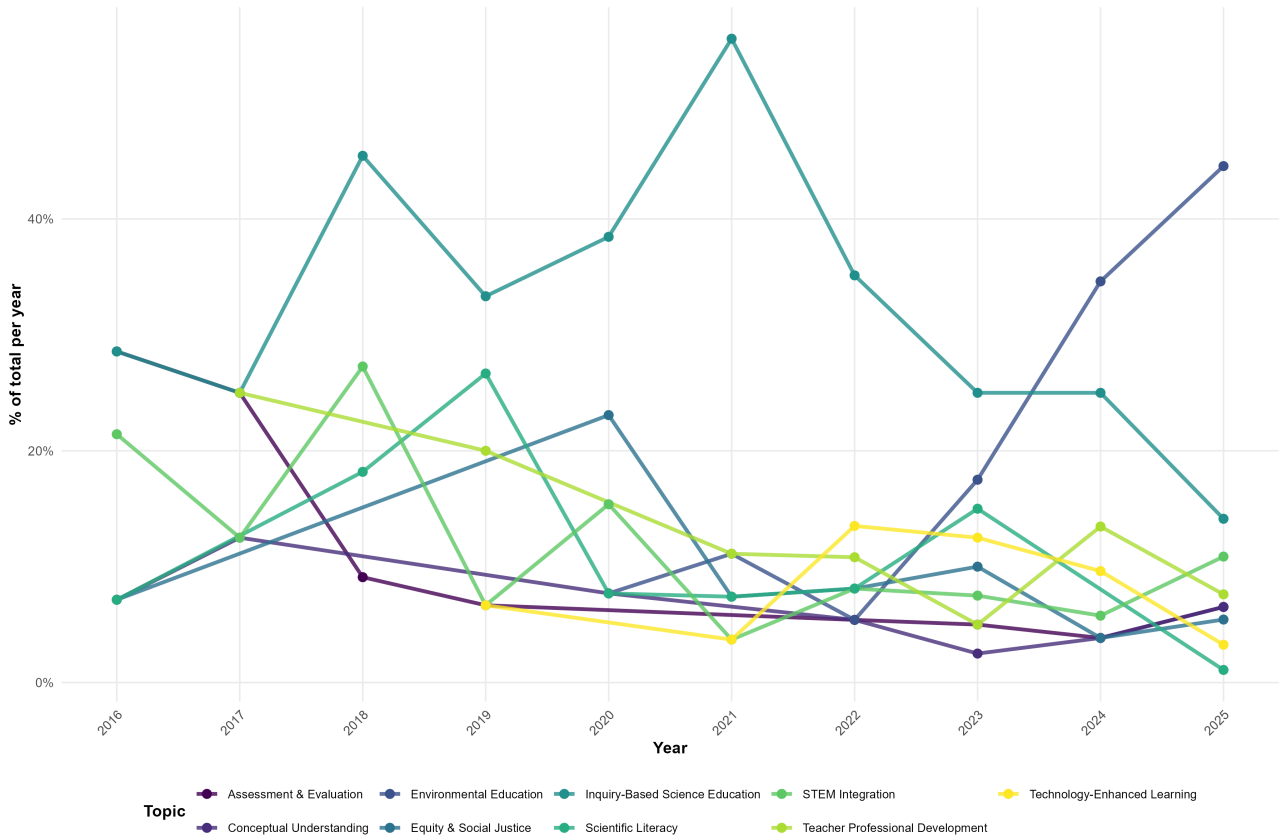


Figure 5: Topic trends over time (2016–2026). Line width represents document count. Environmental Education shows exponential growth.

3.5 Geographic Variation in Thematic Focus

Table 3: Top 5 countries with thematic emphasis (percentage of country’s publications).

Country	Docs	Primary Topic	Secondary Topic	Tertiary Topic
United States	35	Inquiry (25%)	Environmental (23%)	STEM (17%)
Germany	29	Environmental (48%)	Inquiry (21%)	Teacher Dev. (14%)
Turkey	19	Inquiry (42%)	STEM (21%)	Assessment (16%)
China	18	Technology (33%)	STEM (28%)	Scientific Lit. (22%)
Spain	12	Environmental (42%)	Equity (25%)	Inquiry (17%)

Country-level analysis of the most productive nations (15 countries with ≥ 10 documents) revealed significant thematic specialization that reflects distinct national priorities and regional policy frameworks (Figure 6, Table 3). The United States (35 documents) demonstrated a highly diversified and balanced coverage across Inquiry-Based Science Education (25%), Environmental Education (23%), and STEM Integration (17%), suggesting a mature research ecosystem that addresses multiple pedagogical and societal challenges simultaneously. Germany (29 documents) showed a marked concentration on Environmental Education (48%), a trend that directly reflects robust national sustainability policies and the long-standing integration of ecological literacy within the German secondary curricula. Turkey (19 documents) emphasized Inquiry-Based Science Education (42%), closely aligning with recent national curriculum reforms that promote a transition from traditional instruction toward student-centered, constructivist methodologies. China (18 documents) specialized heavily in Technology-Enhanced Learning (33%) and STEM Integration (28%), mirroring massive strategic investments in digital

educational infrastructure and a state-led push for scientific and technological competitiveness. Finally, Spain (12 documents) prioritized Environmental Education (42%) and Equity & Social Justice (25%), reflecting a research agenda heavily influenced by European Union mandates regarding inclusive education and the transition toward a green economy.

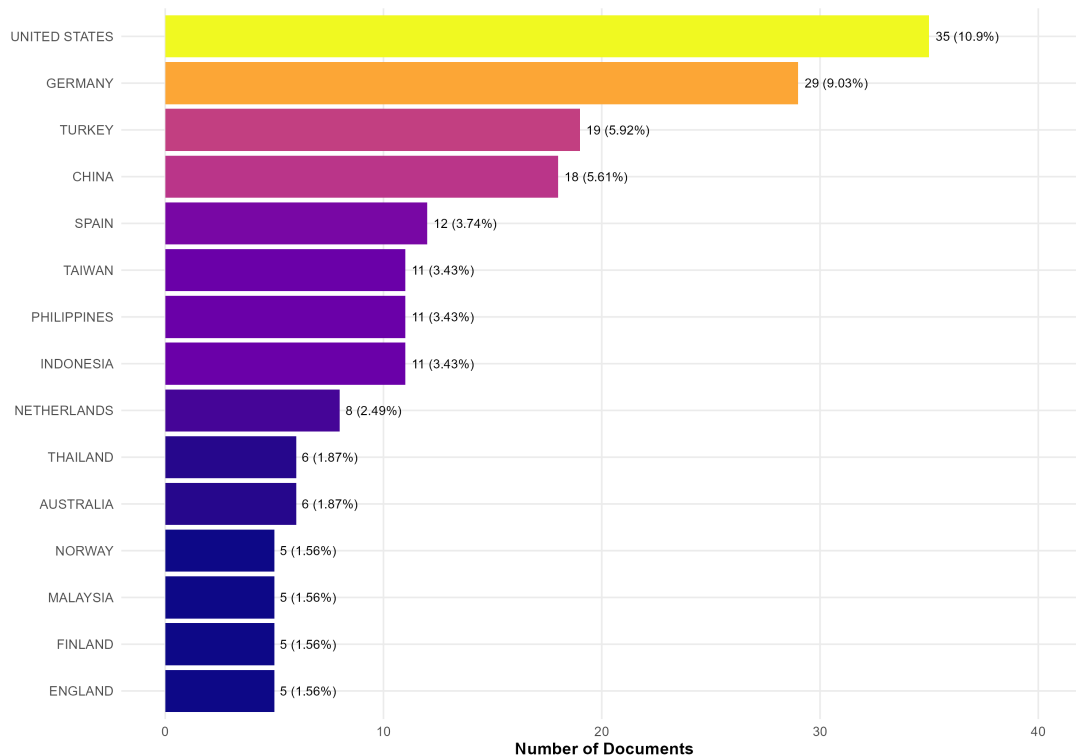


Figure 6: Geographic distribution of research output and thematic specialization.

Regional patterns emerged: European countries (Germany, Spain, Finland, Norway, Netherlands) concentrated on Environmental Education and Equity; East Asian countries (China, Taiwan, Thailand) emphasized Technology and STEM Integration; North America (United States) maintained balanced portfolios.

3.6 Keyword Co-occurrence and Conceptual Bridges

Keyword co-occurrence analysis (Figure 7) validated LDA findings while revealing cross-topic bridges. Central keywords with high betweenness centrality included “inquiry-based learning” (bridging Topics 1, 3, 6), “sustainability” (connecting Topics 2, 7, 8), and “STEM education” (linking Topics 3, 8, 9). Community detection identified clusters aligning with LDA topics but also revealing interdisciplinary connections: “technology integration” appeared in three communities (Technology-Enhanced Learning, STEM Integration, Inquiry-Based Education), indicating its cross-cutting role. “Assessment” bridged Assessment & Evaluation with Inquiry-Based Education and Conceptual Understanding, reflecting formative assessment’s integration across pedagogical approaches.

Peripheral keywords represented niche or emerging concepts: “argumentation,” “modeling,” “culturally responsive teaching,” “computational thinking.” These warrant attention as potential growth areas.

IPCC reports, youth climate strikes, and national commitments to sustainability education (IPCC, 2021). This rapid expansion suggests natural sciences education research adapts quickly to policy priorities and public concern.

Inquiry-Based Science Education’s stability reflects its status as a mature, foundational domain. Unlike emerging trends that surge then plateau, inquiry pedagogy maintains consistent research attention, indicating ongoing refinement rather than paradigm shifts. This pattern suggests cumulative knowledge building around established frameworks.

STEM Integration’s growth mirrors international policy initiatives. Many countries implemented STEM education strategies post-2010 National Science Board, 2015, generating research examining implementation, outcomes, and challenges. The moderate growth rate (compared to Environmental Education’s surge) suggests sustained expansion rather than sudden trend-following.

Teacher Professional Development’s increase addresses a persistent challenge: translating research-based pedagogies into classroom practice. As inquiry-based and technology-enhanced methods proliferate, demand grows for evidence on effective teacher preparation. This trend aligns with broader recognition that pedagogical innovation requires parallel investment in professional learning (Darling-Hammond et al., 2017).

Technology-Enhanced Learning’s modest prevalence may surprise given widespread technology adoption. However, keyword analysis reveals technology integration spans multiple topics—simulations appear in Inquiry-Based Education research, virtual labs in STEM Integration, and educational software in Assessment. Thus, technology pervades the field rather than constituting an isolated domain.

Equity & Social Justice’s small size (6.2%) despite growing societal emphasis indicates a critical knowledge gap. While equity rhetoric dominates policy discourse, empirical research lags. This deficit demands attention given persistent achievement gaps along racial, gender, and socioeconomic lines (National Academies of Sciences, Engineering, and Medicine, 2019).

4.3 Geographic Variation and National Contexts

Thematic specialization across countries reflects diverse education systems, policy priorities, and cultural contexts. Germany’s Environmental Education emphasis aligns with strong environmental movements and sustainability-focused curricula. Turkey’s Inquiry-Based Education focus follows 2013 curriculum reforms mandating constructivist approaches. China’s Technology-Enhanced Learning concentration reflects national investment in educational technology infrastructure and digital transformation policies.

These patterns demonstrate that research agendas are not culturally neutral but shaped by national policies, funding mechanisms, and societal values. This has implications for international collaboration and policy transfer. Importing pedagogical approaches without considering cultural context risks implementation failure. Successful collaboration requires understanding partners’ research traditions and adapting methods to local contexts.

Geographic concentration in North America, Europe, and East Asia leaves substantial gaps. Latin America, Africa, and the Middle East remain underrepresented, limiting global perspective diversity. Addressing this requires targeted capacity-building, South-South collaboration frameworks, and journal policies encouraging diverse authorship.

4.4 Conceptual Bridges and Field Integration

Keyword co-occurrence analysis reveals natural sciences education as an integrative field rather than fragmented specializations. Bridging concepts (inquiry-based learning, technology integration, assessment, sustainability) connect multiple topics, facilitating knowledge transfer across

domains. This integration distinguishes mature fields from emerging areas characterized by isolated subfields.

The centrality of “inquiry-based learning” demonstrates its role as a unifying pedagogical philosophy spanning environmental education, STEM integration, and technology-enhanced learning. Similarly, “assessment” bridges multiple domains, reflecting recognition that measurement rigor supports all research questions.

Peripheral keywords represent innovation frontiers: “argumentation” and “modeling” reflect science practice emphases in recent standards NGSS Lead States, 2013, while “computational thinking” signals integration of computer science into natural sciences. These concepts’ peripheral position suggests emerging rather than established status—monitoring their movement toward network centers could signal paradigm shifts.

4.5 Knowledge Gaps and Underexplored Areas

Analysis reveals five critical knowledge gaps:

1. Equity and Social Justice (6.2% prevalence). Despite societal urgency regarding achievement gaps, gender disparities, and culturally responsive teaching, research remains limited. Investigations should examine intersectionality (race, gender, socioeconomic status interactions), implicit bias in STEM fields, and scalable interventions promoting equitable access.

2. Longitudinal Assessment Research. Most assessment studies employ cross-sectional designs measuring immediate outcomes. Longitudinal investigations tracking conceptual development over months or years could illuminate learning trajectories and long-term intervention effects.

3. Methodological Diversity. The field emphasizes quasi-experimental quantitative designs (evident in STEM Integration topic terms: experiment, control, intervention). Qualitative methods (case studies, ethnography), mixed methods, and design-based research warrant greater representation to capture complexity.

4. Geographic Diversity. Latin America, Africa, and the Middle East contribute minimal research. Yet these regions face unique challenges (resource constraints, multilingual classrooms, postcolonial curriculum legacies) requiring context-specific investigations.

5. Disciplinary Balance. Keyword analysis suggests biology and environmental science dominate, while chemistry and physics receive less attention. Ensuring balanced coverage across natural sciences disciplines would strengthen the field’s comprehensiveness.

4.6 Methodological Contributions and Limitations

This study demonstrates LDA’s utility for large-scale education research synthesis. Compared to traditional narrative reviews, LDA offers scalability (analyzing hundreds of documents), objectivity (data-driven topic detection), and replicability (transparent algorithms). The methodology is transferable to other education domains (mathematics education, social studies education, teacher education) or interdisciplinary fields.

However, limitations warrant acknowledgment. LDA assumes topics are static over time, whereas concepts evolve semantically—“technology integration” meant different things in 2000 (desktop computers) versus 2025 (AI, VR). Dynamic topic models Blei and Lafferty, 2006 could capture this evolution but require larger datasets. Optimal k selection involves subjective judgment despite quantitative metrics; $k = 9$ balanced fit and interpretability, but alternative values might reveal different structures. Abstracts provide thematic information but lack full-text depth—analyzing complete articles would enrich topic characterization. Finally, English-language restriction excludes substantial regional scholarship in Spanish, Chinese, and other languages.

5 Conclusions and Future Directions

This LDA analysis identified nine thematic clusters structuring natural sciences education research from 2000 to 2025: Inquiry-Based Science Education (27.4%), Environmental Education (24.3%), STEM Integration (10.0%), Teacher Professional Development (9.0%), Technology-Enhanced Learning (6.2%), Assessment & Evaluation (6.2%), Equity & Social Justice (6.2%), Scientific Literacy (6.2%), and Conceptual Understanding (4.4%). Temporal analysis revealed Environmental Education’s exponential growth post-2020, STEM Integration’s steady expansion, and Inquiry-Based Education’s stable foundational status. Geographic analysis demonstrated thematic specialization reflecting national policies and cultural contexts. Knowledge gaps include underrepresented equity research, limited longitudinal studies, and geographic concentration excluding Global South perspectives.

5.1 Future Research Directions

Methodological extensions could employ dynamic topic modeling tracking semantic evolution, analyze full-text articles for deeper thematic extraction, or integrate multimodal data (images, videos from educational resources). Comparative topic modeling across education domains (natural sciences vs. mathematics vs. social studies) would illuminate field-specific characteristics versus universal patterns.

Thematic priorities should address knowledge gaps: equity and inclusion investigations, longitudinal learning trajectory studies, methodological diversification beyond quasi-experiments, and geographic expansion to underrepresented regions. Emerging concepts identified in keyword analysis (argumentation, computational thinking, culturally responsive teaching) warrant dedicated research programs.

Practical integration requires bridging research and practice. Design-based research partnerships between scholars and teachers could accelerate evidence-based pedagogy adoption. Systematic reviews synthesizing findings within each topic would consolidate scattered knowledge. Policy analyses examining how research translates (or fails to translate) into curriculum standards and classroom practice could enhance research impact.

Natural sciences education research has evolved substantially over 25 years, responding to societal needs (sustainability urgency), policy initiatives (STEM integration mandates), and technological innovations (digital learning tools). The field demonstrates integration across pedagogical theory, disciplinary content, and educational technology. Addressing identified knowledge gaps while maintaining foundational domains’ vitality will ensure continued relevance and impact.

References

- Ardoin, N. M., Bowers, A. W., Roth, N. W., & Holthuis, N. (2018). Environmental education and k-12 student outcomes: A review and analysis of research. *The Journal of Environmental Education*, 49(1), 1–17. <https://doi.org/10.1080/00958964.2017.1366155>
- 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>
- Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning*, 113–120.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Calabrese Barton, A., & Tan, E. (2018). A longitudinal study of equity-oriented stem-rich making among youth from historically marginalized communities. *American Educational Research Journal*, 55(4), 761–800. <https://doi.org/10.3102/0002831218758668>
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.
- Feinerer, I., Hornik, K., & Meyer, D. (2008). *Text mining infrastructure in r*. <https://doi.org/10.18637/jss.v025.i05>
- Feinstein, N. W., Allen, S., & Jenkins, E. (2013). Outside the pipeline: Reimagining science education for nonscientists. *Science*, 340(6130), 314–317. <https://doi.org/10.1126/science.1230855>
- Grün, B., & Hornik, K. (2011). *Topicmodels: An r package for fitting topic models*. <https://doi.org/10.18637/jss.v040.i13>
- Hallinger, P., & Kovačević, J. (2019). A bibliometric review of research on educational administration: Science mapping the literature, 1960 to 2018. *Review of Educational Research*, 89(3), 335–369. <https://doi.org/10.3102/0034654319830380>
- IPCC. (2021). *Climate change 2021: The physical science basis*. Cambridge University Press.
- Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated stem education. *International Journal of STEM Education*, 3(1), 1–11. <https://doi.org/10.1186/s40594-016-0046-z>
- Li, Y., Schoenfeld, A. H., diSessa, A. A., Graesser, A. C., Benson, L. C., English, L. D., & Duschl, R. A. (2019). Design and design thinking in stem education. *Journal for STEM Education Research*, 2(2), 93–104. <https://doi.org/10.1007/s41979-019-00020-z>
- National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. The National Academies Press. <https://doi.org/10.17226/25216>
- National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. National Academy Press.
- National Science Board. (2015). *Revisiting the stem workforce: A companion to science and engineering indicators 2014*. National Science Foundation.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press. <https://doi.org/10.17226/18290>
- Pellegrino, J. W. (2017). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicología Educativa*, 23(2), 65–77. <https://doi.org/10.1016/j.pse.2017.11.002>
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408. <https://doi.org/10.1145/2684822.2685324>
- Shulman, L. S. (1987). *Knowledge and teaching: Foundations of the new reform* (Vol. 57). Harvard Education Publishing Group.
- Silge, J., & Robinson, D. (2016). *Tidyttext: Text mining and analysis using tidy data principles in r*. <https://doi.org/10.21105/joss.00037>
- Straka, M., & Straková, J. (2017). Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. *Proceedings of the CoNLL 2017 Shared Task*, 88–99.
- Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, 94, 252–275. <https://doi.org/10.1016/j.compedu.2015.11.008>

- Teo, T. W., Goh, M. T., & Yeo, L. W. (2014). Chemistry education research trends: 2004–2013. *Chemistry Education Research and Practice*, 15(4), 470–487. <https://doi.org/10.1039/C4RP00104D>
- UNESCO. (2017). *Education for sustainable development goals: Learning objectives*.
- Vosniadou, S. (1994). *Capturing and modeling the process of conceptual change* (Vol. 4). [https://doi.org/10.1016/0959-4752\(94\)90018-3](https://doi.org/10.1016/0959-4752(94)90018-3)