NJCIE Nordic Journal of Comparative and International Education

ISSN: 2535-4051 Vol 10, No 4 (2026) https://doi.org/10.7577/njcie.6585

Mapping the Research Landscape on Learning Outcomes: A Structural Topic Modelling Analysis

Tiago Neves
Faculdade de Psicologia e de Ciências da Educação da Universidade do Porto/CIIE-UP
Email: tiago@fpce.up.pt

Luís Filipe de Araújo Pessoa
Universität Münster

Jozef Zelinka
Universität Münster

Marcelo Parreira do Amaral
Universität Münster

Tanja Sturm
Universität Hamburg

José Pedro Amorim
Faculdade de Psicologia e de Ciências da Educação da Universidade do Porto/CIIE-UP

Abstract

Learning outcomes (LOs) have become central across the field of education, from the political to the academic realms, through to the professional. Yet, their definition and application remain contested, ranging from narrowly measurable skills to broad developmental goals. This is further complicated by their tight entanglement with assessment, accountability, and cross-national comparability agendas. This study addresses this complex situation by, following a configurative synthesis review, employing

Structural Topic Modelling (STM) to map the research landscape in LOs in eight European countries. The corpus comprises 176 peer-reviewed journal articles. Model optimisation driven by semantic coherence and exclusivity identified 16 topics, which were later organised into five thematic categories. The findings reveal a diversified and uneven research landscape. Instructional design and educational technology and cognitive-affective dimensions were the most prevalent categories. In turn, social equity received comparatively limited attention. This points to a strong dominance of studies on how particular instructional configurations - namely technology-mediated pedagogy and individual-level learning optimisation – can boost engagement and performance. That is, within the literature that explicitly mobilises the vocabulary of LOs, measurable outcomes and instructional efficacy are prioritised, while the structural determinants of educational achievement remain at the margins. The analysis, then, reveals a persistent misalignment between prevailing research priorities in this literature and the enduring realities of educational inequality. Addressing this imbalance requires a tighter integration of micro-level instructional mechanisms with macro-level governance, accountability, and equity conditions, and a better scrutiny of how LO metrics travel and metamorphose — across national contexts.

Keywords: learning outcomes, structural topic modelling, instructional design, educational equity, comparative education

Introduction

The concept of Learning Outcomes (LOs) has become central across the field of education and is now widely used at different levels, from the political to the academic, through to the professional. This contemporary prominence is closely tied to late-twentieth-century developments in assessment, measurement, and accountability regimes (Mau, 2019; Tröhler, 2013). In mainstream usage, LOs typically refer to achievements – what learners are expected to know or be able to do after a learning experience (Adam, 2004; Hussey & Smith, 2002). Yet, both their precise definition and application are variable and contested. In recent years, a quantifiable understanding of LOs has increasingly operated as a proxy for "educational achievement," contributing to a conceptual contraction in which what counts as "educational" is equated with what is academically measurable (e.g., grades, retention, examinations, large-scale test scores) (Biesta, 2009; Hussey & Smith, 2002). Such narrowing aligns with the broader institutional fact that schooling functions as a "people-processing organisation," in which classification and credentialing enable other organisations (labour markets, higher education, welfare systems) to process individuals through standardised categories (Luhmann, 1978). Conversely, a potential advantage of this contraction is that it makes LOs amenable to standardised evaluation and cross-context comparison (Adam, 2004; Hussey & Smith, 2002). There are, then, relevant philosophical, sometimes ideological, clashes regarding the purpose and effects of LOs: while some emphasise their positive role as organisers of the learning processes, others criticise what they deem to be a utilitarian approach to education and the imposition of a one-size-fits-all approach.

The definition of learning outcomes as statements that describe what a learner is expected to know, understand, and be able to do after a learning process is grounded in attempts at transparency, standardisation, comparability and benchmarking (Adam, 2004; CEDEFOP, 2009). These attempts are frequently underpinned by neoliberal ideologies and/or governance goals, which emphasise marketable skills, employability, and economic utility. To be sure, these tensions live at the crossroads between the political and the professional. Indeed, in professional (and vocational) contexts, LOs are often closely linked to employability, practical skills and alignment with industry standards. However, such emphasis can come at the expense of broader educational aims like equity, diversity, and social justice (Cairney & Kippin, 2021). Also, it risks reducing teaching inputs to measurable outputs and having supposedly pedagogical contents and goals being shaped mostly by shifting political priorities (Caspersen et al., 2017). A broader critique to this utilitarian approach lies in what is seen as its disregard for non-measurable aspects of education, a process that Biesta calls the "learnification" of education: the transformation of all educational discourse into the language of learning and measurable outcomes (Biesta, 2009; Murtonen et al., 2017). This development risks inverting educational priorities – valuing what is measured rather than measuring what is valued (Biesta, 2009) – and aligns with broader movements toward technocratic educational cultures premised on programmed instruction and human capital enhancement (Tröhler, 2013).

These deep, structural divergences are readily apparent at the broad European scale. An example is the fact that, despite attempts to align LOs across EU countries – such as the European Qualifications Framework (EQF), the European Network for Quality Assurance in Higher Education (ENQA), the Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG), or the European Quality Assurance Reference Framework for Vocational Education and Training (EQAVET) – the approaches to LOs can differ significantly from country to country. The professional sector's involvement in defining LOs can also vary by country, with some systems granting more influence to industry bodies and others to academic institutions (Prøitz et al., 2017). This reflects varying educational philosophies, policies, and labour market needs (initiatives such as the *OECD Learning Compass 2030* and the *CALOHEE - Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe* attest this fact).

Tensions also exist in the academic field, where LOs are subject to debate and reinterpretation within disciplines and institutions. While they may be viewed and designed as a means to clarify curriculum goals and guide assessment, there is resistance to rigid, pre-specified outcomes, especially from those who value emergent or holistic forms of learning that cannot be fully anticipated or measured (Caspersen et al., 2017; Prøitz et al., 2017). These tensions may be amplified by disciplinary differences: while some fields prioritise abstract, conceptual knowledge, others emphasise practical skills and professional competencies (Prøitz et al., 2017).

At the intersection between the academic and the professional dimensions, Zelinka and colleagues (2025) have identified a flagrant mismatch between expectations regarding LOs and their actual definition: while students, teachers and other involved actors are pushed to constantly deliver high-quality LOs, they nonetheless remain unsure about what exactly that means. Not only do they frequently fail to make sense of an endless process of data collection, evaluation and comparison, they also tend to consider that LOs have been associated with a flat understanding of learning which sidelines the life course of people, the intricacies of policy implementation, the inequalities which divide societies, and the disparities between European educational contexts.

Against this backdrop, critical scholarship identifies three strands of concern regarding LOs (see Zelinka et al., 2025). The pedagogical-scientific critique highlights behaviourism's inability to capture higher-order learning and complex processes that cannot be predetermined (Murtonen et al., 2017). The economic-managerialist critique traces the assessment of LOs to Taylorist efficiency models that emphasise employability over broader educational purposes (Clarke, 2018; Holmes, 2013). The political-technocratic critique focuses on large-scale assessments like PISA, which Gorur (2016) argues flatten education into standardised, portable metrics. Despite these critiques, some argue that student-centered learning demands clear LOs (Adam, 2004). It is across all these tensional layers that the ongoing contestation over what the concept entails, how it should be operationalised, and whose interests its measurement and use ultimately serve is played.

In this article, we attempt an exploratory mapping of the research landscape on learning outcomes by answering the question: what are the main research areas, and how do they both intersect and diverge? It is expected that this endeavour highlights the dominant research areas and recurring thematic clusters, as well as the extent to which (and the ways in which) the aforementioned tensions operate to produce given configurations of learning outcomes. Ultimately, this endeavour should contribute to advance the understanding of how the interplay of different research strands enters the construction of learning outcomes.

Method

To fulfil the goals stated above, we began by conducting a configurative synthesis review of the literature on LOs covering eight European countries (Newman & Gough, 2020). The details of this study, conducted within a Horizon Europe project named *CLEAR - Constructing Learning Outcomes in Europe: A multi-level analysis of (under)achievement in the life course*, can be found in Zelinka et al. (2025). In brief, this type of review uses systematic search strategies across multiple scholarly databases to identify relevant studies (Zawacki-Richter et al., 2020). Its aim is to pattern data into configurations to develop a richer conceptual understanding of the phenomenon under investigation.

The corpus of publications identified and characterised by the configurative synthesis review served as the direct input for a complementary, data-driven analysis using Structural Topic Modelling (STM); the two stages of the study are thus sequentially connected STM is a probabilistic, unsupervised text-analysis method that infers latent topics from patterns of word co-occurrence across large sets of documents and that, unlike conventional topic models, allows document-level metadata to be incorporated as covariates of topic prevalence and content (Roberts et al., 2016: Roberts et al. 2019). It thereby enables the detection of latent semantic structures that would be difficult to identify through close reading alone, as well as the investigation of covariate effects and topic correlations. A fuller account of STM as a research approach is provided in the next section, while the methodological details concerning corpus construction, pre-processing, metadata selection, and topic estimation are presented in the Case selection subsections below. As mentioned above, it is expected that this article will enable mapping out constellations of the main themes that characterise the research landscape on learning outcomes, thereby offering a clear(er) picture of their prevalence, intersections and divergences.

The article evolves in four steps: *First*, we describe the research approach of STM and its application in the research on LOs. *Second*, we characterise the case selection, including data collection, methodological considerations, pre-processing, and modelling issues such as the procedure to identify a suitable number of topics, the evaluation metrics, and the covariates considered. *Third*, we present the findings of the STM analysis, in which we describe the resulting themes, categories, topic correlations, and covariate effects. *Fourth*, we discuss the findings in relation to the current debates on LOs and education at large.

Structural Topic Modelling as a research approach

The literature on LOs is highly diverse: even texts that are all about learning outcomes address them from quite different angles – focusing, for instance, on their meaning or on their use – and the words making up each text, and their connections, vary substantially.

STM is a topic modelling method that extends the Latent Dirichlet Allocation (LDA) method by representing texts as distributions over topics (topic prevalence) and topics as distributions over words (topic content), both of which can be affected by the metadata collected for each document in the corpus (Roberts et al., 2016; Ulstein, 2024). Unlike LDA, which infers topics solely based on the co-occurrence of words, STM incorporates document-level metadata (e.g., author information, date, location, age, or other relevant data) into the model as predictors of topic prevalence and content, thereby allowing for the investigation of covariate effects on topical structure. This technical capability of STM, together with the need to tackle the challenge of increasing volumes of textual data, (Chakrabarti & Frye, 2017; Roberts et al., 2019; Rodriguez & Storer, 2019), justifies our option for using it to conduct what we regard as an innovative exploration of research literature on LOs. Furthermore, the considerable document length provides sufficient context for meaningful word co-occurrence patterns to emerge within each publication, which supports the models' underlying assumptions for identifying latent topics.

Besides uncovering patterns that would otherwise remain unrecognised, STM's quantitative framework also enhances the replicability and transparency of social research while metadata integration improves interpretability.

The workflow in STM consists of four main steps: (1) Document Ingestion and Pre-processing, by which the documents and the respective metadata are loaded and the documents pre-processed (including lemmatisation, i.e., reducing inflected or derived words to their base dictionary form (the lemma) and the removal of punctuation, numbers, stop words, and infrequent or very frequent words); (2) Model Estimation, by which a statistical model for topical and content prevalence is computed taking into account the co-occurrence of words and the influence of the document-level metadata (covariates) on the prevalences; (3) Evaluation, by which several models are inferred and assessed through various metrics, aiming at determining the one with the best fit for the underlying data; and lastly (4) Analysis and Interpretation, by which the topics content and relationships between topics and covariates are analysed through several visualisations and tables, enabling researchers to interpret the results and draw insights into how different factors influence topic prevalence and content (Roberts et al., 2016).

Case selection

In this section, we briefly describe how the data was collected and pre-processed for the STM analysis and how the topics for the modelling were selected.

Data collection

The synthesis review on which the application of STM is grounded addressed the following questions: How are learning outcomes defined across the literature? What theoretical assumptions underpin different definitions? How are these definitions operationalised empirically? What functions do learning outcomes serve? Who are the primary stakeholders in defining learning outcomes? (see Zelinka et al., 2025).

The data collection was conducted on April 2, 2024, and included a comprehensive search across publicly accessible databases (N=28). The search was based on pre-established criteria to ensure that the research articles covered the countries participating in the CLEAR project (Austria, Bulgaria, Finland, Germany, Greece, Italy, Portugal, Spain), within which this research was carried out. All selected articles are written in English and have an explicit reference to LOs. The search considered both qualitative and quantitative studies, published in articles available in research databases, that address the definition, implementation, assessment and measurement of LOs. No date constraints were applied.

The initial search was conducted across 28 databases[1]. Based on the inclusion criteria, the search string was: “learning outcomes” AND (“educational achievement” OR “academic achievement” OR “underachievement” OR “educational attainment” OR “educational success”) AND (“Germany” OR “Spain” OR “Finland” OR “Austria” OR “Portugal” OR “Italy” OR “Bulgaria” OR “Greece”). It should be noted that this search string necessarily bounds the corpus: it captures research that explicitly mobilises the vocabulary of LOs and educational achievement. The implications of this delimitation are addressed in the Discussion.

Where appropriate, we also applied proximity operators, and truncated search terms using the conventions of each database or search engine to capture spelling variants and alternative word endings. Within each term category we combined keywords with OR, and we linked categories using AND.

[1] ERIC, Education Source Ultimate, Academic Search Ultimate, APA PsycInfo, APA PsycArticles, and 23 further databases accessed via EBSCOhost; the full list is available from the authors upon request.

The initial search yielded 220 texts, 44 of which were excluded from the sample. Reasons for exclusion were the following: 13 articles had no focus on CLEAR countries; 17 texts were not scientific articles, but rather policy briefs, book articles, or news articles; 8 articles were not written in English; and 6 articles could not be obtained due to institutional safety restrictions or were otherwise unavailable. The remaining 176 articles were collected and saved on institutional servers.

In the next step, the articles' contents were extracted and saved in tabular files. For the STM analysis, we used only the main text body, except abstracts, keywords, titles, sub-titles, image captions, tables, figures, mathematical formulas, or references. We have, however, included footnotes whenever they contained relevant information.

In addition to the main text, we also collected the articles' metadata. This included a) the name of the first author, b) the name of the scientific journal, c) the source (DOI or URL, if available), d) publication year, e) abstract, and f) keywords. Not all metadata were available for the *corpus*: 16 articles had no keywords, and one article was missing the abstract.

Methodological considerations

We implemented several measures in our methodological design in order to mitigate the risks of instability in the STM model due to a relatively small corpus (N=176) and the inherent trade-off between the number of topics and the number of documents available to estimate them. First, the appropriate number of topics (K) was determined by weighing the trade-offs across candidate models for a wide range of K values, based on two topic quality metrics (i.e., semantic coherence and exclusivity) and one model fit metric (i.e., residual dispersion). Second, the models were estimated using a deterministic initialisation, i.e., spectral initialisation (Roberts et al., 2016), thus avoiding a common source of instability in topic models applied to small corpora. Third, the covariates were modelled with additive effects only (i.e., no interaction terms) to keep the number of estimated parameters manageable considering the limited number of documents. These aspects are further detailed in the *Modelling* subsection. Fourth, the eight focal countries operated exclusively as criteria for the selection of publications, not as analytical dimensions: 'country' was not included in the model as a covariate of topic prevalence or content, and the analysis was not designed to estimate or compare country-specific topic structures. This design choice is relevant because topic models elaborated on sparse corpora are known to converge on cross-cutting common structures while underrepresenting country-level institutional heterogeneity; accordingly, our claims concern latent themes in the corpus as a whole; no inferences about national variation are drawn from the results.

Pre-processing

In the dataset cleaning process, the column names and data types of the raw data were standardised, and the articles' content, which spread through multiple columns, was merged.

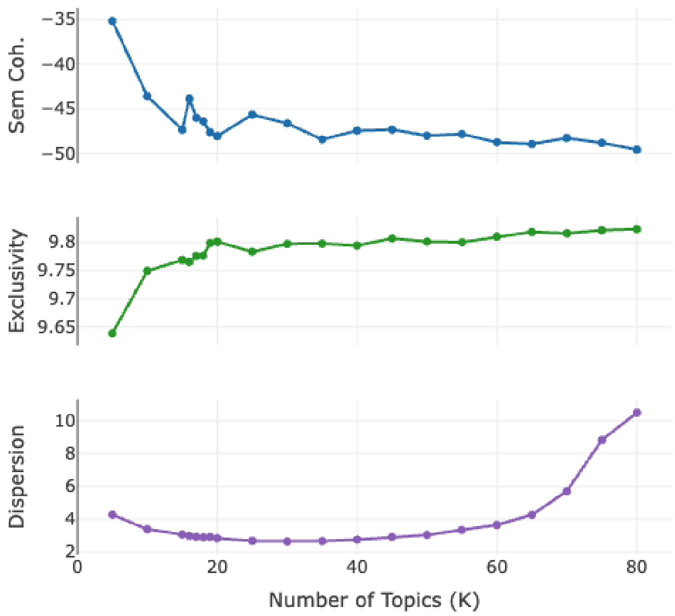
After cleaning the dataset, the documents were pre-processed to standardize terms, remove irrelevant ones, and construct the terms-document matrix for the *corpus*, which was then used as the basis for the model estimation. The main procedures employed in this step were: lower case conversion; removal of stop words, numbers, punctuations, non-alphanumeric characters, and terms with less than three letters; and word lemmatisation. The pre-processing of the dataset resulted in a vocabulary containing 16,272 words.

The next step was to explore different threshold values for rarely occurring terms, aiming at reducing the vocabulary (and thereby noise and estimation time) without losing relevant contextual information. We tested threshold values within the interval [2, 9] for the minimal occurrence of terms in different documents and assessed the number of terms and documents removed for each value (the maximum threshold value corresponds to 5% of the *corpus* size, i.e., 5% of 176). This analysis suggested that removing terms occurring in 7 or fewer documents was the best choice: a higher value would yield a relatively smaller reduction in the total number of words while potentially hindering the discovery of niche, emerging, or relevant topics. Terms that occur in more than 70% of the documents were also excluded, as they are less likely to discriminate different topics. After removing 13851 of 16272 terms (62106 of 145268 tokens) due to frequency, our corpus was constituted by 176 documents, 2421 terms, and 83162 tokens.

Modelling

The estimation of the topic model requires setting *a priori* the number of topics (K) to be modelled. To find an optimal value for K, we estimated several models for different values of K, starting at 5 and increasing its value by 5 until a maximum of 80 topics. The models were assessed based on their mean semantic coherence, mean exclusivity, and residual dispersion. The metrics for the different number of topics are depicted in Figure 1. Note that Figure 1 does not depict the topics themselves: rather, it displays the values of the three evaluation metrics computed for candidate models estimated across the range of K values, thereby documenting the evidential basis for the selection of K. The results suggested a stagnation of the metrics roughly at $K = 15$, $K = 20$; from $K > 25$, the dispersion starts to increase, suggesting a worse fit of the models and no significant quality improvement of the topics for larger K values. For smaller values of K, we notice a trade-off between semantic coherence and exclusivity, indicating that topics become more exclusive, but less coherent. An additional investigation of models for K within [15, 20] reveals that $K = 16$ leads to increased semantic coherence, while exclusivity remains high (although not at the maximum) and the residual dispersion low. It suggests, therefore, that $K = 16$ offers the best balance of the metrics.

Figure 1. Evaluation metrics for STM models across a range of topic numbers (K)



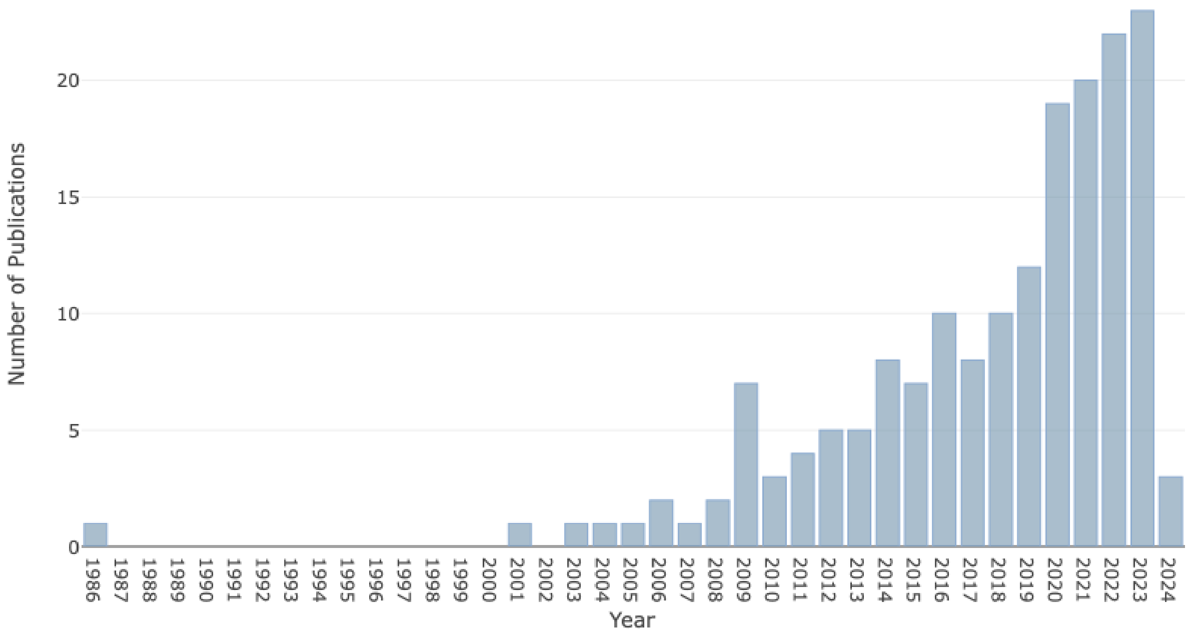
Moreover, we estimated the additive effects of publication type (empirical vs. non-empirical, review vs. non-review), Web-of-Science impact factor (IF), and publication year on the prevalence of the identified topics. The publication type was manually assigned during a screening process of all publications. Note that empirical and review are not fully orthogonal in the corpus, although most reviews are non-empirical. The impact factor was classified into four categories according to the quantiles of the observed distribution in the corpus: *low* (below 25%), *medium* (25% to 75%), *high* (above 75%), and *none*, when the information is missing. The publications are distributed as follows: empirical studies (147), reviews (27), IF-high (38), IF-medium (79), IF-low (41), and IF-none (18). We additionally applied the Benjamini-Hochberg (BH) procedure to control the false discovery rate (FDR), as the number of comparisons is large (16 topics x 6 covariate levels = 96).

Results

The STM analysis produced several types of findings. In this article, we focus primarily on the contents identified in the analysed publications – i.e., the topics, their organisation into categories, and their correlations – because they most directly serve the article's exploratory aim: mapping thematic clusters, intersections and tensions that reveal how learning outcomes are constructed. Complementarily, we also report the effects of the collected metadata (publication type, impact factor, and year) on topic prevalence.

The striking evolution in the number of publications on LOs since 2020 – as depicted in Figure 2 – cannot be overlooked. Even though the number of publications is limited in bibliometric terms, the steady rise between 2008 and 2019, along with the steep increase since 2020, suggests that the research interest in this topic is considerably growing.

Figure 2. Number of publications per year



Source: Authors' own elaboration; the data were collected on April 2, 2024.

With regard to contents, the analysis conducted identified 16 topics. Below we present their name (generated by a local Large Language Model based on the available topic information[2]), their

[2] The LLM used was Llama-3.3-70B, hosted at https://gpt.uni-muenster.de/v1/. The system prompt used was the following: You are an expert on Education Sciences, with a particular focus on Learning Outcomes. You know how to interpret and label topics based on the information provided.
Your task is to interpret a topic based on the information provided. The topic is related to Education Sciences, specifically to Learning Outcomes.
The information is provided in a structured format, including terms with high probabilities and representative documents.
The terms are categorized into two groups:
- Regular probabilities
- FREX terms (Frequent and Exclusive)
The representative documents include:
- Title

prevalence (i.e., their percentage in the overall corpus content), and the terms most frequently associated with each topic.

- Year
- Impact factor
- Type
- Relevant excerpts that mention one or more terms that are highly relevant for this topic (i.e., those that have the highest probability for the topics)

Your task is to break down the information and perform a step-by-step analysis of the information about the topic in order to:
- to identify the actual topic content,
- provide a short, suitable, clear, and precise label for the topic, and
- write a short description for this topic, based on the excerpts of the relevant documents and their relations to the relevant terms. You can also take the metadata of the relevant documents into account. This description/summary of the topic must be at most 1000 words long.
- provide an evaluation of coherence of the topic, based each excerpts from the different relevant documents. This evaluation must be at most 400 words long.
- provide an overall assessment of the relevance of the topic in the context of Education Sciences and Learning Outcomes. This assessment must be at most 400 words long.
- identify the interdisciplinary aspects and areas of the topic, if any. This assessment must be at most 400 words long.
- identify the mentioned and related theories, if any. This assessment must be at most 400 words long.
- identify the dimensions, the measures, and the aspects of learning outcomes used or mentioned in the relevant documents, if any. Only include the dimensions, the measures, and the aspects that are mentioned in the excerpts. This assessment must be at most 400 words long.
- suggest the dimensions, the measures, and the aspects of learning outcomes related to this topics, if any. This assessment must be at most 400 words long.

You also need to provide a sound, grounded explanation and reasoning for the provided topic label and description, and how the relevant documents fit into the topic.
Your output MUST be in the following JSON format
{format}
You SHOULD NOT include any other text in the response. You MUST format the response according to the provided JSON schema.
If you are not able to provide a label or description, just say that you are not able to do so. Do not make up a label or description. If you are not able to provide an explanation, just say that you are not able to do so. Do not make up an explanation.

Table 1. Topic content details

Nr.	Prevalence	Content	Top terms
1.	5.5%	Dealing with Mistakes in the Classroom	percept, mathemat, complex, mistak, school
2.	4.5%	Learning Outcomes in Medical Education	patient, intervent, adult, treatment, attent
3.	4.4%	Personalised Scaffolding in Self-Regulated Learning	scaffold, srl, mobil, metacognit, regul
4.	7.3%	Interpersonal Relationships and Academic Achievement	school, parent, children, peer, adolesc
5.	5.1%	Nursing Education and Student Experience	countri, nurs, non, programm, graduat
6.	4.7%	Inclusive Education for Students with Moderate, Severe, and Complex Disabilities	inclus, belief, post, pre, disabl
7.	6.8%	Competence-Based Learning Outcomes in Higher Education	game, compet, busi, project, team
8.	3.8%	Inquiry-Based Science Instruction and Learning Outcomes	style, scienc, inquiri, children, divers
9.	4.9%	Personalised Feedback Strategies in Computer-Based Learning Environments	learner, feedback, condit, script, argument
10.	8.3%	Motivation and Learning Outcomes in Education	orient, action, gender, theori, deep
11.	5.6%	Emotions and Learning Outcomes in Online Learning Environments	emot, perceiv, anxieti, onlin, adapt
12.	6.2%	Second Language Learning Outcomes and Personality Traits	languag, domain, meta, size, learner
13.	13.6%	Flipped Classroom and Online Learning	onlin, face, technolog, flip, collabor
14.	5%	Virtual Reality in Education	virtual, condit, solut, prior, solv
15.	8.7%	The Use of PISA Results in Finnish Education Policy	school, countri, pisa, pupil, nation
16.	5.7%	Elaborate Feedback and Self-Regulated Learning	feedback, write, learner, journal, prompt

Source: Authors' own elaboration.

Pushing the analysis forward, we then organised these 16 topics into broader thematic categories, considering their more theoretical or empirical orientation:

- *Category 1 – Cognitive and Affective Dimensions of Learning*: This category has a more theoretical orientation and focuses on features such as motivation, emotion, and personality as internal processes with an important role in learning (and learning outcomes).

- *Category 2 – Instructional Designs, Educational Technologies and their Evaluation:* This category is the largest and has a mostly applied or empirical orientation, focusing on the design and evaluation of innovative instructional models and educational technologies, mostly geared towards digital learning environments.

- *Category 3 – Learner Relationships and Social Contexts: Equity Concerns:* This category, although small in size, blends empirical analysis with socio-educational theory, focusing on social and developmental factors.

- *Category 4 – Higher Education: Student Experiences and Competences*: This category is mostly empirical and focuses specifically on the assessment of educational interventions in the higher education sector, regarding both learning outcomes and workforce-oriented goals.

- *Category 5 – Education Policy and Large-Scale Assessment*: This category is constituted by only one topic. Focused on the measurement of academic performance across educational systems, it has an empirical nature but also policy-theoretical implications.

The STM analysis reveals, then, a diversified research landscape, which can be structured around five thematically distinct clusters (see Table 2).

Table 2. Category per prevalence and content details

Category	**Prevalence**	**Constituent Topics**
2. Instructional Designs, Educational Technologies and their Evaluation	37.4%	Topic 9 – *Personalised Feedback Strategies in Computer-Based Learning Environments;* Topic 16 – *Elaborate Feedback and Self-Regulated Learning;* Topic 3 – *Personalised Scaffolding in Self-Regulated Learning;* Topic 13 – *Flipped Classroom and Online Learning;* Topic 14 – *Virtual Reality in Education;* Topic 8 – *Inquiry-Based Science Instruction and Learning Outcomes*
1. Cognitive and Affective Dimensions of Learning	25.6%	Topic 10 – *Motivation and Learning Outcomes in Education;* Topic 11 – *Emotions and Learning Outcomes in Online Learning Environments;* Topic 12 – *Second Language Learning Outcomes and Personality Traits;* Topic 1 – *Dealing with Mistakes in the Classroom*
4. Higher Education: Student Experiences and Competences	16.4%	Topic 5 – *Nursing Education and Student Experience;* Topic 7 – *Competence-Based Learning Outcomes in Higher Education;* Topic 2 – *Learning Outcomes in Medical Education*
3. Learner Relationships and Social Contexts: Equity Concerns	12.0%	Topic 4 – *Interpersonal Relationships and Academic Achievement;* Topic 6 – *Inclusive Education for Students with Moderate, Severe, and Complex Disabilities*
5. Education Policy and Large-Scale Assessment	8.7%	Topic 15 – *The Use of PISA Results in Finnish Education Policy*

Source: Authors' elaboration based on the results of STM analysis

It is important to note that nearly two thirds of the literature (63% of the combined topic prevalence) revolve around Category 2 (37.4%) and Category 1 (25.6%). Interestingly, higher education (Category 4) constitutes a significant field by itself, totalling 16.4%. Finally, it is also revealing that broader social and political concerns amount to little more than a fifth (20.7%) of the combined topic prevalence (Category 3, with 12.0%, and Category 5, with 8.7%).

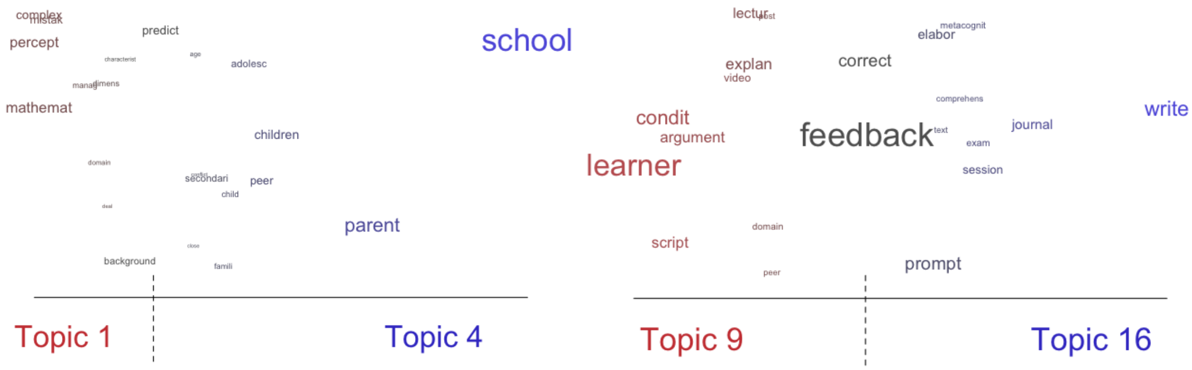
These figures reflect a clear predominance of applied research on instructional design and digital learning environments over other themes: a strong interest in how specific instructional configurations may enhance learner engagement and performance, visible in the increasing emphasis on technologically mediated personalised instruction and on the performance of individual learners, considering their specificities.

It must be noted that the organisation of the 16 constituent topics into 5 thematic categories goes beyond – and sometimes differs from – the correlations between topics identified via the structural modelling analysis. The computed correlations are based on the co-occurrence of topics within documents: correlated topics might be related, but not necessarily similar, since two topics may appear together in texts while remaining semantically distinct (e.g., a topic about assessment metrics and a topic about policy governance may simply be frequently discussed in tandem). Therefore, correlations are not a proxy for similarity: the relationships must be interpreted based on the actual content and semantics of the topics.

In our case, only two topic pairs show positive correlations: topics 1 and 4 ($r = 0.004$) and topics 9 and 16 ($r = 0.054$), suggesting that the identified topics are mostly thematically distinct, with little co-occurrence within the same document. Nonetheless, we carried out a qualitative, interpretative analysis to identify meaningful thematic relationships between topics beyond their statistical correlation, thus enabling their grouping into the appropriate categories.

This qualitative analysis identified that topics 9 and 16 belong to the same category, while topics 1 and 4 articulate with distinct topic groups – a pattern also supported by the different correlation magnitudes and the perspective plots in Figure 3. Both topics 9 and 16 concern instructional and feedback mechanisms in learning contexts (e.g., *correct*, *feedback*): Topic 9 focuses on explanatory formats and learner argumentation within computer-based environments (e.g., *lecture*, *explanation*, *learner*, *argument*) and Topic 16 is centred on the elaborative and metacognitive dimensions of feedback that support self-regulation (e.g., *feedback*, *elaboration*, *metacognition*, *comprehension*). In turn, Topic 1 focuses on the cognitive and perceptual aspects of error handling in instructional settings (e.g., *complex*, *mistake*, *perception*, *mathematics*), whereas Topic 4 concern the social and relational context surrounding students, including peers and family (e.g., *school*, *children*, *peer*, *parent*, *family*). According to the perspective plot, the shared terms *background* and *predict* seem to bridge both topics, which is confirmed by a manual inspection of documents where both topics have prevalence greater than 5%. These publications mainly investigate how individual cognitive or psychological factors (e.g., *perception*, *cognitive*, *ability*, *emotional engagement*) are related to educational outcomes within a social or equity-oriented context. Altogether, these results emphasise the importance of examining the nature of correlations instead of taking them as sufficient evidence of thematic relationship – even though, despite their minimal magnitude, they successfully uncovered relevant relationships between the topics.

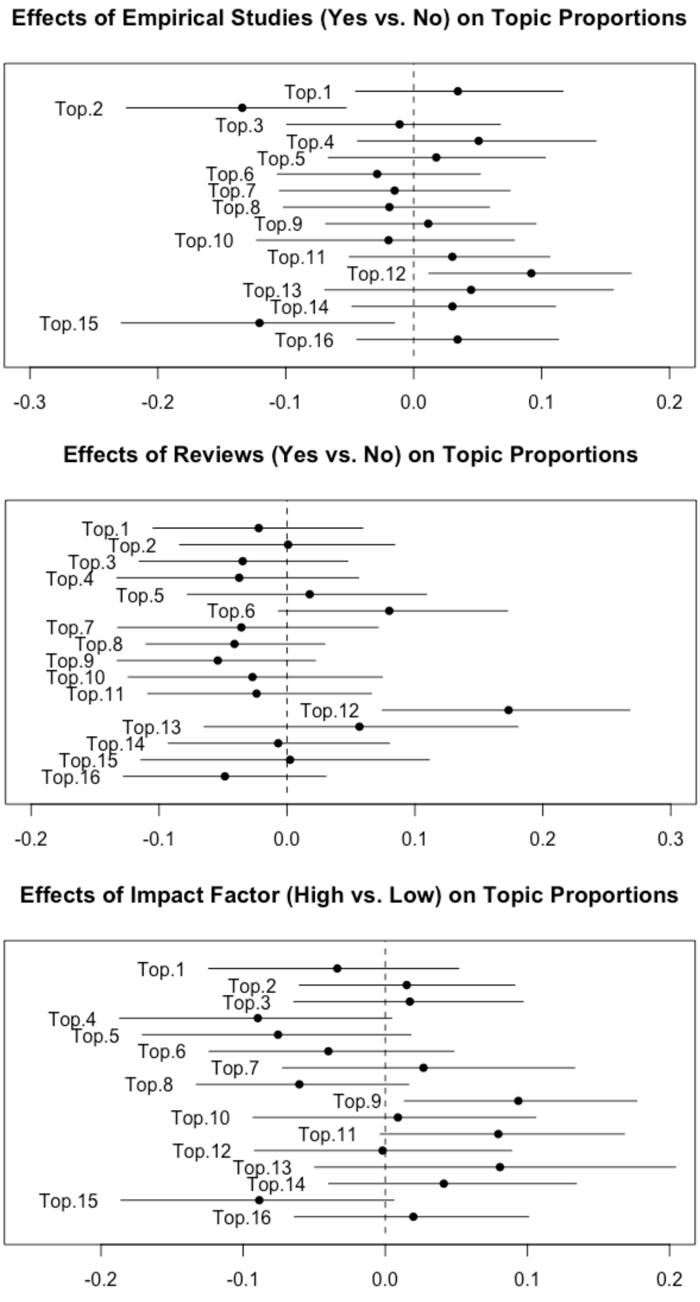
Figure 3. Perspective plots for the positive correlated topics



From a methodological perspective, this points to the importance of combining different approaches to make sense of the data, particularly when correlations are weak. In this regard, integrating STM with topic modelling methodologies that account for semantic similarity (e.g., BERTopic) is a promising direction for future work.

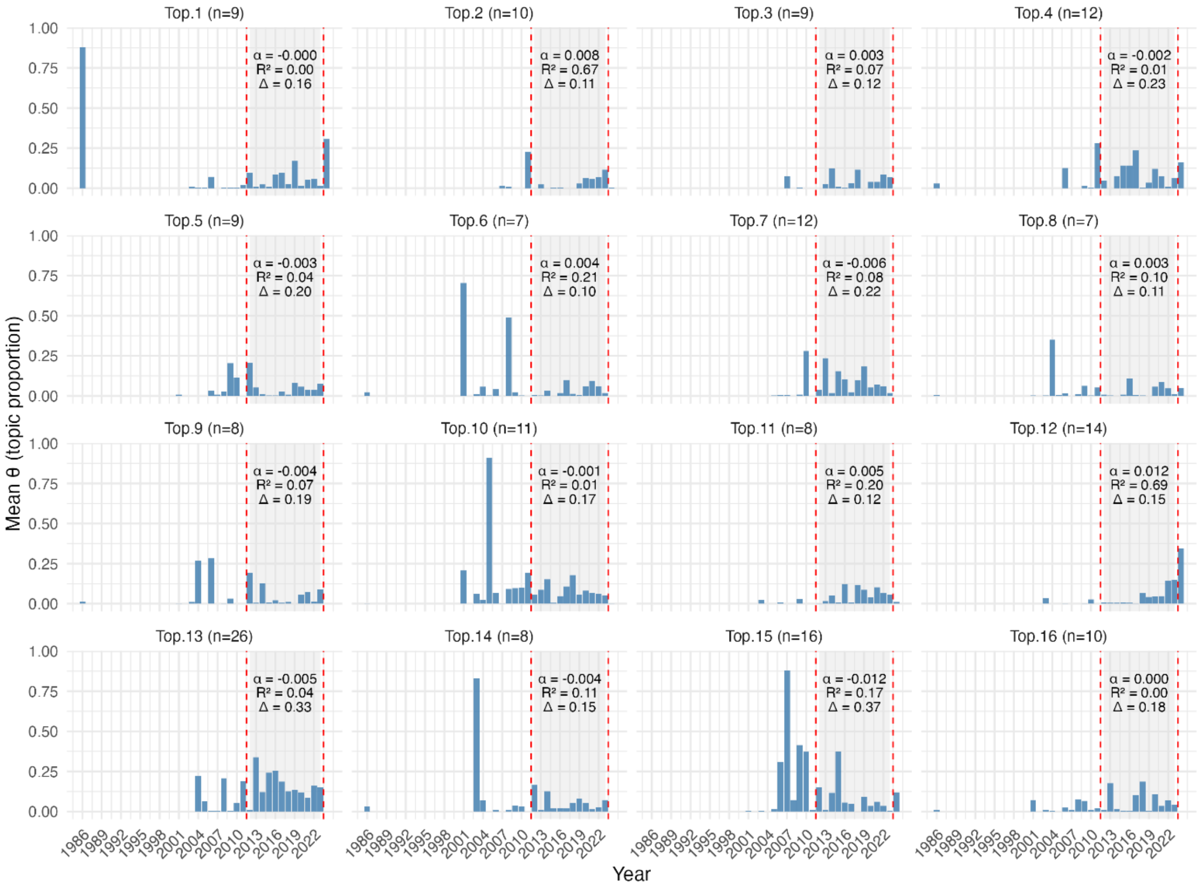
Additionally, we investigated the effects of covariates on the topic proportions. The estimated difference in topic prevalence associated with two levels of empirical studies, reviews, and impact factor are depicted in Figure 4. In all of them, most topics show wide confidence intervals that crosses the zero, reflecting a large estimation uncertainty due to the sample size. For empirical studies, Topic 2 and Topic 15 present the largest negative associations, with confidence intervals not crossing zero, which suggests they are more prevalent among non-empirical publications. However, the width of these confidence intervals indicates considerable variability in the estimate of the difference in prevalence. Conversely, Topic 12 is more prevalent among empirical studies and reviews, with the latter showing a considerably larger positive effect size. In turn, Topic 9 is more prevalent in publications with higher impact factor than in those with lower impact factors. After applying the BH correction procedure for multiple comparisons, only two effects remained statistically significant: reviews are positively associated with Topic 12 ($\beta = 0.183$, $p_{adj} = 0.038$), and empirical studies negatively associated with Topic 2 ($\beta = -0.135$, $p_{adj} = 0.038$). This result is an additional indicator that a larger dataset would be needed to reliably estimate the covariate effects.

Figure 4. Effects of the covariates empirical studies, reviews, and impact factor in the topics proportions



The analysis revealed that the effects of year on the topic proportions is not statistically robust, with large confidence intervals for years below 2012 and p-values > 0.05. Complementarily, we carried out a descriptive analysis of mean topic proportion by year averaging the topic's proportion over all documents published in that year, as illustrated in Figure 5. In this analysis, we fitted a linear regression for the period with more consistent and higher number of publications per year, 2012–2023. The slope (α), coefficient of determination (R^2), and amplitude of variation (Δ) for the respective topics are also reported in Figure 5.

Figure 5. Mean topic proportions per year: *n* – number of documents in which the topic has the highest prevalence; $\boldsymbol{\alpha}$ – slope of the linear regression fitted between 2012 and 2023; $\boldsymbol{R^2}$ – coefficient of determination of the linear regression; Δ - amplitude of variation within the same period



Outside this window, several peaks with considerably high mean topic proportions ($\geq 50\%$) are direct consequences of the small number of publications in those years (≤ 5 and often a single document) rather than evidence of shifts or early thematic prominence (e.g., topics 1, 6, 10, 14, 15). Within the window, most topics present irregular fluctuations instead of a clear trend, which is consistent with the absence of a significant year effect and reflected by low explained variance (R^2) and, for several topics, large amplitudes of variation (e.g., topics 15, 13, 4, 7, and 5). Mean topic proportions also tend to be lower and more stable in the last third of the time window, when the number of publications is two to four times higher than before: as expected, the larger and more heterogeneous sets of documents yield more stable topic proportions.

Conversely, topics 2 and 12 are the only cases for which there is a moderate, positive linear trend ($\alpha = 0.008$ and $\alpha = 0.012$, respectively) with relatively high R^2 values (0.67 and 0.69, respectively), which may indicate that they represent emerging topics gaining traction in recent years. Furthermore, by taking the continuity of mean topic proportions between 2012 and 2023 into consideration, some topics (e.g., Topic 13, Topic 10) seem to be consistently addressed themes (i.e., well spread, almost no gaps), whereas others (e.g., Topic 3, Topic 8, Topic 9) seem to appear more marginally and intermittently in the analysed literature, not as a sustained research focus.

In the following section, we delve into the most prevalent topics in each of the five categories described above to better illustrate how STM can illuminate the specificities of the factors that shape the quality and construction of learning outcomes in the focal European countries.

Delving into the topics

Topic 13, titled Flipped Classroom and Online Learning, is the most prevalent topic in Category 2 - Instructional Designs, Educational Technologies and their Evaluation, which is the most prevalent category: it has an overall 13.6% prevalence within a category that accounts for 37.4% of the topics. The fact that a single topic within such a diversified research landscape amounts to nearly 15% of all cases is relevant in and of itself, and points to the tightness of the links between education, pedagogy and technology, being paradigmatic of the weight that innovative, technologically mediated instruction strategies take on in the current landscape of learning opportunities. A flipped classroom reverses the conventional teaching structure: students first get acquainted with new contents through self-directed study, namely using digital materials, and then attend class to engage in interactive exercises, peer collaboration, and teacher-facilitated application of concepts.

Textual evidence reveals an emphasis on the capacity of this methodology to strengthen student participation and achievement, alongside the imperative for educators to embrace innovative practices – a relevance amplified by the COVID-19 pandemic, which catalysed the widespread adoption of remote learning.

The retrieved excerpts typically highlight the pedagogical benefits (increased engagement, motivation, satisfaction, and improved learning outcomes) and the technical challenges (teacher adaptation, technical support) of the flipped classroom, providing examples of successful implementation in various contexts, particularly higher education.

The most prevalent topic in the second most prevalent category (Cognitive and Affective Dimensions of Learning) is Topic 10, titled Motivation and Learning Outcomes in Education, with an overall prevalence of 8.3%. The core of this topic – the relationship between motivation and academic performance – was also touched upon in the previous topic, but in the context of a particular instructional design; here it is explored more broadly and from multiple disciplinary standpoints (psychology, education, sociology), including the role of achievement goals, self-determination, and interest in promoting achievement. Also noticeable is that its prevalence is quite lower than that of the most prevalent topic in the previous category (8.3% versus 13.6%).

The texts examine multiple motivational frameworks – expectancy-value, social cognitive, self-determination, and achievement goal theories – in their relationship with learning outcomes, emphasising deep learning approaches and intrinsic motivation as crucial factors in educational achievement. Importantly, learning outcomes are understood as having different dimensions, including theoretical knowledge, practical knowledge, and generic skills.

The most prevalent topic in the third most prevalent category (Higher Education: Student Experiences and Competences) is Topic 7, titled Competence-Based Learning Outcomes in Higher Education, with an overall prevalence of 6.8%. The documents present research on innovative educational strategies – business simulation games, portfolio methodologies, and game jam activities – as tools for improving student achievement, with the development of competences, the efficacy of non-traditional strategies, and implementation obstacles as transversal themes. The topic is interdisciplinary, drawing on education sciences, psychology, sociology, and economics, and the combination of competences and higher education suggests a concern with the relationship between education and the labour market.

The textual data encompass social constructivist, experiential learning, and self-determination theories, which highlight the importance of active learning, social interaction, and self-directed learning in the development of competences. Like the two previous topics, this one also touches upon motivation – namely through self-efficacy and flow theories – making it increasingly clear that motivation is a transversal theme on which others hinge. The facets of LOs identified range from universal competencies to targeted managerial skills (analytical problem-solving, effective communication, team collaboration), with critical thinking, creativity, and metacognitive skills such as self-reflection and self-regulation deemed essential.

Last in prevalence come, as noted above, topics related to broader social and political concerns. The most prevalent topic in the fourth most prevalent category (Learner Relationships and Social Contexts: Equity Concerns) is Topic 4, titled Interpersonal Relationships and Academic Achievement, with an overall prevalence of 7.3%. It is clear from the topic title that it revolves around a somewhat watered-down take on social issues, which are approached from an individual-based stance focused on interpersonal relationships. In this context, the more political dimension, although not absent, is only slightly touched upon. Literature in this topic examines how social connections – including family, educator, and peer relationships – influence academic performance. Research indicates that nurturing bonds with parents, teachers, and classmates contribute positively to educational success, whereas strained or antagonistic relationships tend to undermine achievement. The research nonetheless underscores the complexity of academic prediction, which requires analysing interconnected social-emotional variables such as adult guidance relationships, peer belonging, and accessible support structures; in disciplinary terms, it draws on psychology, sociology, and education.

The texts present studies informed by several theories, including attachment theory, self-determination theory, and social support theory, which offer a framework for understanding the mechanisms by which interpersonal relationships influence academic achievement. In these studies, the evaluation of learning outcomes spans academic performance, social connectivity, and mental-emotional health domains. Regarding measurement strategies, these combine objective testing, subjective reporting, and behavioural documentation approaches.

Finally, the fifth category (Education Policy and Large-Scale Assessment) includes only one topic: Topic 15 – The Use of PISA Results in Finnish Education Policy, which has an 8.7% overall prevalence. Interestingly, this is the topic with the second highest prevalence in the set of 16 topics, surpassed only by Topic 13, on Flipped Classroom and Online Learning (which, as seen above, has a prevalence of 13.6%). Therefore, although this is the category with the lowest prevalence, its only topic is the second most prevalent across all topics. This seemingly paradoxical situation may be explained by the attention attracted by what was regarded, at some point, as the Finnish miracle in education – Finland's consistently high performance in international assessments like PISA, combined with a focus on equity and teacher autonomy. Although this view has meanwhile been complexified, the Finnish case appeared at one point as something to be studied in order to be replicated elsewhere. The literature in the topic analyses how Finnish education authorities employ PISA data in policy development – how policymakers interpret these results and incorporate them into rationales for reforms – drawing on concepts from education policy, sociology, and psychology.

The topic draws upon different theoretical strands: governmentality theory, for analysing power mechanisms in policy formation, and assessment and evaluation theories, for identifying how systematic evaluation can enhance effectiveness – that is, a more political approach is combined with a more technical one.

Discussion and conclusion

Our goal in this article was to identify dominant research areas and thematic clusters in the literature on learning outcomes for selected European countries. By clarifying which specific configurations of learning outcomes emerge, we have sought to advance the understanding of how the interplay of different research strands enters the construction of learning outcomes. We note that the results and analysis presented in this work are bounded in two important ways. First, they concern the eight focal countries and the topics covered by the search string used to retrieve publications from the scientific databases. Second, and relatedly, the corpus comprises research that explicitly mobilises the vocabulary of LOs and educational achievement. Within European academic contexts, scholarship concerned with equity, class, or migrant education is often developed under adjacent framings – such as ‘inclusive education’, ‘Bildung’, or ‘critical pedagogy’ – that, to the extent that do not mobilise LOs as a core keyword, fall outside our corpus. Our findings should thus be read as statements about the literature that explicitly engages the vocabulary of LOs in specific countries, not about European educational research at large. From this angle, the comparatively limited prevalence of equity concerns within the corpus may itself be diagnostic: it suggests a discursive divide whereby equity-oriented scholarship develops largely outside – and possibly at a deliberate distance from – the measurement-oriented vocabulary of LOs. It must be stressed that mobilising this vocabulary does not entail endorsing it: the corpus includes critically oriented work – particularly in the governmentality-informed analyses of assessment-based policy within Topic 15 – yet such critical engagement remains marginal in prevalence terms. Probing this divide, for instance by comparing the present corpus with corpora built around those adjacent framings, is a promising avenue for future research.

Despite its limits, our analysis indicates the existence of clear research priority imbalances. Indeed, a massive emphasis on instructional effectiveness through technology (37.4%) coexists with limited attention to structural equity (12.0%). This suggests that, within this literature, research prioritises individual optimisation over systemic inequalities, alongside optimising existing structures over questioning or transforming them. The predominant question is “*How can we make learning more effective?*” rather than “*What systemic conditions enable or constrain learning?*”. Educational challenges, then, tend to be framed as technical problems solvable through better design and pedagogical innovation rather than as manifestations of deeper social, economic, and political inequalities in need of systemic transformation. This pattern is consistent with a form of technological determinism in which research, while examining implementation effectiveness, rarely asks whether technological solutions tackle the root causes of inequity or reproduce existing hierarchies instead. Two caveats are nonetheless in order. First, our analysis provides a cross-sectional map of thematic prevalence: while it documents an imbalance, it does not authorise causal claims to the effect that the emphasis on technology displaces socio-political critique. Second, alternative – and not mutually exclusive – explanations for the observed imbalance must be acknowledged. The prominence of instructional and technological research may reflect the political economy of academic publishing rather than researchers' indifference to social equity: intervention studies of instructional design tend to rely on more accessible data and more standardised methods, and to yield readily measurable results, making them more amenable to publication under prevailing incentive and accountability regimes. Yet, it should be highlighted that this alternative reading does not weaken our broader argument but rather reinforces it: the very pressures towards measurability that shape the definition of LOs may also be shaping the research agenda about them. This techno-pedagogical approach is integral to an emphasis on the individualisation of learning reminiscent of the warnings by Biesta (2009) and Tröhler (2013). Yet again, the focus on individual psychological characteristics or personalised interventions risks oversight of collective factors, namely of how LOs are shaped by resource distribution, teacher conditions, curriculum pressures, and socioeconomic stratification. In the same vein, higher education's emphasis on competence-based learning (Category 4, 16.4%) reflects the pressure to align competences and outcomes with labour market demands. Here, it is the broader purposes of education (democratic citizenship, critical thinking, aesthetic appreciation) that risk relegation. Thus, it can be argued that our findings – even if limited by a corpus focused on eight European countries – suggest that the economic-managerialist and political-technocratic critiques are far from unfounded.

Thus, from the strains between different forces, tensions materialise: first, while LOs do emerge as multidimensional constructs encompassing cognitive, practical, emotional, social, and metacognitive domains, their assessment remains predominantly quantitative and struggles to capture less tangible dimensions like critical thinking or ethical reasoning – reflecting ongoing tensions between the expansive scope of LOs and the narrower operationalisation demands of accountability systems. Second, there are tensions between empirical and conceptual approaches, individual-focused and system-level perspectives, technological determinism and socio-political critique. Overcoming such polarisations requires integrative studies examining instructional effectiveness and structural conditions simultaneously, as well as more interdisciplinary dialogue – namely between educationalists, cognitive scientists and education policymakers – to build a more comprehensive theory of learning outcomes. A realistic way forward, one that satisfies proponents of both narrower and broader approaches to LOs, may anchor itself on the acknowledgment that motivation and self-regulation, two fundamental features for successful learning, are related to the background of students and, therefore, associated with equity issues.

Finally, a note on the methodology. Although the STM was applied to a relatively small corpus (N=176), the methodological design enabled the training of a stable and interpretable topic model for a carefully chosen number of topics. Based on the STM model, (i) relevant themes were identified, (ii) topic correlations were uncovered and characterised through qualitative assessment, (iii) covariate effects on topic prevalence were assessed, and (iv) the temporal distribution of the mean topic proportions was exploratorily investigated. However, the wide confidence intervals of most covariate effects (only two of which survived the correction for multiple comparisons) suggest that a larger corpus would be required to increase the statistical power of these estimates, especially if interaction effects must be considered. Altogether, the generated artefacts allowed for a multi-faceted, data-driven characterisation of the literature on LOs for the focal EU countries, delivering a broad, plausible and original picture of the research landscape. Nonetheless, understanding the substantive meaning of topic structures, their correlations and covariate effects required interpretive human work; combine computational analysis must be combined with the contextual, qualitative interpretation of domain expertise.

References

Adam, S. (July 1−2, 2004). *Using learning outcomes: a consideration of the nature, role, application and implications for European education of employing "learning outcomes" at the local, national and international levels.* United Kingdom Bologna Seminar, Heriot-Watt University (Edinburgh Conference Centre), Edinburgh.

Biesta, G. (2009). Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21*, 33–46. https://doi.org/10.1007/s11092-008-9064-9

Cairney, P., & Kippin, S. (2021). The future of education equity policy in a COVID-19 world: a qualitative systematic review of lessons from education policymaking. *Open Research Europe*, 1. https://doi.org/10.12688/openreseurope.13834.2

Caspersen, J., Frølich, N., & Muller, J. (2017). Higher education learning outcomes – Ambiguity and change in higher education. *European Journal of Education*, 52, 8-19. https://doi.org/10.1111/ejed.12208

CEDEFOP. (2009). *The shift to learning outcomes. Conceptual, political and practical developments in Europe*. https://www.cedefop.europa.eu/en/publications/4079

Chakrabarti, P., & Frye, M. (2017). A mixed-methods framework for analyzing text data: Integrating computational techniques with qualitative methods in demography. *Demographic Research, 37*(42), 1351–1382. https://doi.org/10.4054/DemRes.2017.37.42

Clarke, M. (2018). Rethinking graduate employability: The role of capital, individual attributes and context. *Studies in Higher Education*, *43*(11), 1923–1937. https://doi.org/10.1080/03075079.2017.1294152

Gorur, R. (2016). Seeing like PISA: A cautionary tale about the performativity of international assessments. *European Educational Research Journal*, 15(5), 598–616. https://doi.org/10.1177/1474904116658289

Holmes, L. (2013). Competing perspectives on graduate employability: Possession, position or process? *Studies in Higher Education*, *38*(4), 538–554. https://doi.org/10.1080/03075079.2011.587140

Hussey, T., & Smith, P. (2002). The trouble with learning outcomes. *Active Learning in Higher Education*, 3(3), 220–233. https://doi.org/10.1177/1469787402003003003

Luhmann, N. (1978). Erleben und handeln. In H. Lenk (Ed.), *Handlungstheorien interdisziplinär II. Handlungserklärungen und philosophische Handlungsinterpretationen. Erster Halbband* (pp. 235–253). Fink.

Mau, S. (2019). *The metric society: On the quantification of the social*. Polity Press.

Murtonen, M., Gruber, H., & Lehtinen, E. (2017). The return of behaviourist epistemology: a review of learning outcomes studies. *Educational Research Review*, *22*, 114–128, https://doi.org/10.1016/j.edurev.2017.08.001

Newman, M., & Gough, D. (2020). Systematic reviews in educational research: Methodology, perspectives and application. In O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Eds.), *Systematic reviews in educational research* (pp. 3–22). Springer VS. https://doi.org/10.1007/978-3-658-27602-7_1

Prøitz, T., Havnes, A., Briggs, M., & Scott, I. (2017). Learning Outcomes in Professional Contexts in Higher Education. *European Journal of Education*, *52*, 31-43. https://doi.org/10.1111/ejed.12207

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for Experimentation in the Social Sciences. *Journal of the American Statistical Association, 111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software, 91*(2), 1–40. https://doi.org/10.18637/jss.v091.i02

Rodriguez, M. Y., & Storer, H. L. (2019). A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data. *Journal of Technology in Human Services, 38*(1)*,* 54–86. https://doi.org/10.1080/15228835.2019.1616350

Tröhler, D. (2013). The OECD and the Cold War culture: thinking historically about PISA. In H.-D. Meyer & A. Benavot (Eds.). *PISA, power, and policy: the emergence of global educational governance* (pp. 141-161). Routledge.

Ulstein, J. (2024). Structural topic modelling as a mixed methods research design: a study on employer size and labor market outcomes for vulnerable groups. *Quality & Quantity, 58*, 4331–4351. https://doi.org/10.1007/s11135-024-01857-2

Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (Eds.). (2020). *Systematic reviews in educational research: Methodology, perspectives and application*. Springer VS. https://doi.org/10.1007/978-3-658-27602-7

Zelinka, J., Amorim, J. P., Parreira do Amaral, Jablonczay, N., Kovacheva, S., Kyttälä, M., Papadimitriou, A., Rambla, X., Siri, A., & Tikkanen, J. (2025). A Synthesis Review of Education Research on Learning Outcomes. In J. Tikkanen, M. Parreira do Amaral, T. Järvinen, & N. Alves (Eds.), *Constructing learning outcomes: problematizing (under)achievement in Europe* (pp. 103–125). Bloomsbury Academic. http://dx.doi.org/10.5040/9781350446717.ch-4