ISSN 2411-7390

jle

JOURNAL OF LANGUAGE & EDUCATION

Volume 11 Issue 3, 2025

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

EDITOR-IN-CHIEF

EDITORIAL BOARD

EDITORIAL TEAM

CONTENT

EDITORIAL

RESEARCH PAPERS

OPINION PAPERS

BOOK REVIEWS

https://doi.org/10.17323/jle.2025.28633

Prevailing Themes and Future Research Agendas in Higher Education: A Systematic Scoping Review of Reviews (2024-2025)

Elena Tikhonova [1, 2], **Lilia Raitskaya** [3]

[1] HSE University, Moscow, Russia

[2] Peoples' Friendship University of Russia (RUDN University), Moscow, Russia

[3] Moscow State Institute of International Relations (MGIMO University), Moscow, Russia

ABSTRACT

Introduction: The existing scholarship related to higher education, while extensive, lacks comprehensive analyses of prevailing research themes. Numerous reviews provide detailed charts of niche topics, yet they fall short of offering a comprehensive overview. To accurately chart the major tendencies within the broader field of higher education, our analysis includes all review types, arguing that they collectively provide a more complete picture and reveal the field's dominant themes.

Purpose: To provide a high-level synthesis of the current scholarly landscape by analyzing the primary themes and recommended research agendas outlined in secondary literature, including systematic, scoping, bibliometric, and conceptual reviews.

Method: We employed a systematic scoping review methodology to map the breadth of literature. The review questions were formulated using the PCC (Problem, Concept, Context) framework to ensure a comprehensive scope. The conduct and reporting of this review followed the guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR). 128 reviews of diverse types published in 2024-2025 were selected from the Scopus database. The included reviews were based on publications within the timeframe from 2010 to 2025.

Results: This systematic scoping review reveals that systematic reviews (55%) constitute the dominant methodological approach in higher education research, followed by bibliometric (12%) and scoping reviews (9%). The analysis identifies eight prevailing thematic clusters that directly inform future research agendas, consolidating them into four synthesized directions: technology and digital transformation; curriculum and learning outcomes; equity, ethics, and education's social role; and systemic shifts including internationalization and sustainability. Notably, the field shows concentrated emphasis on artificial intelligence integration and AI competency development, signaling a pivotal direction for future higher education research.

Conclusion: This review has synthesized the current intellectual landscape of higher education to map its prevailing themes and, consequently, to define a unified research agenda. The primary implication of this work is to focus the efforts of the research community by highlighting the most pertinent and promising directions for further investigation. Although this review provides valuable insights, its findings are limited by its reliance on a single database and its restriction to English-language reviews, potentially omitting significant literature from regional or non-English sources. Therefore, we recommend future research pursue a more expansive systematic mapping with a longer time span, multiple databases, and incorporate reviews published in various languages to achieve a truly comprehensive perspective.

KEYWORDS

higher education; systematic scoping review; review of reviews; university; artificial intelligence (AI); digital literacy; digital competence

Citation: Tikhonova, E., & Raitskaya, L. (2025). Prevailing themes and future research agendas in higher education: A systematic scoping review of reviews (2024-2025). *Journal of Language and Education, 11*(3), 5-59. https://doi.org/10.17323/jle.2025.28633

Correspondence:
Elena Tikhonova,
etihonova@hse.ru

Received: August 1, 2025
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

Higher education is a complex social institution, characterized by a long history and a distinct academic culture. Its social significance is twofold: it reproduces a professionally stratified workforce and functions as a primary engine of innovation, contributing approximately 17% of total global research and development[1]. With over 235 million students enrolled worldwide[2], the sector's scale is immense. Consequently, it has become a multi-disciplinary field of study in its own right, attracting scholarly attention from education, sociology, psychology, and the humanities (Lund et al., 2020).

Despite a vast body of literature, the higher education research landscape lacks comprehensive, cross-cutting thematic analyses, with most studies restricted to specific national, regional, or disciplinary contexts (Pandey et al., 2025; You et al., 2024). Recent attempts to provide synthesis have emerged, though they remain focused on particular sub-fields. For instance, Eaton (2025) explores the impact of artificial intelligence on global education, advocating for a "post-plagiarism" framework, while Alzahmi et al. (2025) provide a bibliometric review of 25 years of research on organizational change in higher education. Similarly, Mahrishi et al. (2025) systematically review global trends of outcome-based education in engineering and its alignment with UN Sustainable Development Goals. While these studies represent a step forward, they collectively highlight that a comprehensive synthesis of the field's overarching research themes is still absent, encouraging further scholarly effort.

To delineate the intellectual landscape of contemporary higher education research, this scoping review of reviews analyses a comprehensive collection of recent (2024–2025) Scopus-indexed literature syntheses. The corpus encompasses both evidence-based reviews (e.g., meta-analyses, systematic, and scoping reviews) and traditional scholarly reviews (e.g., bibliometric, conceptual), which collectively cover primary research published between 2010 and 2025. The resulting synthesis of prevailing themes and emergent research agendas offers a foundational roadmap for the broader scholarly community. The timeframe (2024-2025) of our scoping review was determined by the launch of ChatGPT in November 2022, as it brought an avalanche of publications on the consequent transformations in higher education.

In the Social Sciences, the synthesis of research must navigate a spectrum of evidence. Although systematic reviews are traditionally valued for their methodological rigor, this preference can marginalize other vital forms of scholarship, such as narrative, descriptive, and bibliometric reviews. These approaches capture the nuanced, "softer" evidence necessary for charting broad intellectual tendencies. Therefore, our scoping review of higher education literature incorporates all review types to comprehensively delineate the field's prevailing themes.

This scoping review of reviews aims to map the current intellectual landscape of higher education research by synthesizing findings from review articles published within two distinct but purposeful timeframes: 2024–2025 and 2010–2025. The delimitation to the immediate two-year period, 2024–2025, is directly and purposely tied to the public release of ChatGPT in late 2022. This boundary was set to capture the consequent proliferation of literature examining the ensuing paradigm shifts within the higher education sector following this pivotal technological event.

Simultaneously, the selection of the broader 2010 to 2025 timeframe is justified by its unique positioning to capture a complete and transformative epoch in global higher education. The period around 2010 represents a critical inflection point, marking the convergence of several paradigm-shifting trends that collectively redefined the university's mission and operations. This era was characterized by a "great transition" where institutions were learning to operate in a new, multi-faceted paradigm. Structurally, this involved the full global implementation of the Bologna Process, moving beyond policy adoption to the tangible restructuring of curricula around competency-based approaches and the widespread establishment of the two-cycle degree system.

Digitally, the dawn of the modern EdTech revolution was catalyzed by the advent of Massive Open Online Courses (MOOCs) and the systematic institutional integration of Learning Management Systems (LMS) and blended learning, fundamentally challenging traditional pedagogical delivery. Concurrently, universities navigated a matured agenda of internationalization and embedded the global sustainability framework of the Sustainable Development Goals (SDGs) into their core functions, all while operating under sustained post-financial crisis pressure to demonstrate economic value and employability outcomes. By extending the timeframe to 2025, the review ensures the capture of the full maturation, interplay, and long-term consequences of these disruptive forces, including the recent impact of generative Artificial Intelligence, thereby providing a comprehensive lens to assess a definitive fifteen-year period of unprecedented and consolidated change. By analyzing reviews across these two timeframes, this study seeks to identify the field's prevailing themes and articulate a consolidated agenda for future re-

[1] OECD. (2023). Main Science and Technology Indicators. https://www.oecd.org/en/data/datasets/main-science-and-technology-indicators.html

[2] UNESCO Institute for Statistics (UIS). (2023). Total enrollment in tertiary education, regardless of age, both sexes. https://databrowser.uis.unesco.org/

search. The review is guided by the following research questions:

RQ 1: What are the key characteristics (e.g., methodological approaches, geographic focus, topics) of review articles on higher education published in 2024-2025?
RQ 2: What prevailing themes are synthesized in these reviews as central to contemporary higher education?
RQ 3: What gaps and priorities for further research are identified across these reviews, forming a collective future research agenda?

Method

This study employed a systematic scoping review methodology, guided by the frameworks of Arksey & O'Malley (2005), Levac et al. (2010), and Peters et al. (2020a), and reported in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) (Tricco et al., 2018; Pollock et al., 2022). We conducted a scoping overview of reviews to synthesize meta-level evidence: the dominant themes and future research agendas identified across systematic, scoping, bibliometric, and conceptual reviews. This design is particularly suited for mapping cross-methodological tendencies and comparing the convergence or divergence of conclusions across different review genres. This meta-level approach is especially relevant for the period following 2023, as it allows for the examination of how accelerated digital transformation and AI integration may have reconfigured thematic priorities in the literature. The unit of analysis is the individual review article. Key themes and research agendas were extracted verbatim and subsequently consolidated using a predefined coding framework. To address heterogeneity, findings were stratified by review genre and, where applicable, by methodological quality indicators.

Protocol

While starting the systematic scoping review of reviews, we meticulously developed a research protocol. The authors hereby certify that this review report constitutes a faithful, precise, and transparent description of the conducted systematic scoping review of reviews. No deviations from the protocol were registered. Any departures from the original study design have been duly elucidated.

Search Eligibility Criteria

In this review, the problem, concept, and context (PCC) framework was engaged to devise an effective search strategy where each criterion was justified (Table 1).

Search Strategies

To identify prevailing themes in higher education research, this study analyzes review articles indexed in the Scopus database. A systematic search was performed on August 21, 2025, using the keyword "higher education" and related concepts. The search string included "higher education" AND "artificial intelligence" OR "education for sustainable development" OR "competence" OR "digital competence" OR "digital literacy" OR "critical thinking" OR "blended learning" OR "online learning". An initial filter was applied to include only documents indexed as "review" by Scopus. To ensure methodological consistency, the full text of each result was then screened, with particular attention to the methods section, to verify that it described a recognized review methodology. Although preliminary searches tested a wider set of keywords, they failed to retrieve a significant number of additional relevant publications; consequently, the final search strategy was streamlined to the most productive terms.

Study Selection

After the searches had been made and the Scopus filters (areas of research, language, types of sources, timespan) had been applied, the authors identified reviews subject to the eligibility criteria enumerated in Table 1. Duplicate publications were identified and removed using Zotero. Each reviewer independently screened the titles, and then the abstracts of the identified documents. The reviewers tagged each document with "to include" or "to exclude" marks to compare the outcome. When occasional disagreements arose, they were settled by mutual consent. No disputed issue required lateral expertise. Title/abstract screening $\kappa = 0.71$; full-text $\kappa = 0.82$. The full texts of potentially eligible studies were obtained through online databases, publisher platforms, or by request from the corresponding authors. Each article underwent a full-text review, during which it was thoroughly examined and independently assessed for relevance by two reviewers.

Quality of the Synthesis Included in the Review

The quality of the synthesis of the included reviews tends to be considered in the context of the type of reviews, frameworks applied, and rigour of all the review stages as the frontrunners. As we included in our synthesis reviews of all type that conformed to the eligibility criteria, we outlined the following features to address the issue of synthesis quality:

(1) type of the review;
(2) appliance of traditional frameworks that provide for better standards of identification, selection, screening of the publications and reporting of the findings as well as frameworks for guiding reviews to more focused research questions (PRISMA, PICO - PICOT, PCC, etc.);
(3) guidelines and recommendations beyond traditional evidence-based frameworks that were mainly published in methodological articles;

Table 1

Eligibility Criteria

Criterion	Inclusion	Exclusion	Rationale
Problem	Prevailing themes of reviews on higher education and further research agendas; themes are recurrent field-level topic explicitly reported as a primary theme in Results/Discussion of the review; research agendas are explicit statements of future research directions, recommendations, or priorities	Reviews going beyond higher education	Prevailing themes of reviews on higher education accumulate the trends in the subject area
Concept	Educational and closely associated concepts, including artificial intelligence, education for sustainable development, competence, digital competence, digital literacy, critical thinking, blended learning, and online learning	Other concepts	Themes of reviews may be based on educational or closely related concept. Pre-protocol searches found the most frequently used concepts
Context	All higher education institutions	Any other contexts	The focus on higher education determines the review context
Language	English	Other languages	The choice of English is justified by its status as a lingua franca of international science
Timeframe	All reviews on higher education published and indexed in the Scopus database for 2024-2025 whose own inclusion criteria for primary studies were limited to a timeframe between 2010 and 2025	All reviews published before 2024 or with a timeframe before 2010	The search was limited to reviews published in 2024–2025. The delimitation of the review to 2024–2025 is directly tied to the public release of ChatGPT in late 2022. This boundary was set to capture the consequent proliferation of literature examining the ensuing paradigm shifts within the higher education sector. The temporal scope of the primary literature within these included reviews was constrained to the period of 2010–2025.
Types of sources	In the Scopus database: full texts of reviews of all types	Unavailable full texts	This review aims to get a comprehensive understanding of the prevailing themes
Geography of affiliation	Any location	None	To avoid a geographic bias and to assemble a body of evidence that reflects the worldwide scope of scholarly discourse on this topic, enhancing the generalizability of the review's findings
Database	Scopus	Other bases	The study utilized the Scopus database because it provides robust indexing of peer-reviewed literature in the relevant field. Its strong coverage of Social Sciences and Education journals makes it well-suited for capturing the breadth of thematic developments in higher education research
Areas of Research	Social Sciences	Other areas	Analysis of publication categories reveals that the discourse is centered in the Social Sciences but is fundamentally interdisciplinary, with many works jointly classified under both Social Sciences and other related fields

(4) identified implications for higher education stakeholders based on a scale embracing "not identified", "identified as a separate sub-section", and "identified in the Discussion, Conclusion or other sections.

Extracting and Charting the Data

To reply to the research questions, we require the following data:

(1) the information for evaluation of the quality of the included reviews as stated herein;
(2) the themes of the currently published reviews of higher education that constitute the prevailing trends in research;
(3) further research agendas in higher education as defined in the reviews.

A data-charting form was cooperative developed based on the data extraction limitations. The items were chosen according to the research questions. The data extraction process was conducted entirely by Reviewer 2, with a thorough revision of the first ten completed forms by Reviewer 1 to ensure the consistency in the extraction and the accuracy of the extracted data. Both reviewers were in constant contact with one another to discuss and settle any issues arising at this stage.

The raw data extracted from the reviews comprise the synthesized coding matrix, including themes, categories, and relevant excerpts from the included literature. To ensure transparency and facilitate replication, the data extraction forms and the detailed coding framework used in the analysis have been included as tables, figures and appendices to this publication.

Visualisation and Thematic Clusters

We made a visualization based on author keywords, applying VOSviewer. The resulting visualization with a threshold of three occurrences per keyword grouped the documents into thematic clusters. In addition, the themes of all publications were coded while we were extracting the raw data from the reviews. At the stage of the analysis, we distributed the reviews among the thematic codes that essentially coincided with those suggested by the VOSviewer, with some overlapping as most themes and subsequently clusters are rather complex and far-reaching. The outcome gave an overview of the prevailing themes and details of the research focus.

To triangulate the findings and identify predominant themes, we conducted a comparative analysis between the thematic groupings generated by VOSviewer (derived from author keywords) and the thematic clusters identified through our manual coding process.

Table 2

A Data-Charting Form

	Data to be extracted
1.	Title of review
2.	Author(s)
3.	Country/ Countries of affiliation(s)
4.	Journal Title
5.	Journal Ranking in Social Sciences (Education) if applicable (Scopus Q1-Q4)
6.	Journal Ranking in Subject Areas other than Social Sciences (Education) if no ranking in Social Sciences (Education) – Scopus Subject Area(s)
7.	Type of synthesis as defined by the author(s) in the title, abstract or methods
8.	Theme of the review (Problem)
9.	Timeframe of the review
10.	Population of the review if applicable
11.	Type of higher education institution
12.	Further research agenda
13.	Implications (identified as a sub-section; identified, but within various sections; not identified)
14.	Targeted stakeholders of the implications
15.	Synthesis frameworks and protocols if stated (PICO-PICOT, PCC, PRISMA)
16.	Applied guidelines and recommendations referred to in the reviews

RESULTS

Search and Selection Results

A systematic search was executed in the Scopus database on August 21st, 2025. The initial search string, incorporating the keywords "higher education", "artificial intelligence", "education for sustainable development", "competence", "digital competence", "digital literacy", "critical thinking", "blended learning", and "online learning" returned 5,562 records. These records were filtered according to the pre-defined inclusion criteria, limiting them to reviews published in English within the Social Sciences subject area from 2024 to the present. This process identified 961 potentially relevant reviews for title and abstract screening. After screening, 694 records were excluded, resulting in 267 publications for which full-text retrieval was initiated.

The 260 full-text reviews were thoroughly read by each reviewer. Another 35 reviews were eliminated in screening as irrelevant to higher education research and another 97 reviews were eliminated as their timeframes did not comply with the inclusion criteria (2010-2025) or the timeframe was not identified. The total of 128 reviews were finally selected for further analysis. The PRISMA flow-chart (Figure 1) depicts the whole identification and screening procedure.

Key Characteristics of Review Articles

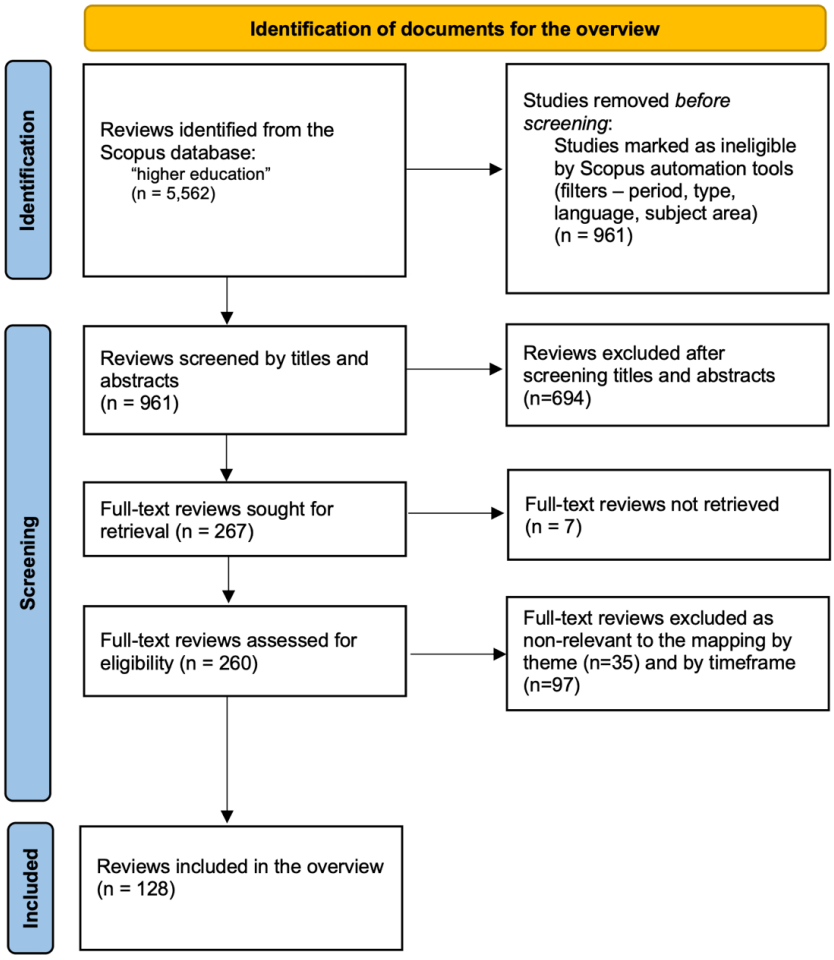
Demographics of the Sample

Eighty reviews out of the 128 reviews ultimately selected for the review were published in 2024, forty-eight came out in 2025. The most cited review in the sample is "A Systematic Review of Generative AI for Teaching and Learning Practice" (Ogunleye et al., 2024), with 61 citations in the Scopus database. Thirty-three reviews published in 2025 are not cited so far.

All 128 reviews are attributed to Social Sciences in compliance with the inclusion criteria. Simultaneously, the reviews are distributed among the following subject areas: Computer Sciences with 24 documents, Energy, Environmental Science, and Psychology 15 reviews each; Arts & Humanities

Figure 1

PRISMA Flow-Chart: Selection of Reviews



with 12 publications; Health Profession with 8 documents; 6 publications in Nursing; Decision Sciences and "Business, Management and Accounting" four reviews each; Engineering, Medicine, and Neuroscience with three reviews each. Other areas accounted for seven reviews.

The leading journals embrace Frontiers in Education (17 reviews), Cogent Education (13 reviews), Sustainability Switzerland (11 reviews), Education Sciences (6 reviews), Educational Research Review (4 reviews), Discover Education (3 reviews), Interactive Technology and Smart Education (3 reviews), International Journal of Learning Teaching and Educational Research (3 reviews), Journal of Language and education (3 reviews), Nurse Education Today (3 reviews), Sage Open (3 reviews), and Social Sciences and Humanities (3 reviews). The remaining 47 journals brought out fewer than three reviews. Of the 128 reviews in the dataset, 96 (75%) were published in journals ranked within the Scopus Social Sciences (Education) category. The quartile distribution of these is as follows: Q1 (23.6%), Q2 (36.7%), Q3 (7.8%), and Q4 (7.0%). The remaining journals, while not specifically categorized under Education, predominantly fall within other Social Sciences domains, most notably Social Sciences (miscellaneous) and Social Sciences (Geography, Planning and Development).

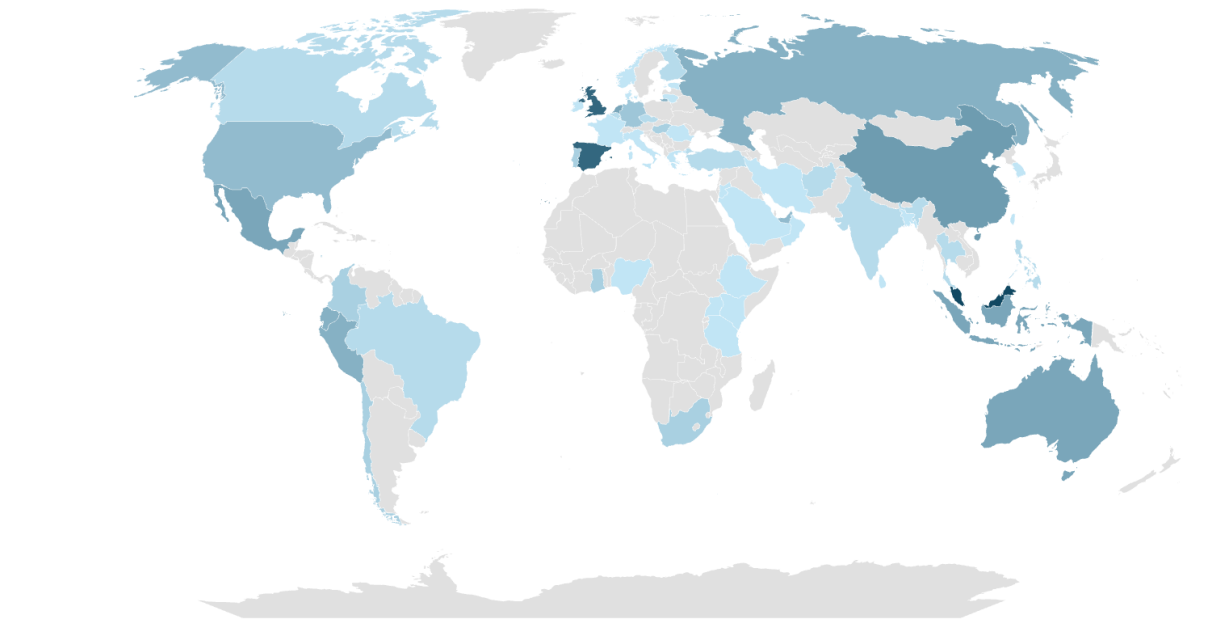
The most prolific authors in the sample were Bervell, B. (Mireku & Bervell, 2024; Mireku et al., 2024a; Mireku et al., 2024b), Mireku, D.O. (Mireku & Bervell, 2024; Mireku et al., 2024a; Mireku et al., 2024b), Raitskaya, L. (Raitskaya & Tikhonova, 2024; Raitskaya & Tikhonova, 2025; Tikhonova & Raitskaya, 2024), and Tikhonova, E. (Raitskaya & Tikhonova, 2024; Raitskaya & Tikhonova, 2025; Tikhonova & Raitskaya, 2024), with three reviews each. Six authors contributed two publications each: Alias, B.S., Barreiro-Collazo, A., Dzamesi, P.D., Glasserman-Morales, L.D., Oo, T.Z., and Tarake, T.G. The remaining researchers authored a single review.

The leading country by affiliations in the review is Malaysia (16 affiliations). Spain and the United Kingdom follow with 13 affiliations. China has 8 affiliations. Australia, Indonesia, Mexico, and Netherlands account for seven affiliations each. They are followed by Ecuador, Peru, and Russia, with six affiliations each. The geographical breakdown is shown in Figure 2. The top three of universities by affiliations embrace Tecnológico de Monterrey and RUDN University (six affiliated authors each) and Universiti Sains Malaysia (five affiliated authors). Consequently, the most conspicuous results that the reviews were conducted throughout the world, with a less impressive representation of Africa and Asia. International collaborations, involving researchers from two or three countries, accounted for 30 reviews (23.4%) in the sample.

Data pertaining to the methodological classification of the reviews were extracted and categorized. The distribution,

Figure 2

A Worldwide Breakdown of Review Authors' Affiliations



Note. The legend includes the density of publications per country, with a maximum of 16 (the darkest shade) to a minimum of 1 (the lightest shade). the map was made up in the microsoft excel by the authors

detailed in Figure 3, was as follows: systematic reviews (n=70, 55%), bibliometric reviews (n=15, 12%), scoping reviews (n=12, 9%), other review types (n=21, 16%), and reviews of an unidentified type (n=10, 8%).

The field's diverse and mixed methodological approaches were evident in the variety of review types identified. Alongside the dominant methodologies, the sample contained several less common types, as illustrated by the following examples: integrative (Amavasi & Zimmerman, 2024), exploratory (Barua & Lockee, 2024), narrative (Buele & Llerena-Aguirre, 2025), and realist (van der Wee et al., 2024) reviews, meta-ethnographic (Nieminen et al., 2024), pragmatic scoping (Marano et al., 2024), systematic scoping (Tikhonova & Raitskaya, 2024) among others (see Appendix 2 for full citations).

The reviews included in this scoping review demonstrated a strong adherence to established methodological standards, as evidenced by their consistent citation of authoritative guides and protocols (Table 3). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, particularly the seminal work by Moher et al. (2009), were the most frequently referenced framework, underscoring a prevailing commitment to transparency and rigor in systematic reviewing. For scoping reviews, the foundational framework by Arksey & O'Malley (2005) and its subsequent updates were commonly employed. Furthermore, authors utilized a diverse toolkit of specialized methodologies to suit their research questions, drawing on handbooks for meta-analysis (e.g., Schwarzer et al., 2015), bibliometric analyses (e.g., van Eck & Waltman, 2010; Donthu et al., 2021), and qualitative synthesis approaches like meta-ethnography (Noblit & Hare, 1988) and narrative synthesis (Popay et al., 2006). This comprehensive referencing reflects a mature and methodologically conscientious field that strategically selects and applies a wide spectrum of review typologies to synthesize higher education research.

Characteristics of the Reviews by Included Primary Publication

The analysis of the 128 reviews reveals a highly skewed distribution in the number of primary publications they synthesize, with a small number of large-scale reviews accounting for a substantial portion of the total evidence base (Appendix 4).

The data is characterized by a clear concentration, where a minority of reviews are responsible for synthesizing the majority of the primary literature. Specifically, just 11 reviews (8.6% of the sample) incorporate over 100 primary studies each, and together they account for 7,259 publications, or 43.9% of the total. This trend is further emphasized by the three largest reviews in the dataset, which each encompass over 500 primary studies and collectively contribute 3,990 publications (24.1% of the total).

In contrast, the typical review in this sample is far more limited in scope. The vast majority - 72.7% of the reviews (n=93) - synthesized fewer than 50 primary studies. This indicates that while the field is being consolidated through numerous focused, smaller-scale reviews, the overall evidence base is heavily influenced by a few exceptionally comprehensive studies that integrate knowledge from a thousand or more primary sources each.

Figure 3

Breakdown of the Reviews by Type

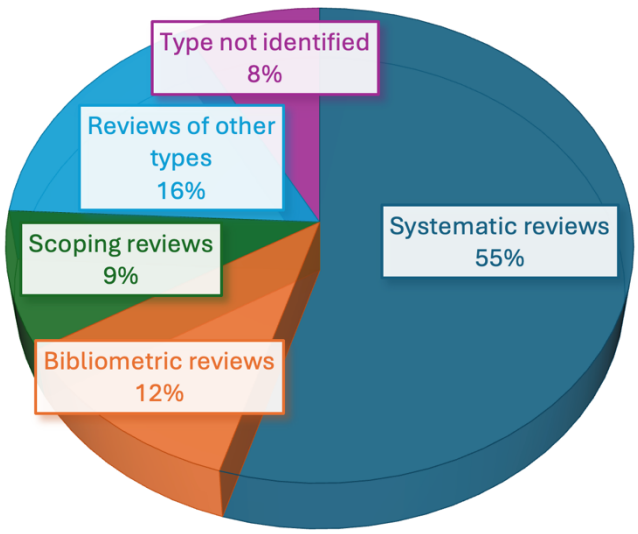


Table 3

References to Guides & Recommendations on Synthesis and Review Methodology, Frameworks, Protocols in the Reviews

Guides & Recommendations on Synthesis and Review Methodology, Frameworks, Protocols, etc.	References in the Reviews	
Systematic Reviews & recommendations on PRISMA	Abusaada & Elshater, 2022 Alexander, 2020 Bearman et al., 2012 Borrego et al., 2014 Briner & David, 2012 García-Peñalvo, 2017* Gough et al., 2017 Haddaway et al., 2022 Higgins & Green, 2011 Kitchenham & Charters, 2007 Kitchenham, 2004 Knobloch et al., 2011 Kraus et al., 2020 Marshall & Wallace, 2019 Moher et al., 2009 Moher et al., 2010 Moher et al., 2014 Moher et al., 2015	Moreno et al., 2018 Newman & Gough, 2020 Okoli, 2015 Page et al., 2016 Page et al., 2021 Papaioannou et al., 2016 Peters et al., 2020a Petticrew & Roberts, 2008 Pussegoda et al., 2017 Sánchez-Meca, 2022 Sarkis et al., 2021 Selçuk, 2019 Serrano et al., 2022* Sohrabi et al., 2021 Stansfield et al., 2016 Transfield et al., 2003 Urrútia et al., 2010* Xiao & Watson, 2019
Systematic Mapping Studies	Petersen et al., 2008	Petersen et al., 2015
Narrative synthesis in systematic reviews	Popay et al., 2006	Turnbull et al., 2022
Scoping Reviews	Arksey & O'Malley, 2005 Bradbury-Jones et al., 2022 Landa et l., 2011	Pollock et al., 2022 Tricco et al., 2018
Choice between Systematic and Scoping reviews	Munn et al., 2018	
Bibliometric Review	Aria & Cuccurullo, 2017 van Eck & Waltman, 2010 Belinchón Romero, 2008* Dávila et al., 2009*	Donthu et al., 2021 Gregorio-Chaviano, 2021* Joshi, 2014
Mapping Studies	Budgen et al., 2008	Novak, 1990
Literature Reviews	Fink, 2019	Templier & Pare, 2015
Meta-ethnography	France et al., 2019 Noblit & Hare, 1988	Noblit, 2019
Meta-analysis	Schwarzer et al., 2015	
Integrative Reviews	Toracco, 2005	Whittemore & Knafl, 2005
Realist Synthesis	Wong et al., 2013	

Populations of the Primary Publications

The analysis of the populations within the scoping review of 128 articles reveals a distinct concentration on specific stakeholder groups within the higher education landscape (Article 4). The field is overwhelmingly dominated by a focus on students, who constitute the central unit of analysis. This is evidenced by the fact that a significant majority of the reviews explicitly target various student cohorts. For instance, this encompasses general student populations, as well as more specific groups such as the medical students examined by Fengye et al. (2025), the student nurses in Amavasi & Zimmerman (2024), the STEM students studied by Bustamante-Mona et al. (2025), and students with disabilities, who are the focus of reviews like Nieminen et al. (2024) and Solis-Garcia et al. (2025).

Beyond the student body, a significant secondary focus is placed on educators and faculty, with numerous reviews examining the roles, practices, and development of teachers, lecturers, and HE academics. A smaller, yet notable, segment of the literature adopts a broader institutional or multi-stakeholder perspective. For example, Castillo-Martinez et al. (2024) analyzes the perspectives of "HE stakeholders," while Tikhonova & Raitskaya (2024) and Kaymakcioglu & Thomas (2024) focus on researchers and academics, respectively. Finally, a portion of the reviews, such as Barua & Lockee (2024), did not specify a particular population, in-

stead concentrating on theoretical or methodological trends within the literature itself. In summary, the intellectual focus of this corpus is firmly centered on the student experience, while also acknowledging the critical roles of academic staff and institutional frameworks in the ecosystem of higher education.

Higher Education Institutions as a Focus in the Primary Publications

The analysis of the institutional contexts within the scoping review reveals a strong predominance of generalist higher education institutions (HEIs) as the primary setting for research (Appendix 4). The vast majority of the reviews are focused on "HEI" as a broad category, indicating that the findings are intended to be applicable across universities and colleges without a specific professional or technical orientation. This general focus underscores a field concerned with universal themes in higher education. Alongside this dominant trend, a distinct and significant subset of reviews concentrates on specialized institutions. This is particularly evident in the health sciences, where a notable number of reviews, such as those by Amavasi & Zimmerman (2024) and Burton et al. (2024), are explicitly situated within nurse education institutions. Similarly, other reviews narrow their scope to specific academic units, including engineering faculties, medical universities, and STEM departments, as seen in works like Shahjahan & Seinn (2025) and Fengye et al. (2025). A small number of reviews also highlight institutions with a unique mission, such as those focusing on "green campuses" or entrepreneurship. In conclusion, while the literature is largely anchored in the context of the comprehensive university, a substantial and important thread of research is dedicated to understanding the unique dynamics and challenges within specialized professional and disciplinary educational environments.

Settings of the Primary Studies

An analysis of the geographical scope of the 128 reviews reveals a distinctly international character in the body of literature. The overwhelming majority of the reviews, comprising 99 articles or 77.3% of the sample, synthesized primary studies from an international setting. This indicates the field's strong emphasis on cross-border research and its pursuit of findings with global applicability.

Alongside this dominant global perspective, a notable secondary trend is the presence of regional studies, which account for 11 reviews (8.6%). These works concentrate on specific geographical or economic areas, such as developing countries or particular continents. Furthermore, a significant portion of the reviews, 17 in total (13.3%), are anchored in a national context, delving deeply into the higher education systems and policies of single countries.

A small number of the international reviews (1.6%) are supplemented with clarifying notes, such as "global, mainly Asian," indicating that while the scope is worldwide, the underlying evidence base may have a distinct geographical concentration. In summary, the field is overwhelmingly international in its outlook, yet maintains important, more focused lines of inquiry at both the regional and national levels.

Prevailing Themes

The co-occurrence analysis of author keywords from the 128 included reviews was conducted applying the VOSviewer software, based on full counting and minimum threshold set to three occurrences per keyword. The resulting visualisation (Figure 4) reveals nine initial groupings[3], automatically generated via VOSviewer's clustering algorithm based on keyword proximity and frequency.

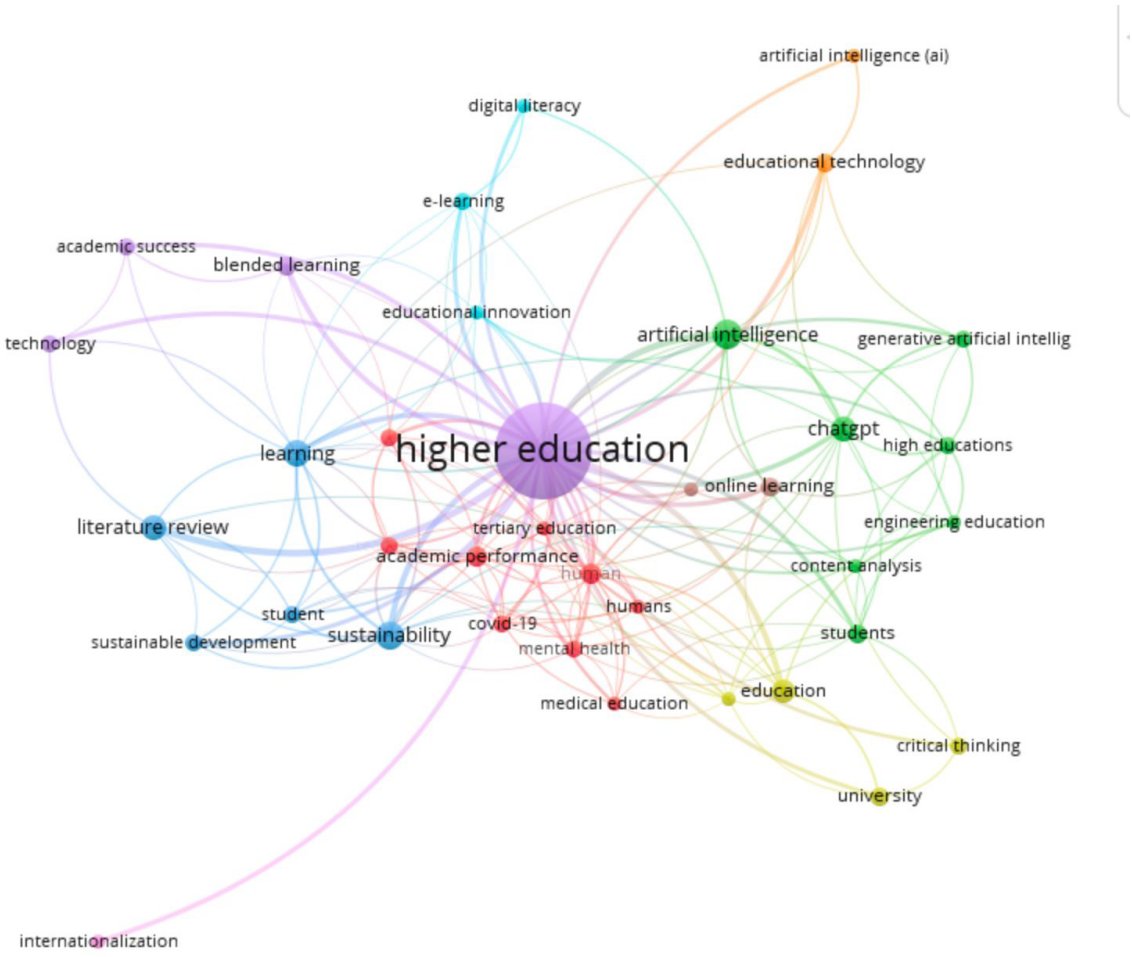
We coded all publications based on the clusters we extracted from the raw data (Appendix 1). The manually extracted clusters (codes) essentially coincided with groupings indicated by VOSviewer (Table 4). We categorised the reviews into eight clusters, or thematic codes (Table 5). The themes of the reviews partially overlapped, with a few entering more than one group. The biggest thematic group "*Digital literacy and educational innovation*" (Code 2) entails 40 reviews. The theme is wide in scope and scattered. Several reviews dwell upon digital competence, including the competence among higher education teachers (Karimi & Khawaja, 2025; Trujillo-Juárez et al., 2025), evaluation of digital competence (López-Nuñez et al., 2024), and a digital literacy gap in Sub-Saharan Africa universities (Ndibalema, 2025).

This cluster covers mobile-assisted language learning in teaching English in higher education (Álvarez-Martinez et al., 2025); teachers' behaviours and practices in various settings, including strategies fostering creativity of students (Brauer et al., 2024), lecturers' organizational commitment (Chen et al., 2024), and implementation of the Teachers Square concept suggested by Monica Vesely in 2001 as non-judgmental, evaluative process to enhance teaching (Gudoniene et al., 2025); challenge-based learning (Galdames-Calderón et al., 2024); use of technology-enhanced learning applications (Estaiba et al., 2025), self-regulated learning in the context of modern technologies (Junaštíkova, 2024), immersive learning platforms (Cabrera-Duffaut et al., 2024), virtual, augmented, and mixed reality (Balalee, 2025).

[3] *VOSviewer groupings based on the author keywords*: education, health and social challenges; artificial intelligence and transformation of educational practices; education, literature systemization and synthesis, and sustainable development; cognitive competencies and university education; academic success and blended learning; digital literacy and educational innovation; artificial intelligence in educational technologies; online learning and student engagement; and internationalization of higher education.

Figure 4

VOSviewer Visualization of Keyword Co-Occurrence and Thematic Groupings



Note. Circle size indicates keyword frequency; spatial distance reflects co-occurrence strength; colours represent cluster groupings by VOSviewer's modularity-based algorithm.

Table 4

Thematic Clusters (Codes) and VOSviewer Thematic Groupings

Code Number	Reviewers' Thematic Cluster (Code)	VOSviewer Thematic Groupings
1	Artificial intelligence in higher education (AI)	Artificial intelligence and transformation of educational practices Artificial intelligence in educational technologies
2	Digital literacy and educational innovation (DL)	Digital literacy and educational innovation
3	Education, health and social challenges (SC)	Education, health and social challenges
4	Cognitive competencies and university education (CC)	Cognitive competencies and university education
5	Online and blended learning in view of student engagement (OLBL)	Academic success and blended learning Online learning and student engagement
6	Education for sustainable development (ESD)	Education, literature systematization and synthesis, and sustainable development
7	Internationalization and transformation of higher education (INT)	Internationalization of higher education
8	Education: literature systematization and synthesis (SYN)	Education, literature systematization and synthesis, and sustainable development

Table 5

Thematic Distribution of Reviews

Thematic Cluster (Code)	Number of Reviews	Included Reviews (References)
1	24	Abdallah et al., 2025; Adauto Medina et al., 2024; Alotaibi, 2025; Andrade-Girón et al., 2025; Bayly-Castaneda et al., 2024; Buele & Llerena-Aguirre, 2025; Castillo-Martinez et al., 2024; Chashmyazdan et al., 2024; Chugh et al., 2025; Cui & Alias, 2024; Fernandez-Batanero et ak., 2024; Isiaku et al., 2024; Kalim et al., 2025; Klimova & Chen, 2024; Kovari, 2025; Lopez-Chila et al., 2024; Ma, 2025; Ogunleye et al., 2024; Peláez-Sánchez et al., 2024a; Raitskaya & Tikhonova, 2024; Raitskaya & Tikhonova, 2025; Sahar & Munawaroh, 2025; Schei et al., 2024; Tillmanns et al., 2025
2	41	Ahmad & Khurizan, 2024; Álvarez-Martinez et al., 2025; Balalee, 2025; Barua & Lockee, 2024; Brauer et al., 2024; Buele & Llerena-Aguirre, 2025; Cabrera-Duffaut et al., 2024; Cadiz, 2024; Castillo et al., 2024; Chen et al., 2024; Din Eak & Annamalai, 2024; Eltaiba et al., 2025; Galdames-Calderón et al., 2024; Harmon et al., 2024; Henry et al., 2024; Hidayat et al., 2024; Karimi & Khawaja, 2025; Liang et al., 2025; Lim & Lee, 2024; Liu et al., 2025; López-Nuñez et al., 2024; Lou & Zhou, 2024; Mashingaidze & Mayayise, 2025; Ndibalema, 2025; Olivares-De la fuente et al., 2025; Ortega-Ruiperez & Correa-Gorospe, 2024; Oulamine et al., 2025; Pallaris et al., 2024; Parambil et al., 2024; Peláez-Sánchez et al., 2024b; Renfors, 2024; Roy et al., 2025; Saez-Zevallos & Montalvo-Apolín, 2025; Sergeeva et al., 2024; Shahjahan & Seinn, 2025; Singun, 2025; Tang et al., 2024; Trujillo-Juárez et al., 2025; Vuoriainen et al., 2024; Zhao & Selvaratnam, 2024; Fernandez-Batanero et al., 2024
3	25	Amavasi & Zimmerman, 2024; Bannigan et al., 2025; Bustamante-Mona et al., 2025; Coleman et al., 2025; Dukes III et al., 2024; Edvardsen Tonheim et al., 2024; El Aatik et al., 2024; Fengye et al., 2025; Junaštíkova, 2024; Kalocsányiová et al., 2024; Kaymakcioglu & Thomas, 2024; Lim & Lee, 2024; Mireku et al., 2024a; Mireku et al., 2024b; Mursalzade et al., 2025; Ncube & Ngulube, 2024; Nieminen et al., 2024; Punch et al., 2025; Rahajeng et al., 2024; Rosales-Ricardo & Caceres-Manzano, 2024; Solis-Garcia et al., 2024; Solis-Garcia et al., 2025; Stamou et al., 2024; Wang & Ishak, 2025; Tikhonova & Raitskaya, 2024
4	7	Gomez et al., 2025; Marano et al., 2024; Mare & Mutezo, 2025; Mohammed & Ozdamli, 2024; Raitskaya & Tikhonova, 2025; Setiamurti & Kurniawati, 2024; Wang & Abdullah, 2024
5	15	Alonso et al., 2025; Alvarado, 2025; Barikzai et al., 2024; Benson et aL., 2024; Burton et al., 2024; Goncalves et al., 2024; Gudoniene et al., 2025; Ishmuradova et al., 2024; Lou & Zhou, 2024; Marano et al., 2024; Mare & Mutezo, 2025; Masalimova et al., 2024; Oulamine et al., 2025; Tareke et al., 2025; Valencia Quecano et al., 2024
6	12	Basheer et al., 2024; Bonilla-Jurado et al., 2024; Ghazian & Lortie, 2024; Liu et al., 2025; Ncube & Ngulube, 2024; ParedesCanencio et al., 2024; Rosario & Raimundo, 2024; Roy et al., 2025; Subki, 2025; van der Wee et al., 2024; Veres et al., 2025; You et al., 2024
7	14	Esteban, 2025; Ferreira Santos, 2024; Frez-Pulgar et al., 2025; Jaxin et al., 2024; Kaymakcioglu & Thomas, 2024; Marques et al., 2024; Mireku & Bervell, 2024; Sukjairungwattana et al., 2024; Tang et al., 2025; Tareke et al., 2024; Teng & Cosier, 2024; Tikhonova & Raitskaya, 2024; Vinueza-Morales et al., 2025; You et al., 2024
8	6	Banarjee et al., 2024; Heintalu et al., 2025; Mireku et al., 2024a; Nalweyiso et al., 2025; Thiedig & Wegner, 2024; Tikhonova & Raitskaya, 2024

Some focus is made on gamification and game-based learning (Galdames-Calderón et al., 2024; Hidayat et al., 2024; Henry et al., 2024).

A wide spread of mobile phones resulted in pervasive learning, in which students are placed in a real-world setting with smart phones and other wireless devices (Lim & Lee, 2024). The thematic cluster also spreads to barriers affecting e-learning (Oulamine et al., 2025), and barriers to digital transformation (Singun, 2025). In view of Education 4.0, the reviews synthesize education-industry collaboration (Vuoriainen et al., 2024), Internet of Things (Fernandez-Batanero et al., 2024) and the connections between AI and Education 4.0 (Peláez-Sánchez et al., 2024a).

As higher education is going through great transformation, causing great change in educational practices and learning technologies, the thematic cluster "*Artificial intelligence in higher education*" (Code 1) includes reviews ranging from general issues of Artificial Intelligence complex impact on higher education with all benefits and challenges (Abdallah et al., 2025; Buele & Llerena-Aguirre, 2025; Castillo-Martinez et al., 2024; Cui & Alias, 2024; Isiaku et al., 2024; Lopez-Chila et al., 2024; Ma, 2025; Ogunleye et al., 2024; Peláez-Sánchez et al., 2024a; Sahar & Munawaroh, 2025; Tillmanns et al., 2025) to diverse aspects of appliances of generative artificial intelligence, including cybersecurity (Parambil et al., 2024), AI-powered collaborative learning (Kovari, 2025), AI in academic work (Buele & Llerena-Aguirre, 2025), generative artificial intelligence in academic writing (Adauto Medi-

na et al., 2024; Raitskaya & Tikhonova, 2024), AI in fostering intercultural competence (ParedesCanencio et al., 2024), and critical thinking (Raitskaya & Tikhonova, 2025), and AI in learning management systems (Alotaibi, 2025).

The thematic cluster (Code 3) "*Education, health and social challenges*" entails aspects of education regarding the social function of university, health-related issues, medicine and nurse education (Amavasi & Zimmerman, 2024; Edvardsen Tonheim et al., 2024; Fengye et al., 2025), well-being features of higher education (Bannigan et al., 2025), inclusive education (Bustamante-Mona et al., 2025; Dukes III et al., 2024; Nieminen et al., 2024; Punch et al., 2025; Rahajeng et al., 2024; Solis-Garcia et al., 2024), students' migrant, minority, gender, displaced and refugee backgrounds in higher education (Coleman et al., 2025; Kaymakcioglu & Thomas, 2024; Kalocsányiová et al., 2024; Stamou et al., 2024). The thematic cluster also focuses on ethical issues of academic activities, including plagiarism (Mireku et al., 2024a, examination malpractice behaviours (Mireku et al., 2024b), ethical imperatives of data privacy (Ncube & Ngulube, 2024), and ethical aspects of research (Tikhonova & Raitskaya, 2024). The cluster (Code 3) totals 25 reviews.

The thematic cluster "*Cognitive competencies and university education*" (Code 4) works out at six reviews. As cognitive competence as other meta-competences factor in self-regulated learning and academic success of students, this theme attracts researchers' attention. Critical thinking as higher order thinking skill gets most focus (Gomez et al., 2025; Raitskaya & Tikhonova, 2025). Cognitive competence is also approached through shared metacognition in communities of inquiry in online learning (Mare & Mutezo, 2025). In some research, cognitive competence is considered within soft skills (Gomez et al., 2025; Mohammed & Ozdamli, 2024). Critical thinking is studied in the context of artificial intelligence impact (Raitskaya & Tikhonova, 2025) and through Mathematics (Wang & Abdullah, 2024). One more direction of study is creativity that is regarded as a metacognitive process (Setiamurti & Kurniawati, 2024).

The next thematic cluster titled "*Online and blended learning in view of student engagement*" (Code 5) has been popular for most of the 21st century, with a spike during the COVID-19 Pandemic (Kumar & Pande, 2021). Though, the timeframe of this synthesis includes reviews published in 2024-2025, the timeline of the included reviews is limited to 2010-2025. Thus, the Pandemic practices were reflected on in several reviews (Burton et al., 2024; Gudoniene et al., 2025; Lou & Zhou, 2024; Oulamine et al., 2025). Online learning and e-learning is the topic of a few reviews (Mare & Mutezo, 2025; Masalimova et al., 2024; Barikzai et al., 2024; Burton et al., 2024), with more reviews of blended learning (Alonso et al., 2025; Benson et aL., 2024; Ishmuradova et al., 2024; Lou & Zhou, 2024). This thematic cluster also covers hybrid learning (Gudoniene et al., 2025) and remote proctoring (Marano et al., 2024).

Sustainable Development Goals (SDGs) accepted in 2015 included Goal 4 - Quality Education. The focus on the topic reached its peak after 2015 but the theme is still presented in educational research. The thematic cluster "*Education for sustainable development*" (Code 6) totals 12 reviews. They dwell upon assessment of sustainability in higher education (Basheer et al., 2024), benefit of education for sustainable development at universities (Bonilla-Jurado et al., 2024; Veres et al., 2025), in incorporating the UN Sustainable Development Goals (SDGs) into Environmental and Natural Science Courses (Ghazian & Lortie, 2024), contribution of students' entrepreneurship competence to Sustainable Development (Liu et al., 2025; Rosario & Raimundo, 2024), carbon footprint of higher education institutions (ParedesCanencio et al., 2024) and green campus management (Subki, 2025).

Another thematic cluster titled "*Internationalization and transformation of higher education*" (Code 7) is rather wide and entails various aspects of transformations of higher education, including impact of internationalization issues. The cluster consists of 14 reviews. Some researchers align digital transformation with internationalization (Ferreira Santos, 2024). Aspects of student mobility still forms an integral part of internationalization of higher education (Esteban, 2025). As internationalization leads to a global educational dimension, issues of global competence (Jaxin et al., 2024; Teng & Cosier, 2024) and regional or national differences (Sukjairungwattana et al., 2024; Tang et al., 2025; Tareke et al., 2024) arise. Higher education change at large highlights academic identities (Marques et al., 2024).

Distinguished by its meta-research focus, the thematic cluster "*Education: literature systematization and synthesis*" (Code 8) encompasses methodological and literature or review aspects of educational inquiry. The six reviews comprising this cluster examine a range of foundational issues, such as conceptual frameworks for goal orientation (Heintalu et al., 2025), methodological inconsistencies in reporting STEM workforce skills (Banarjee et al., 2024), and the application of evidence in both teaching practice (Nalweyiso et al., 2025) and institutional decision-making (Thiedig & Wegner, 2024). Additional reviews explore the culture of research

(Tikhonova & Raitskaya, 2024) and trends in higher education leadership development (You et al., 2024).

Gaps and Priorities for Further Research

Of the 128 reviews analysed, 115 included recommendations for further research, which were primarily located in the conclusion (mainly) or discussion (rarely) sections. We catalogued the specific wordings of these agendas (Appendix 3) and subsequently categorized the recommendations into six overarching themes.

Technology Integration & Emerging Tools

This theme encompasses research on the implementation, effectiveness, and ethical implications of new technologies like AI, VR/AR, and digital platforms in education. The sub-theme of Long-term & Longitudinal Impact focuses on assessing the long-term effects on learning, critical thinking, career readiness, and skill retention, as seen in the work of Abdallah et al., 2025; Alotaibi, 2025; Balalee, 2025; Cabrera-Duffaut et al., 2024; Eltaiba et al., 2025; Jaxin et al., 2024; Liu et al., 2025; Nalweyiso et al., 2025; Oulamine et al., 2025; Pallaris et al., 2024; Punch et al., 2025; Raitskaya & Tikhonova, 2025; Vinueza-Morales et al., 2025; and Wang & Ishak, 2025. Research on AI & Generative AI, such as that by Girón et al., 2025; Chugh et al., 2025; Cui & Alias, 2024; Isiaku et al., 2024; Kovari, 2025; Ma, 2025; Ogunleye et al., 2024; Tillmanns et al., 2025; and Trujillo-Juárez et al., 2025, seeks to explore discipline-specific applications, develop ethical frameworks and guidelines, evaluate effectiveness compared to other tools, and investigate impacts on critical thinking and student dependency.

The sub-theme of Immersive Technologies (VR, AR, MR) is concerned with determining long-term applications and effectiveness across different subject domains like STEM, humanities, and health, as investigated by Balalee, 2025 and Cabrera-Duffaut et al., 2024. Studies on Digital Platforms & Tools, including those by Lim & Lee, 2024; López-Nuñez et al., 2024; and Olivares-De la fuente et al., 2025, investigate the use of specific tools like Twitter and YouTube, pervasive tools, and the impact of digital competence. Finally, the critical area of Cybersecurity & Data Privacy involves research aimed at developing benchmark datasets for cybersecurity in online education and exploring the efficacy of privacy-preserving techniques, as highlighted in the works of Ncube & Ngulube, 2024 and Parambil et al., 2024.

Pedagogy, Curriculum & Skill Development

This theme focuses on teaching methods, curriculum design, and the development of specific competencies in students and educators. A significant area of inquiry involves Teaching Methodologies & Frameworks, which investigates the effectiveness of specific pedagogical approaches like challenge-based learning (CBL), team-based learning (TBL), gamification, and translanguaging, as examined by Benson et al., 2024; Burton et al., 2024; Galdames-Calderón et al., 2024; Henry et al., 2024; Renfors, 2024; Sergeeva et al., 2024; and Tang et al., 2024. Closely related is the sub-theme of Skill Development, which focuses on fostering essential competencies in learners such as critical thinking, creativity, self-regulated learning, entrepreneurial competence, and digital literacy, a priority in the research of Brauer et al., 2024; Heintalu et al., 2025; Karimi & Khawaja, 2025; Liu et al., 2025; Lou & Zhou, 2024; Ndibalema, 2025; Peláez-Sánchez et al., 2024; and Wang & Abdullah, 2024.

Furthermore, the theme encompasses Curriculum Integration, exploring the systematic embedding of cross-cutting priorities like sustainability (SDGs), sustainable entrepreneurship, and evidence-based practice into curricula, as seen in the work of Bonilla-Jurado et al., 2024; Ghazian & Lortie, 2024; Nalweyiso et al., 2025; and Rosario & Raimundo, 2024. Finally, recognizing the central role of the educator, research on Faculty Development & Academic Identity examines the formation of academic identities, the enhancement of digital competencies among teachers, and the overall well-being of faculty, addressed by Harmon et al., 2024; López-Nuñez et al., 2024; and Marques et al., 2024.

Equity, Inclusion & Diverse Contexts

This theme highlights the critical need for educational innovations to be inclusive, accessible, and effective across diverse populations, cultures, abilities, and geographic regions. A primary focus is on Inclusive Education & Accessibility, which involves investigating robust support mechanisms for students with disabilities (including non-visible disabilities) and promoting inclusive pedagogical methods and peer support programs, as researched by Bustamante-Mona et al., 2025; Dukes III et al., 2024; Punch et al., 2025; Rahajeng et al., 2024; and Solis-Garcia et al., 2025. Further refining this focus, the sub-theme of Gender & Demographic Studies examines the complex relationship between technology adoption and factors like gender, age, and economic situation, with particular attention to underrepresented regions, as seen in the work of Bannigan et al., 2025; Bustamante-Mona et al., 2025; Coleman et al., 2025; and Kalim et al., 2025.

To counter a Western-centric bias, the theme emphasizes Cross-Cultural & Global Perspectives, calling for research conducted in diverse cultural, regional, and transnational contexts, including the Global South and non-Anglophone countries, exemplified by studies from Alonso et al., 2025; Chen et al., 2024; Kaymakcioglu & Thomas, 2024; Álvarez-Martinez et al., 2025; Setiamurti & Kurniawati, 2024; Shahjahan & Seinn, 2025; Sukjairungwattana et al., 2024; Tang et al., 2025; and Vuoriainen et al., 2024. Finally, ensuring that no cluster is overlooked, research on Support for Specific Populations concentrates on the unique needs of refugee-background students and the factors influencing their success throughout the entire higher education journey, a key area of investigation for Kalocsányiová et al., 2024.

Sustainability & Institutional Management

This theme addresses the environmental, social, and governance dimensions of educational institutions, focusing on their long-term viability and impact. A core area of investigation is Green Campus & Sustainable Development, which delves into the behavioral, cultural, and policy-related dimensions of green campus management and works to as-

sess the carbon footprint of Higher Education Institutions (HEIs), as explored by Basheer et al., 2024; Paredes-Canencio et al., 2024; Subki, 2025; Tareke et al., 2024; and Veres et al., 2025. The operational effectiveness of these institutions falls under Quality Assurance & Governance, which involves research on promoting quality assurance, understanding the impact of legislative frameworks on digital transformation, and enhancing the use of evidence in institutional decision-making, a key focus for Mireku & Bervell, 2024; Singun, 2025; and Thiedig & Wegner, 2024. Finally, bridging theory with practice, the theme explores Living Labs & Interdisciplinary Collaboration, examining how Living Labs function as hybrid learning spaces that merge education with real-world application and assessing the impact of interdisciplinary education on achieving tangible sustainability outcomes, as studied by van der Wee et al., 2024 and Veres et al., 2025.

Assessment, Ethics & Academic Integrity

This theme centers on the critical challenges of evaluating learning effectively, navigating the ethical implications of new educational tools, and preserving foundational academic standards. A key area of development is in Assessment Methods & Feedback, which involves the creation and evaluation of innovative automated assessment tools, such as AI for grading, alongside investigations into the effectiveness of feedback across diverse learning styles and explorations of how assessment practices themselves shape student identities, as examined by Bui & Barrot, 2024; Din Eak & Annamalai, 2024; Lu et al., 2024; Makarova et al., 2024; and Nieminen et al., 2024. In response to technological advancements, the theme critically addresses Ethics & Academic Integrity, focusing on studying the long-term effects of Generative AI on academic honesty, developing robust ethical guidelines and policies for AI use in educational settings, and investigating related issues such as plagiarism detection and technostress among students and educators, a priority in the research of Bayly-Castaneda et al., 2024; Adauto Medina et al., 2024; Mireku et al., 2024b; and Tillmanns et al., 2025.

Research Gaps & Methodological Approaches

This theme critically identifies limitations within current research practices and advocates for more comprehensive, rigorous, and diverse methodological strategies in future scholarly inquiry. A primary recommendation involves Broadening Research Scope, which entails utilizing multiple literature databases, incorporating multilingual sources to mitigate Anglophone bias, expanding inclusion criteria regarding time periods and publication types, and considering a wider range of stakeholder perspectives, as suggested by Ahmad & Khurizan, 2024; Esteban, 2025; Ferreira Santos, 2024; Gudoniene et al., 2025; Mashingaidze & Mayayise, 2025; Ortega-Ruiperez & Correa-Gorospe, 2024; Raitskaya & Tikhonova, 2024; Sahar & Munawaroh, 2025; Stamou et al., 2024; and Tikhonova & Raitskaya, 2024.

To enhance the depth and validity of findings, the theme calls for the adoption of Rigorous & Diverse Methodologies, specifically promoting the use of mixed-methods approaches, longitudinal studies, in-depth case studies, and experimental designs to build a more robust and reliable evidence base, a approach exemplified by Andrade-Girón et al., 2025; Karimi & Khawaja, 2025; Marques et al., 2024; Masalimova et al., 2024; and Tareke et al., 2025. Finally, to adequately address the complex nature of contemporary educational challenges, the theme underscores the necessity of Interdisciplinary Collaboration, encouraging active partnership between traditionally siloed fields such as education, computer science, ethics, and psychology, as highlighted in the works of Frez0Pulgar et al., 2025; Isiaku et al., 2024; Ma, 2025; Ogunleye et al., 2024; Parambil et al., 2024; and Vinueza-Morales et al., 2025.

To integrate the coded themes from our analysis and the research agendas from the reviews, we analyzed their relationship, resulting in the following consolidated list (Table 6). The distribution displayed the interrelation of the themes and complexity of educational research.

Implications of the Reviews Included in the Scoping Review

An analysis of the results of the coding for implications, where 2 denotes a dedicated implications section, 1 indicates implications embedded within the Discussion or Conclusion, and 0 signifies no stated implications, reveals a significant gap in explicit reporting. Only a minority of studies feature implications clearly stated in a separate section (coded as 2[4]), while a larger portion embed them within other sections (coded as 1[5]). However, nearly half of the reviewed literature is coded as 0, failing to state any implications at all. This prevalent

[4] *Implications extracted from the reviews* with Code 2 (Implications in a dedicated section): Alonso et al., 2025; Amavasi & Zimmerman, 2024; Barikzai et al., 2024; Bonilla-Jurado et al., 2024; Brauer et al., 2024; Buele & Llerena-Aguirre, 2025; Chen et al., 2024; Chugh et al., 2025; Coleman et al., 2025; Cui & Alias, 2024; Frez0Pulgar et al., 2025; Ghazian & Lortie, 2024; Harmon et al., 2024; Isiaku et al., 2024; Kalim et al., 2025; Karimi & Khawaja, 2025; Klimova & Chen, 2024; Ma, 2025; Mare & Mutezo, 2025; Masalimova et al., 2024; Mireku & Bervell, 2024; Mireku et al., 2024a; Mireku et al., 2024b; Nalweyiso et al., 2025; Ndibalema, 2025; Nieminen et al., 2024; Ogunleye et al., 2024; Oulamine et al., 2025; Rahajeng et al., 2024; Raitskaya & Tikhonova, 2025; Roy et al., 2025; Sahar & Munawaroh, 2025; Schei et al., 2024; Thiedig & Wegner, 2024.

[5] *Implications extracted from the reviews* with Code 1 (Implications within Discussion/Conclusion): Alotaibi, 2025; Bannigan et al., 2025; Basheer et al., 2024; Burton et al., 2024; Cabrera-Duffaut et al., 2024; Din Eak & Annamalai, 2024; El Aatik et al., 2024; Eltaiba et al., 2025;

Table 6

Consolidated List of Prevailing Research Themes and Themes Further Studies in Higher Education

Number of the theme	Themes for Further Research	Related Coded Clusters
1	Technology Integration & Emerging Tools	Code 1: Artificial Intelligence in Higher Education (partly)
		Code 2: Digital Literacy and Educational Innovation (partly)
2	Pedagogy, Curriculum & Skill Development	Code 4: Cognitive Competencies and University Education (partly)
3	Equity, Inclusion & Diverse Contexts	Code 3: Education, Health and Social Challenges (partly)
4	Sustainability & Institutional Management	Code 6: Education for Sustainable Development (partly)
		Code 7: Internationalisation and Transformation of Higher Education (partly)
5	Assessment, Ethics & Academic Integrity	Code 3: Education, Health and Social Challenges (partly)
		Code 5: Online and Blended learning in View of Student Engagement (partly)
6	Research Gaps & Methodological Approaches	Code 8: Education: Literature Systematization and Synthesis (partly)
7	Online and Blended learning in View of Student Engagement	Code 5: Online and Blended learning in View of Student Engagement (completely). The theme aligns with broader "Technology Integration" and "Pedagogy" themes
8	Education, Health and Social Challenges	Code 3: Education, Health and Social Challenges (completely). The theme aligns with the "Equity, Inclusion & Diverse Contexts" and "Ethics"
9	Cognitive Competencies and University Education	Code 4: Cognitive Competencies and University Education (completely). The theme aligns with "Pedagogy, Curriculum & Skill Development"
10	Internationalisation and Transformation of Higher Education	Code 7: Internationalisation and Transformation of Higher Education (completely). The theme aligns with "Sustainability & Institutional Management"
11	Education for Sustainable Development	Code 6: Education for Sustainable Development (completely), being a subset of "Sustainability & Institutional Management"
12	Education: Literature Systematisation and Synthesis	Code 8: Education: Literature Systematization and Synthesis (completely), being a subset of "Research Gaps & Methodological Approaches"

absence of a dedicated implications section results in vague conclusions and leaves the evidence base for practical application seeming doubtful and underdeveloped. For future research to provide more explicit and actionable guidance, the inclusion of a distinct implications section is strongly recommended to clarify the translational value and real-world impact of the findings.

DISCUSSION

This scoping review of reviews reveals a dynamic and rapidly evolving field of higher education research, characterized by a maturation in methodology, a complex and interconnected thematic landscape, and a distinct duality in the structure of its evidence synthesis (Figure 3; Table 3). A central finding is the field's methodological conscientiousness. The widespread application of established protocols like PRISMA and Arksey & O'Malley's framework, coupled with a sophisticated diversity of review typologies (from bibliometric to meta-ethnographic) signals a move beyond simple literature summaries towards rigorous, transparent, and nuanced forms of evidence aggregation. This methodological maturity strengthens the credibility of the field's collective findings and provides a robust foundation for addressing its complex research questions.

Galdames-Calderón et al., 2024; Jaxin et al., 2024; Kaymakcioglu & Thomas, 2024; Kovari, 2025; Liang et al., 2025; Liu et al., 2025; Marano et al., 2024; Mashingaidze & Mayayise, 2025; Mohammed & Ozdamli, 2024; Mursalzade et al., 2025; Ortega-Ruiperez & Correa-Gorospe, 2024; Parambil et al., 2024; ParedesCanencio et al., 2024; Peláez-Sánchez et al., 2024; Pelaez-Sanchez et al., 2024; Punch et al., 2025; Renfors, 2024; Rosario & Raimundo, 2024; Singun, 2025; Tang et al., 2024; Tareke et al., 2025; Teng & Cosier, 2024; Quecano et al., 2024; van der Wee et al., 2024; Veres et al., 2025.

The thematic architecture of the field reflects this complexity, dominated by the pervasive force of digital integration. The "wide and scattered" nature of the "Digital literacy and educational innovation" cluster shows its influence from pedagogy to institutional strategy. This broad transformation is now being defined by the rapid rise of "Artificial intelligence" as a major, distinct sub-field, as evidenced by a cluster of reviews with overlapping scopes and objectives (Abdallah et al., 2025; Buele & Llerena-Aguirre, 2025; Castillo-Martinez et al., 2024; Ogunleye et al., 2024). The breakneck pace of technological change has led to a compressed synthesis cycle, where reviews covering periods as short as a year quickly follow one another. The significant focus on AI in academic writing, noted in syntheses like those by Adauto Medina et al. (2024) and Raitskaya & Tikhonova (2024) despite our review's exclusion criteria, underscores its status as a rapidly maturing area of inquiry. This technological focus, however, exists in a necessary balance with strong clusters around social challenges, equity, and sustainable development, illustrating a field deeply engaged with its ethical responsibilities.

This interconnectedness is further clarified by the structured mapping of themes for further research and corresponding clusters/ codes (Table 6), which demonstrates that key issues like digital literacy (Cluster/ Code 2) or ethics (Cluster/Code 3) are not siloed but are integral components across technological, pedagogical, and social dimensions. For instance, the theme of "Online and Blended learning" (Theme 7) is a direct instantiation of a specific code but also contributes to broader discussions on "Technology Integration" (Theme 1) and "Pedagogy" (Theme 2). This structure supports an interdisciplinary approach, suggesting that the most significant advances may occur at the intersection of these categories. Furthermore, the inclusion of meta-level themes like "Research Gaps & Methodological Approaches" (Theme 6) indicates a self-reflective field that is systematically critiquing its own development.

Underpinning this thematic and methodological landscape is a dual structure of evidence synthesis. The field is populated by a large number of focused, small-scale reviews (nearly three-quarters synthesizing fewer than 50 studies) that provide detailed charts of niche topics. Conversely, the overarching evidence base is disproportionately shaped by a small number of monumental works (just 11 studies account for over 40% of the total primary publications analyzed), which provide broad maps of the territory. This duality presents a central challenge: integrating these different levels of analysis to form a coherent narrative.

Based on the analysis of the included reviews, the breakdown of how implications are presented in the reviews reveals a significant gap in explicit reporting: a minority of studies (approximately 30%) feature implications stated in a dedicated section, while a larger portion (approximately 25%) embed implications within the Discussion or Conclusion sections. However, nearly half of the reviewed literature fails to state any practical or research implications at all. This absence of a dedicated implications section often results in vague conclusions and leaves the evidence base for practical application seeming doubtful and underdeveloped.

Finally, a tension exists between the field's aspirations and its contextual grounding. The predominantly international scope of the reviews and their overwhelming focus on the student experience aim for global applicability. However, this evidence is often grounded in "generalist" HEI contexts, raising questions about its transferability to specialized professional schools. While the significant representation of countries like Malaysia, Spain, and the UK suggests a shifting geographical centre of gravity beyond Anglo-American dominance, the under-representation of regions like Africa and Asia points to persistent gaps in the global knowledge ecosystem. To end it up, the trajectory of higher education research will be shaped by its ability to integrate its dualistic evidence base, navigate the interplay between technological capability and ethical commitment, and broaden its geographical and institutional scope to build a truly inclusive and applicable body of knowledge.

CONCLUSION

The findings of our systematic scoping review of reviews of higher education show a field dominated by rigorous, protocol-driven methodologies; 76% of the reviews were systematic, bibliometric, or scoping in nature. While this demonstrates a broad commitment to evidence-based frameworks, a minority of studies utilized narrative or other qualitative methodologies.

The analysis identified eight prevalent thematic clusters in higher education research: (1) AI in higher education; (2) digital literacy and innovation; (3) education, health, and social change; (4) cognitive competencies; (5) student engagement in online/blended learning; (6) education for sustainable development; (7) internationalisation and transformation; and (8) literature systematisation and synthesis. Our review synthesized the research agendas proposed across the reviewed literature: (a) technology integration & emerging tools; (b) pedagogy, curriculum & skills development; (c) equity, inclusion & diverse contexts; (d) sustainability & institutional management; (e) assessment, ethics & academic integrity; (f) research gaps & methodological approaches.

Implications for Researchers

The review outlined the prevailing themes in research on higher education that entail (1) research on AI appliances in higher education and transformations which AI may lead to; (2) digital and AI literacy and competencies essential for facing innovations in higher education and society; (3) so-

cial function of universities, issues of medicine and nurse education, health-related aspects of higher education and inclusive education; (4) metacognition, with cognitive and creative competencies as the frontrunners for self-regulated life-long learning and knowledge society; (5) online, blended, hybrid and e-learning; virtual, augmented and mixed reality in higher education; (6) education for sustainable development; (7) transformation of higher education and internationalization, global issues of higher education; (8) epistemological, methodological ans. evidence-based aspects of educational research. For future research to provide more explicit and actionable guidance, the inclusion of a distinct implications section is strongly recommended to clarify the translational value and real-world impact of the findings.

Limitations and Further Research

The valuable insights from this review of higher education research themes must be considered alongside its methodological constraints. These include the potential omission of relevant reviews due to the search being limited to one database. The restriction to English-language publications may have excluded relevant insights from regional and national reviews, potentially limiting the geographical and contextual diversity of the synthesized evidence. To address these limitations and shed further light on the field's evolution, a subsequent study could employ an overview of reviews methodology. This would involve a longer time span, searches in multiple databases, and the inclusion of literature in other languages.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTIONS

Elena Tikhonova: conceptualization; data curation; formal analysis; investigation; methodology; resources; software; validation; visualization; writing – original draft; writing – review & editing.

Lilia Raitskaya: conceptualization; data curation; formal analysis; investigation; methodology; resources; software; validation; visualization; writing – original draft; writing – review & editing.

REFERENCES

Abdallah, N., Katmah, R., Khalaf, K., & Jelinek, H. F. (2025). Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration. *Social Sciences & Humanities Open, 12*, Article 101866. https://doi.org/10.1016/j.ssaho.2025.101866

Abusaada, H., & Elshater, A. (2022). Notes on developing research review in urban planning and urban design based on PRISMA statement. *Social Sciences*, 11(9), 391. https://doi.org/10.3390/socsci11090391

Adauto Medina, W. A., Flores Velásquez, C. H., Olivares Zegarra, S. del R., Morales Romero, G. P., Quispe Andia, A., Aybar Bellido, I. E., Palacios Huaraca, C. R., Aldana Trejo, F. H., & Alvarado Bravo, N. (2024). Using ChatGPT in university academic writing: A bibliometric review study on the implications for writing reports, papers, essays, and theses. *International Journal of Learning, Teaching and Educational Research, 23*(12), 360–381. https://doi.org/10.26803/ijlter.23.12.19

Ahmad, A. M., & Khurizan, N. S. N. (2024). Data envelopment analysis and higher education: A systematic review of the 2018–2022 literature and bibliometric analysis of the past 30 years of literature. *SAGE Open*, *14* (3), 1-28. https://doi.org/10.1177/21582440241245678

Alexander, P. A. (2020). Methodological guidance paper: The art and science of quality systematic reviews. *Review of Educational Research, 90*(1), 6-23. https://doi.org/10.3102/0034654319854352

Alzahmi, R. A., Syed, R. T., Singh, D., Arshi, T. A., & Kutty, S. V. (2025). Organizational change in higher education institutions: Thematic mapping of the literature and future research agenda. *Humanities and Social Sciences Communications*, *12*(282). https://doi.org/10.1057/s41599-025-05650-w

Alonso, R. K., Velez, A., Diego Mantecón, M. Á., & Heredia Oliva, E. (2025). Blended learning in higher education for the development of intrinsic motivation: A systematic review. *Cogent Education, 12*(1), Article 2487572. https://doi.org/10.1080/2331186X.2025.2487572

Alotaibi, N. S. (2024). The impact of AI and LMS integration on the future of higher education: Opportunities, challenges, and strategies for transformation. *Sustainability, 16*(23), Article 10357. https://doi.org/10.3390/su162310357

Alvarado, L. F. (2025). Design thinking as an active teaching methodology in higher education: A systematic review. *Frontiers in Education, 10*, Article 1462938. https://doi.org/10.3389/feduc.2025.1462938

Álvarez Martínez, J. A., Gómez, J. F., & Restrepo Botero, J. C. (2025). Mobile-assisted language learning (MALL) for improving English language teaching and learning in higher education: A systematic literature review. *Knowledge Management & E-Learning, 17*(2), 258–277. https://doi.org/10.34105/j.kmel.2025.17.012

Amavasi, B., & Zimmerman, P.-A. (2024). Infection prevention and control continuous education and training in pre-registration nursing programmes. *Nurse Education Today, 133*, Article 106051. https://doi.org/10.1016/j.nedt.2023.106051

Andrade-Girón, D., Marín-Rodriguez, W., Sandivar-Rosas, J., Carreño-Cisneros, E., Susanibar-Ramirez, E., Zuñiga-Rojas, M., Angeles-Morales, J., & Villarreal-Torres, H. (2024). Generative artificial intelligence in higher education learning: A review based on academic databases. *Iberoamerican Journal of Science Measurement and Communication, 4*(1), 1–16. https://doi.org/10.47909/ijsmc.101

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

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*, 19–32. http://doi.org/ 10.1080/1364557032000119616

Balalle, H. (2025). Learning beyond realities: exploring virtual reality, augmented reality, and mixed reality in higher education - A systematic literature review. *Discover Education, 4*(151). https://doi.org/10.1007/s44217-025-00559-7

Banerjee, P., Graham, L., & Given, G. (2024). A systematic literature review identifying inconsistencies in the inclusion of subjects in research reports on STEM workforce skills in the UK. *Cogent Education, 11*(1), Article 2288736. https://doi.org/10.1080/2331186X.2023.2288736

Bannigan, G., McGrath, D., & Matthews, J. (2025). Whole-university approaches to embedding well-being in the curriculum: A scoping review. *Frontiers in Education, 10*, Article 1534244. https://doi.org/10.3389/feduc.2025.1534244

Barikzai, S., Bharathi S, V., & Perdana, A. (2024). Challenges and strategies in e-learning adoption in emerging economies: A scoping review. *Cogent Education, 11*(1), Article 2400415. https://doi.org/10.1080/2331186X.2024.2400415

Barua, L., & Lockee, B. B. (2024). A review of strategies to incorporate flexibility in higher education course designs. *Discover Education, 3* (127). https://doi.org/10.1007/s44217-024-00213-8

Basheer, N., Ahmed, V., Bahroun, Z., & Anane, C. (2024). Exploring sustainability assessment practices in higher education: A comprehensive review through content and bibliometric analyses. *Sustainability, 16*(14), Article 5799. https://doi.org/10.3390/su16145799

Bayly-Castaneda, K., Ramirez-Montoya, M.-S., & Morita-Alexander, A. (2024). Crafting personalized learning paths with AI for lifelong learning: A systematic literature review. *Frontiers in Education*, *9,* Article 1424386. https://doi.org/10.3389/feduc.2024.1424386

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research and Development*, *31*(5), 625–640. https://doi.org/10.1080/07294360.2012.702735

Belinchón Romero, I. (2008). La bibliometría como herramienta de conocimiento de la situación de la investigación clínica espa nola en dermatología y sus implicaciones futuras [Bibliometrics as a tool for understanding the state of Spanish clinical research in dermatology and its future implications]. *Actas Dermo-Sifiliogr*, 109, 2. https://doi.org/10.1016/j.ad.2017.09.002

Benson, H., Williams, K., & Heggart, K. (2024). Quality assurance interventions in blended learning design: A systematic review of the literature. *Australasian Journal of Educational Technology*, *40*(5), 31–46. https://doi.org/10.14742/ajet.9362

Bonilla-Jurado, D., Zumba, E., Lucio-Quintana, A., Yerbabuena-Torres, C., Ramírez-Casco, A., & Guevara, C. (2024). Advancing university education: Exploring the benefits of education for sustainable development. *Sustainability*, *16*(17), Article 7847. https://doi.org/10.3390/su16177847

Borrego, M., Foster, M. J., & Froyd, J. E. (2014). Systematic Literature Reviews in Engineering Education and Other Developing Interdisciplinary Fields*. Journal of Engineering Education*, 103 (1), 45–76. https://doi.org/10.1002/jee.20038Briner, R., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. Oxford University Press, 112–129.

Bradbury-Jones, C., Aveyard, H., Herber, O.R., Isham, L., Taylor, J., & O'Malley, L. (2022). Scoping reviews: the PAGER framework for improving the quality of reporting. *International Journal of Social Research Methodology*, 25(4), 457–470. https://doi.org/10.1080/13645579.2021.1899596

Brauer, R., Ormiston, J., & Beausaert, S. (2024). Creativity-fostering teacher behaviors in higher education: A transdisciplinary systematic literature review. *Review of Educational Research,95*(5), 899-928. https://doi.org/10.3102/00346543241258226

Briner, R.B., & David, D. (2012). Systematic Review and Evidence Synthesis as a practice and scholarship tool. In D.M. Rousseau (Ed.), *The Oxford handbook of evidence-based management*. Oxford University Press.

Budgen, D., Turner, M., Brereton, P., & Kitchenham, B.A. (2008). Using mapping studies in software engineering. IIn PPIG 2008 - 20th Annual Workshop (vol. 8., pp. 195–204). PPIG. https://ppig.org/files/2008-PPIG-20th-budgen.pdf.

Buele, J., & Llerena-Aguirre, L. (2025). Transformations in academic work and faculty perceptions of artificial intelligence in higher education. *Frontiers in Education, 10*, Article 1603763. https://doi.org/10.3389/feduc.2025.1603763

Burton, R., Kellett, U., Mansah, M., & Sriram, D. (2024). A systematic review of online team based learning approaches in health professional education. *Nurse Education Today, 138*, Article 106156. https://doi.org/10.1016/j.nedt.2024.106156

Bustamante-Mora, A., Diéguez-Rebolledo, M., Díaz-Arancibia, J., Sánchez-Vázquez, E., & Medina-Gómez, J. (2025). Inclusive pedagogical models in STEM: The importance of emotional intelligence, resilience, and motivation with a gender perspective. *Sustainability, 17*(10), Article 4437. https://doi.org/10.3390/su17104437

Cabrera-Duffaut, A., Pinto-Llorente, A. M., & Iglesias-Rodriguez, A. (2024). Immersive learning platforms: analyzing virtual reality contribution to competence development in higher education - A systematic literature review. *Frontiers in Education*, *9,* Article 1391560. https://doi.org/10.3389/feduc.2024.1391560

Cadiz, M. C. D., Manuel, L. A. F., Reyes, M. M., Natividad, L. R., & Ibarra, F. P. (2024). Technology integration in Philippine higher education: A content-based bibliometric analysis. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi, 8*(1), 35–47. https://doi.org/10.22437/jiituj.v8i1.31807

Cadiz, M. C. D., Manuel, L. A. F., Reyes, M. M., Natividad, L. R., & Ibarra, F. P. (2024). Technology integration in Philippine higher education: A content-based bibliometric analysis. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *8* (1), 35–47. https://doi.org/10.22437/jiituj.v8i1.31807

Castillo-Martínez, I. M., Flores-Bueno, D., Gómez-Puente, S. M., & Vite-León, V. O. (2024). AI in higher education: A systematic literature review. *Frontiers in Education*, *9*, Article 1391485. https://doi.org/10.3389/feduc.2024.1391485

Castillo, F. L. V., Reyna, A. C. P., Huaraca, C. R. P., Bravo, N. A., Trejo, F. H. A., Condor, M. C. C., Bazalar, R. S., Pérez, O.P.A., Quiroz, A. G. T., & Muñoz, R. P. (2024). Contribution of artificial intelligence to the development of metaverses: A bibliometric review study on its impact on immersive learning in higher education. *International Journal of Learning, Teaching and Educational Research*, *23*(12), 322–341. https://doi.org/10.26803/ijlter.23.12.17

Chashmyazdan, M. R., Sadatmoosavi, A., Bamir, M., Poursheikhali, A., & Masoud, A. (2024). Future requirements and challenges of universities and higher education institutions in the knowledge-based economy: Literature reviews. *Journal of Higher Education Policy and Leadership Studies*, *5*(3), 180–191. https://doi.org/10.61186/johepal.5.3.180

Chen, S., Kadir, S. A., & Kang, E. K. M. S. (2024). Factors influencing lecturers' organizational commitment in higher education: A systematic literature review. *Journal of Curriculum and Teaching*, *13*(4), 192–208. https://doi.org/10.5430/jct.v13n4p192

Chugh, R., Turnbull, D., Morshed, A., Sabrina, F., Azad, S., Mamunur, R. M., Kaisar, S., & Subramani, S. (2025). The promise and pitfalls: A literature review of generative artificial intelligence as a learning assistant in ICT education. *Computer Applications in Engineering Education, 33*(2), Article e70002. https://doi.org/10.1002/cae.70002

Coleman, A., Crawford, M., Shin, H., Dadd, D., & Collins, B. (2025). The career prospects of migrant, female academics from minority ethnic backgrounds in the UK's higher education sector: An integrative review of peer-reviewed literature. *Educational Review*, *77*(4), 1313–1335. https://doi.org/10.1080/00131911.2023.2281240

Cui, P., & Alias, B. S. (2024). Opportunities and challenges in higher education arising from AI: A systematic literature review (2020–2024). *Journal of Infrastructure, Policy and Development*, *8*(11), Article 8390. https://doi.org/10.24294/jpd.v8i11.8390

Dávila, M., Guzmán, R., Macareno, H., Piñeres, D., de la Rosa, D., & Caballero-Uribe, C. (2009). Bibliometría: Conceptos y utilidades para el estudio médico y la formación profesional [Bibliometrics: Concepts and applications for medical research and professional training]. *Salud Uninorte, 25*, 319–330.

Din Eak, A., & Annamalai, N. (2024). Enhancing online learning: A systematic literature review exploring the impact of screencast feedback on student learning outcomes. *Asian Association of Open Universities Journal, 19*(3), 247–263. https://doi.org/10.1108/AAOUJ-08-2023-0100

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070

Dukes, L., III, Faggella-Luby, M., Gelbar, N., Mendoza, M. E., Charles, S., & Madaus, J. (2025). Getting global with it: The rapid growth in higher education and disability-focused professional literature. *Disabilities*, *5*(1), 13. https://doi.org/10.3390/disabilities5010013

Eaton, S. E. (2025). Global trends in education: Artificial intelligence, postplagiarism, and future-focused learning for 2025 and beyond – 2024–2025 Werklund Distinguished Research Lecture. *International Journal for Educational Integrity*, *21*(12). https://doi.org/10.1007/s40979-025-00187-6

Edvardsen Tonheim, L. E., Molin, M., Brevik, A., Gundersen, M. W., & Garnweidner-Holme, L. (2024). Facilitators and barriers to online group work in higher education within health sciences – A scoping review. *Medical Education Online*, *29*(1), Article 2341508. https://doi.org/10.1080/10872981.2024.2341508

El Aatik, Y., El Habchi, M., Gazzaz, H., Elouardi, A., El Feniche, M., & Kisra, H. (2024). State of the art of students psychological needs at the level of higher education institutions in Morocco: A systematic literature review. *The Open Public Health Journal*, *17*, Article e18749445348464. https://doi.org/10.2174/0118749445348464241010080618

Eltaiba, N., Hosseini, S., & Okoye, K. (2025). Benefits and impact of technology-enhanced learning applications in higher education in Middle East and North Africa: A systematic review. *Global Transitions*, *7*, 350–374. https://doi.org/10.1016/j.glt.2025.06.004

Esteban, A. J. (2025). Mapping the research landscape of language use in study abroad through a bibliometric lens. *Cogent Education*, *12*(1), Article 2482392. https://doi.org/10.1080/2331186X.2025.2482392

Fengye, L., Xiaoyang, L., Shuyan, Y., Muying, Y., Junyi, C., & Sun, P. (2025). A comparative analysis of research interest in medical students: An international and Chinese perspective over the past decade using CiteSpace. *Cogent Education*, *12*(1), Article 2510021. https://doi.org/10.1080/2331186X.2025.2510021

Fernández-Batanero, J. M., Montenegro-Rueda, M., Fernández-Cerero, J., & López-Menéses, E. (2024). Adoption of the Internet of Things in higher education: Opportunities and challenges. *Interactive Technology and Smart Education*, *21*(2), 292–307. https://doi.org/10.1108/ITSE-01-2023-0025

Ferreira Santos, L. (2024). Internationalisation in the digital transformation: A scoping review. *Higher Education Quarterly*, *78*(3), 807–824. https://doi.org/10.1111/hequ.12488

Fink, A. (2019). *Conducting Research literature reviews: From the internet to paper*. Sage Publications.

France, E. F., Cunningham, M., Ring, N., Uny, I., Duncan, E. A., Jepson, R. G., Noyes, J. (2019). Improving reporting of meta-ethnography: The eMERGe reporting guidance. *BMC Medical Research Methodology*, *19*(1), 1–13. https://doi.org/10.1111/jan.13809

Frez-Pulgar, G., Valdés-Rodríguez, Y., Negrier-Seguel, V., Gutiérrez-Gutiérrez, F., Caceres-Senn, M., Glasserman-Morales, L., & Hochstetter-Diez, J. (2025). Institutionalization of bioethics in higher education institutions: A systematic mapping. *Bioethical Inquiry*. https://doi.org/10.1007/s11673-025-10442-7

Galdames-Calderón, M., Stavnskær Pedersen, A., & Rodriguez-Gomez, D. (2024). Systematic review: Revisiting challenge-based learning teaching practices in higher education. *Education Sciences, 14*(9), Article 1008. https://doi.org/10.3390/educsci14091008

García-Peñalvo, F. (2017). Revisión sistemática de literatura para artículos. http://repositorio.grial.eu/handle/grial/756

Ghazian, N., & Lortie, C. J. (2024). Ten simple rules for incorporating the UN sustainable development goals (SDGs) into environmental and natural science courses. *Sustainability*, *16*(21), Article 9594. https://doi.org/10.3390/su16219594

Gómez, D. L.J., Álvarez Maestre, A. J., Parada Trujillo, A. E., Pérez Fuentes, C. A., Bedoya Ortiz, D. H., & Sanabria Alarcon, R. K. (2025). Determining factors for the development of critical thinking in higher education. *Journal of Intelligence, 13*(6), Article 59. https://doi.org/10.3390/jintelligence13060059

Gonçalves, G. S., Serra, F. A. R., Storopoli, J. E., Scafuto, I. C., & Rafael, D. N. (2024). Undergraduate student retention activities: Challenges and research agenda. *SAGE Open*. https://doi.org/10.1177/21582440241249334

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews.* Sage Publications.

Gregorio-Chaviano, O., Repiso, R., Calderón-Rehecho, A., León-Marín, J., & Jiménez-Contreras, E. (2021). Dialnet Métricas como herramienta de evaluación bibliométrica: Aportes al análisis de la actividad científica en Ciencias Sociales y Humanidades [Dialnet Métricas as a bibliometric evaluation tool: Contributions to the analysis of scientific activity in social sciences and humanities]. *El Profesional de la Información, 30*. http://doi.org/10.3145/epi.2021.may.18

Gudoniene, D., Staneviciene, E., Huet, I., Dickel, J., Dieng, D., Degroote, J., Rocio, V., Butkiene, R., & Casanova, D. (2025). Hybrid teaching and learning in higher education: A systematic literature review. *Sustainability, 17*(2), Article 756. https://doi.org/10.3390/su17020756

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Review*, 18, e1230. https://doi.org/10.1002/cl2.1230

Haddaway, N.R., Collins, A.M., Coughlin, D., & Kirk, S. (2015).The role of google scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, *10*, e0138237. https://doi.org/10.1371/journal.pone.0138237

Harmon, J., Brown, A., Birbeck, D., Crockett, J., Panadgoo, S., Nawas, A., Stringer, A., & Costabile, M. (2024). Interdisciplinary reflection by higher education academics using teaching squares: A scoping review. *Nurse Education Today, 142*, Article 106353. https://doi.org/10.1016/j.nedt.2024.106353

Heintalu, K., Saks, K., Edisherashvili, N., & Dekker, I. (2025). The conceptualisation of goal setting and goal orientation in higher education: A systematic literature review. *Educational Research Review, 48*, Article 100709. https://doi.org/10.1016/j.edurev.2025.100709

Henry, J., Li, F., & Arnab, S. (2024). On the pre-perception of gamification and game-based learning in higher education students: A systematic mapping study. *Simulation & Gaming, 55*(6), 985–1010. https://doi.org/10.1177/10468781241271082

Hidayat, R., Qi, T. Y., Ariffin, P. N. B. T., Hadzri, M. H. B. M., Chin, L. M., Ning, J. L. X., & Nasir, N. (2024). Online game-based learning in mathematics education among Generation Z: A systematic review. *International Electronic Journal of Mathematics Education, 19*(1), Article em0763. http://doi.org/10.29333/iejme/14024

Higgins, J. P., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions* (version 5.1.0 ed.). The Cochrane Collaboration. https://handbook.cochrane.org.

Ishmuradova, I. I., Chistyakov, A. A., Chudnovskiy, A. D., Grib, E. V., Kondrashev, S. V., & Zhdanov, S. P. (2024). A cross-database bibliometrics analysis of blended learning in higher education: Trends and capabilities. *Contemporary Educational Technology*, *16*(2), Article ep508. https://doi.org/10.30935/cedtech/14478

Isiaku, L., Muhammad, A. S., Kefas, H. I., & Ukaegbu, F. C. (2024). Enhancing technological sustainability in academia: Leveraging ChatGPT for teaching, learning and evaluation. *Quality Education for All, 1*(1), 385–416. https://doi.org/10.1108/QEA-07-2024-0055

Jain, I., & Gulati, R. (2025). Efficiency in higher education: A review of research from 1977 to 2022. *Higher Education Quarterly, 79* (1), Article e70010. https//:doi.org10.1111/hequ.70010

Jiaxin, G. Huijuan, Z., & Md Hassan, H. (2024). Global competence in higher education: A ten-year systematic literature review. *Frontiers in Education, 9*, Article 1404782. http://doi.org/10.3389/feduc.2024.1404782

Joshi, M. A. (2014). Bibliometric indicators for evaluating the quality of scientific publications. *Journal of Contemporary Dental Practice, 15*(2), 258–262. https://doi.org/10.5005/jp-journals-10024-1525

Junaštíková, J. (2024). Self-regulation of learning in the context of modern technology: A review of empirical studies. *Interactive Technology and Smart Education, 21*(2), 270–291. https://doi.org/10.1108/ITSE-02-2023-0030

Kalim, U., Kanwar, A., Sha, J., & Huang, R. (2025). Barriers to AI adoption for women in higher education: A systematic review of the Asian context. *Smart Learning Environments*, *12*(38), 1–21. https://doi.org/10.1186/s40561-025-00390-5

Kalocsányiová, E., Bilici, N., Jenkins, R., Obojska, M., & Samuk Carignani, Ş. (2024). What works to facilitate displaced and refugee-background students' access and participation in European higher education: Results from a multilingual systematic review. *Educational Review*, *76*(6), 1722–1743. https://doi.org/10.1080/00131911.2022.2085670

Karimi, H., & Khawaja, S. (2025). Exploring digital competence among higher education teachers: A systematic review. *International Journal of Learning, Teaching and Educational Research*, *24*(1), 298–314. https://doi.org/10.26803/ijlter.24.1.15

Kaymakcioglu, A. G., & Thomas, M. (2024). Gender inequalities and academic leadership in Nigeria, South Africa and the United Kingdom: A systematic literature review (2013–2023). *Social Sciences & Humanities Open, 10*, Article 101066. https://doi.org/10.1016/j.ssaho.2024.101066

Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Keele. http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. https://docs.edtechhub.org/lib/EDAG684W.

Klimova, B., & Chen, J. H. (2024). The impact of AI on enhancing students' intercultural communication competence at the university level: A review study. *Language Teaching Research Quarterly*, *43*, 102–120. https://doi.org/10.32038/ltrq.2024.43.06

Knobloch, K., Yoon, U., & Vogt, P. (2011). Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *Journal of Cranio-Maxillofacial Surgery*, 39(2), 91–92. https://doi.org/10.1016/j.jcms.2010.11.001

Kovári, A. (2025). A systematic review of AI-powered collaborative learning in higher education: Trends and outcomes from the last decade. *Social Sciences & Humanities Open, 11,* Article 101335. https://doi.org/10.1016/j.ssaho.2025.101335

Kraus, S., Breier, M., & Dasí-Rodríguez, S. (2020). The art of crafting a systematic literature review in entrepreneurship research. *International Entrepreneurship and Management Journal, 16*(3), 1023-1042. https://doi.org/10.1007/s11365-020-00635-4

Kumar, K., & Pande, B.P. (2021). Rise of online teaching and tearning processes during COVID-19 Pandemic. In P.K. Praveen Kumar Khosla, M.D. Mittal, Sharma, & L.M Goyal, *Predictive and Preventive Measures for Covid-19 Pandemic* Singapore (pp. 251-271). Springer Nature. http://doi.org/10.1007/978-981-33-4236-1_14

Landa, A. H., Szabo, I., Le Brun, L., Owen, I., Fletcher, G., & Hill, M. (2011). An evidence-based approach to scoping reviews. *Electronic Journal of Information Systems Evaluation*, *14*, 46–52.

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*, 69. https://doi.org/10.1186/1748-5908-5-69

Liang, T., Singh, C. K. S., Mulyadi, D., & Singh, T. S. M. (2025). Enhancing EFL writing through online peer feedback: A systematic review of higher education studies. *Journal of Curriculum and Teaching, 14*(1), 184–201. https://doi.org/10.5430/jct.v14n1p184

Lim, T. L., & Lee, A. S. H. (2024). A systematic literature review of the benefits of utilizing pervasive tools in higher education. *Journal of Educators Online, 21*(1). http://doi.org/10.9743/jeo.2024.21.2.19

Liu, Y., Alias, B. S., & Hamid, A. H. A. (2025). Student entrepreneurship competence and its contribution to sustainable development: A systematic review in the context of Chinese higher education. *Sustainability*, *17*(7), 3148. https://doi.org/10.3390/su17073148

López-Chila, R., Llerena-Izquierdo, J., Sumba-Nacipucha, N., & Cueva-Estrada, J. (2024). Artificial intelligence in higher education: An analysis of existing bibliometrics. *Education Sciences, 14*(1), Article 47. https://doi.org/10.3390/educsci14010047

López-Núñez, J.-A., Alonso-García, S., Bernal-Ortiz, B., & Victoria-Maldonado, J.-J. (2024). A systematic review of digital competence evaluation in higher education. *Education Sciences, 14*(11), Article 1181. https://doi.org/10.3390/educsci14111181

Lund, K., Jeong, H., Grauwin, S., & Jensen, P. (2020). Research in education draws widely from the Social Sciences and Humanities. *Frontiers in Education, 5*, Article 544194. https://doi.org/10.3389/feduc.2020.544194

Luo, R.-Z., & Zhou, Y.-L. (2024). The effectiveness of self-regulated learning strategies in higher education blended learning: A five years systematic review. *Journal of Computer Assisted Learning*, *40*(6), 3005–3029. https://doi.org/10.1111/jcal.13052

Ma, T. (2025). Systematically visualizing ChatGPT used in higher education: Publication trend, disciplinary domains, research themes, adoption and acceptance. *Computers and Education: Artificial Intelligence, 8*, Article 100336. https://doi.org/10.1016/j.caeai.2024.100336

Mahrishi, M., Ramakrishna, S., Hosseini, S., & Abbas, A. (2025). A systematic literature review of the global trends of outcome-based education (OBE) in higher education with an SDG perspective related to engineering education. *Discover Sustainability, 6,* 620. https://doi.org/10.1007/s43621-025-01496-z

Marano, E., Newton, P. M., Birch, Z., Croombs, M., Gilbert, C., & Draper, M. J. (2024). What is the student experience of remote proctoring? A pragmatic scoping review. *Higher Education Quarterly*, *78*, 1031–1047. https://doi.org/10.1111/hequ.12506

Maré, S., & Mutezo, A. T. (2025). The community of inquiry, shared metacognition, and student engagement in online learning: A systematic review. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2025.2479164

Marques, R. M. G., Lopes, A., & Magalhães, A. M. (2024). Academic identities and higher education change: Reviewing the evidence. *Educational Research*, *66*(2), 228–244. https://doi.org/10.1080/00131881.2024.2334760

Marshall, I. J., & Wallace, B. C. (2019). Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. *Systematic Reviews*, *8*(1), 163. https://doi.org/10.1186/s13643-019-1074-9

Masalimova, A. R., Orekhovskaya, N. A., Pivovarov, V. A., Borovikova, Y. V., Zhirkova, G. P., & Chauzova, V. A. (2024). Analyzing trends in online learning in higher education in the BRICS countries through bibliometric data. *Frontiers in Education*, *9*, Article 1409013. https://doi.org/10.3389/feduc.2024.1409013

Mashingaidze, K., & Mayayise, T. O. (2025). Redefining relevance in information systems curriculum: Insights for IS educators from a systematic literature review. *Cogent Education*, *12*(1), Article 2476291. https://doi.org/10.1080/2331186X.2025.2476291

Mireku, D. O., & Bervell, B. (2024). A decade of quality assurance in higher education (QAIHE) within sub-Saharan Africa: A literature review based on a systematic search approach. *Higher Education, 87*(5), 1271–1316. https://doi.org/10.1007/s10734-023-01064-2

Mireku, D. O., Bervell, B., & Dzamesi, P. D. (2024b). Examination malpractice behaviours in Higher Education (EMALBiFE) in sub-Saharan Africa: A systematic review. *International Journal of Educational Development, 108*, Article 103064. https://doi.org/10.1016/j.ijedudev.2024.103064

Mireku, D. O., Dzamesi, P. D., & Bervell, B. (2024a). Plagiarism in higher education (PLAGiHE) within sub-Saharan Africa: A systematic review of a decade (2012–2022) literature. *Research Ethics, 20*(2), 156–186. https://doi.org/10.1177/17470161231189646

Mohammed, F. S., & Ozdamil, F. (2024). A systematic literature review of soft skills in information technology education. *Behavioral Sciences, 14*(10), Article 894. https://doi.org/10.3390/bs14100894

Moher, D., Liberati, A., Tetzlaff, J., Altlan, D.G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-analyses: The PRISMA Statement. *PLoS Med*, *6*, e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & Grupo Prisma. (2014). Ítems de referencia Para publicar revisiones sistematicas y metaanalisis: La declaracion PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement]. *Revista Española de Nutricion Humana y Dietetica*, *18*(3), 172-181.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery, 8*(5), 336–341. https://doi.org/10.1016/j.ijsu.2010.02.007

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Japanese Pharmacology and Therapeutics*, *1*(4), 1–9.

Moreno, B., Muñoz, M., & Cuellar, J. (2018). Systematic Reviews: Definition and basic notions (Spanish). *Revista Clínica de Periodoncia, Implantología y Rehabilitación Oral, 11*(3), 184–186. https://doi.org/10.4067/s0719-01072018000300184

Munn, Z., Peters, M.D.J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1),143. https://doi.org/10. 1186/s12874- 018- 0611-x.

Mursalzade, G., Escriche-Martínez, S., Valdivia-Salas, S., Jiménez, T. I., & López-Crespo, G. (2025). Factors associated with psychological flexibility in higher education students: A systematic review. *Sustainability, 17*(12), Article 5557. https://doi.org/10.3390/su17125557

Nalweyiso, D. I., Mbabazi, J., Saltikov, J. B., Kabanda, J., Breckon, J., Nnyanzi, L. A., Kawooya, M. G., Mubuuke, A. G., Kinengyere, A.A., & Sanderson, K. (2025). The current educational interventions for teaching and learning evidence-based practice knowledge, skills, attitudes, and behaviours: A systematic review among undergraduate healthcare students in developing countries. *Cogent Education*, *12*(1), 2460254. https://doi.org/10.1080/2331186X.2025.2460254

Ncube, M. M., & Ngulube, P. (2024). A systematic review of postgraduate programmes concerning ethical imperatives of data privacy in sustainable educational data analytics. *Sustainability*, *16*(15), 6377. https://doi.org/10.3390/su16156377

Ndibalema, P. (2025). Digital literacy gaps in promoting 21st century skills among students in higher education institutions in Sub-Saharan Africa: A systematic review. *Cogent Education*, *12*(1), 2452085. https://doi.org/10.1080/2331186X.2025.2452085

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 methodology, perspectives, and application* (pp. 3-22). Springer.

Nieminen, J. H., Morina, A., & Biagiotti, G. (2024). Assessment as a matter of inclusion: A meta-ethnographic review of the assessment experiences of students with disabilities in higher education. *Educational Research Review, 42*, Article 100582. https://doi.org/10.1016/j.edurev.2023.100582

Noblit, G. W. (2019). *Meta-ethnography in education*. Oxford University Press.

Noblit, G. W., & Hare, R. D. (1988). *Meta ethnography: Synthesizing qualitative studies*. Sage. https://doi.org/10.4135/9781412985000

Novak, J.D. (1990). Concept mapping: a useful tool for science education. *Journal of Research in Science Teaching*, *27*(10), 937-949.

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences, 14*(6), Article 636. https://doi.org/10.3390/educsci14060636

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, *37*(1), 879–910. https://doi.org/10.17705/1CAIS.03743

Olivares-De la fuente, P., Jiménez-García, E., & García-López, Ó. (2025). Twitter and YouTube as digital tools in higher education: A systematic review. *Frontiers in Education, 10*, Article 1625803. https://doi.org/10.3389/feduc.2025.1625803

Ortega-Ruipérez, B., & Correa-Gorospe, J. M. (2024). Peer assessment to promote self-regulated learning with technology in higher education: Systematic review for improving course design. *Frontiers in Education*, 9, Article 1376505. https://doi.org/10.3389/feduc.2024.1376505

Oulamine, A., Chakra, R., Ziky, R., Bahida, H., El Gareh, F., Oubihi, I., & Massiki, A. (2025). A systematic literature review of barriers affecting e-learning in higher education. *Educational Process: International Journal*, *17*, Article e2025396. https://doi.org/10.22521/edupij.2025.17.396

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Revista Española de Cardiología, 74*, 790–799. https://doi.org/10.1016/j.recesp.2021.06.016

Page, M. J., Shamseer, L., Altman, D. G., Tetzlaff, J., Sampson, M., Tricco, A. C., Catalá-López,F. Li, L., Reid, E. K., Sarkis-Onofre, R., & Moher, D. (2016). Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. *PLOS Medicine*, *13*(5), ei1002028. https://doi.org/10.1371/journal.pmed.1002028

Pallaris, G., Zaphiris, P., & Parmaxi, A. (2024). Mapping the landscape of Makerspaces in higher education: An inventory of research findings. *Interactive Technology and Smart Education*, *21*(1), 1–20. https://doi.org/10.1108/ITSE-01-2022-0013

Pandey, S., Sahoo, S., & Mishra, S. (2025). Mapping of research on higher education in India: A bibliometric review. *India Higher Education Report 2023: Higher Education Research* (pp. 171-185). Taylor and Francis. https://doi.org/10.4324/9781041025870-14

Papaioannou, D., Sutton, A., & Booth, A. (2016). *Systematic approaches to a successful literature review*. SAGE Publications Ltd.

Parambil, M. M. A., Rustamov, J., Ahmed, S. G., Rustamov, Z., Awad, A. I., Zaki, N., & Alnajjar, F. (2024). Integrating AI-based and conventional cybersecurity measures into online higher education settings: Challenges, opportunities, and prospects. *Computers and Education: Artificial Intelligence, 7*, Article 100327. https://doi.org/10.1016/j.caeai.2024.100327

Paredes-Canencio, K. N., Lasso, A., Castrillon, R., Vidal-Medina, J. R., & Quispe, E. C. (2024). Carbon footprint of higher education institutions. *Environment, Development and Sustainability*, *26*, 30239–30272. https://doi.org/10.1007/s10668-024-04596-4

Pelaez-Sanchez, I. C., Glasserman-Morales, L. D., & Rocha-Feregrino, G. (2024b). Exploring digital competencies in higher education: Design and validation of instruments for the era of Industry 5.0. *Frontiers in Education*, 9, Article 1415800. https://doi.org/10.3389/feduc.2024.1415800

Peláez-Sánchez, I. C., Velarde-Camaqui, D., & Glasserman-Morales, L. D. (2024a). The impact of large language models on higher education: Exploring the connection between AI and Education 4.0. *Frontiers in Education*, *9*, Article 1392091. https://doi.org/10.3389/feduc.2024.1392091

Pereta, K. (2024). Higher education in contemporary Sri Lanka: Key topics. In *Routledge Handbook of Contemporary Sri Lanka* (p. 207-218). Routledge. http://doi.org/10.4324/9781003300991-20

Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020a). Chapter 11: Scoping reviews (2020 version). In E. Aromataris & Z. Munn (Eds.), *JBI manual for evidence synthesis*. JBI.

Peters, M.D.J., Marnie, C., Tricco, A.C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C.M., & Khalil, H. (2020b). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis*, *18*(10), 2119–26. https://doi.org/10.11124/jbies-20-00167

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). *Systematic mapping studies in software engineering*. In *Proceedings of the 12th international conference on evaluation and assessment in software engineering* (pp. 68–77*).* ACM.

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*, 1–18. https://doi.org/10.1016/j.infsof.2015.03.007

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.

Pollock, D., Tricco, A.C., Peters, M.D.J., Mclnerney, P.A., Khalil, H., Godfrey, C.M., Alexander, L.A., & Munn, Z. (2022). Methodological quality, guidance, and tools in scoping reviews: A scoping review protocol. *JBI Evidence Synthesis, 20*(4), 1098–1105. https://doi.org/10.11124/jbies-20-00570

Popay, J., Roberts, H., Sowden, A.,Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews, A product from the ESRC methods programme Version, Vol.1 No.1, p.b92.

Punch, R., Duncan, J., & Talbot-Stokes, R. (2025). Experiences and challenges of students with disability in Australian universities: A scoping review. *International Journal of Inclusive Education*. https://doi.org/10.1080/13603116.2025.2526009

Pussegoda, K., Turner, L., Garritty, C., Mayhew, A., Skidmore, B., Stevens, A., Boutron, I., Sarkis-Onofre, R., Bjerre, L. M., Hrobjartsson, A., Altman, D. G., & Moher, D. (2017). Identifying approaches for assessing methodological and reporting quality of systematic reviews: A descriptive study. *Systematic Reviews*, 6(1), 117. https://doi.org/10.1186/s13643-017-0507-6

Rahajeng, U. W., Hendriani, W., & Paramita, P. P. (2024). Navigating higher education challenges: A review of strategies among students with disabilities in Indonesia. *Disabilities*, *4*(3), 678–695. https://doi.org/10.3390/disabilities4030042

Raitskaya, L., & Tikhonova, E. (2024). Appliances of generative AI-powered language tools in academic writing: A scoping review. *Journal of Language and Education*, *10*(4), 5–30. https://doi.org/10.17323/jle.2024.24181

Raitskaya, L., & Tikhonova, E. (2025). Enhancing critical thinking skills in ChatGPT-human interaction: A scoping review. *Journal of Language and Education*, *11*(2), 5–19. https://doi.org/10.17323/jle.2025.27387

Renfors, S.-M. (2024). Education for the circular economy in higher education: An overview of the current state. *International Journal of Sustainability in Higher Education*, *25*(9), 111–127. https://doi.org/10.1108/IJSHE-07-2023-0270

Rosales-Ricardo, Y., & Cáceres-Manzano, V. (2024). Effects of physical exercise on academic performance in university students: A systematic review. *Health Professions Education*, *10*(3), 174–184. https://doi.org/10.55890/2452-3011.1174

Rosário, A. T., & Raimundo, R. (2024). Sustainable entrepreneurship education: A systematic bibliometric literature review. *Sustainability*, *16*(2), 784. https://doi.org/10.3390/su16020784

Roy, R., Rao, G. A., Pal, D., Anuradha, S., & Mukherjee, S. (2025). Metaverse-based education for sustainable development and improving the performance: Discussing the future research agenda. *Sustainable Futures*, *10*, Article 101091. https://doi.org/10.1016/j.sfrr.2025.101091

Saez-Zevallos, N. S., & Montalvo-Apolin, D. E. (2025). Strategies for learning English in higher education: A systematic mapping. *Frontiers in Education*, *10*, Article 1570602. https://doi.org/10.3389/feduc.2025.1570602

Sahar, R., & Munawaroh, M. (2025). Artificial intelligence in higher education with bibliometric and content analysis for future research agenda. *Discover Sustainability*, *6*(401), 1–32. https://doi.org/10.1007/s43621-025-01086-z

Sánchez-Meca, J. (2022). Systematic reviews and meta-analyses in Education: A tutorial. *RiiTE Revista Interuniversitaria de Investigación En Tecnología Educativa*, 5–40. https://doi.org/10.6018/riite.545451

Sarkis, R., Catalá, F., Aromataris, E., & Lockwood, C. (2021). How to properly use the PRISMA Statement. *Systematic Reviews*, 10, 117. https://doi.org/10.1186/s13643-021-01671-z

Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and use of AI Chatbots among students in higher education: A Scoping review of empirical studies. *Education Sciences, 14*(8), Article 922. https://doi.org/10.3390/educsci14080922

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer. https://doi.org/10.1007/978-3-319-21416-0

Selçuk, A.A. (2019). A guide for systematic reviews: PRISMA. *Turkish Archives of Otorhinolaryngology*, *57*(1), 57–58. https://doi.org/10.5152/tao.2019.4058

Sergeeva, O. V., Zvereva, E. V., Kosheleva, Y. P., Zheltukhina, M. R., Prokopyev, A. I., & Dobrokhotov, D. A. (2024). A comprehensive bibliometric review of gamified learning in higher education. *Contemporary Educational Technology*, *16*(3), Article ep514. https://doi.org/10.30935/cedtech/14708

Serrano, S. S., Navarro, I. P., & González, M. D. (2022). ¿ Cómo hacer una revisión sistemática siguiendo el protocolo PRISMA?: Usos y estrategias fundamentales para su aplicación en el ámbito educativo a través de un caso práctico [How to conduct a systematic review following the PRISMA Protocol?: Fundamental uses and strategies for its application in the educational field through a practical case]. *Bordón. Revista de Pedagogía, 74*(3), 51-66. https://doi.org/10.13042/Bordon.2022.95090

Setiamurti, A., & Kurniawati, F. (2024). Fostering creativity in higher education institution: A systematic review (2018–2022). *Open Education Studies*, *6*, Article 20220221. https://doi.org/10.1515/edu-2022-0221

Shahjahan, R. A., & Seinn, A. P. (2025). Towards 'decolonizing' engineering education: An integrative review. *Cogent Education, 12*(1), 2541075. https://doi.org/10.1080/2331186X.2025.2541075

Singun, A., Jr. (2025). Unveiling the barriers to digital transformation in higher education institutions: A systematic literature review. *Discover Education, 4*(1), Article 37. https://doi.org/10.1007/s44217-025-00430-9

Sohrabi, C., Franchi, T., Mathew, G., Kerwan, A., Nicola, M., Griffin, M., Agha, M., & Agha, R. (2021). PRISMA 2020 statement: What's new and the importance of reporting guidelines. *International Journal of Surgery*, *88*, 105918. https://doi.org/10.1016/j.ijsu.2021.105918

Solís García, P., Real Castelao, S., & Barreiro-Collazo, A. (2024). Trends and challenges in the mental health of university students with disabilities: A systematic review. *Behavioral Sciences, 14*(2), Article 111. https://doi.org/10.3390/bs14020111

Solis-García, P., Barreiro-Collazo, A., Rodríguez-Correa, M., Delgado-Rico, E., & Real-Castelao, S. (2025). Inclusion of students with disabilities in the European Higher Education Area (EHEA): A systematic review. *Cogent Education*, *12*(1), Article 2430880. https://doi.org/10.1080/2331186X.2024.2430880

Stamou, P., Tsoli, K., & Babalis, T. (2024). The role of counseling for non-traditional students in formal higher education: A scoping review. *Frontiers in Education*, *9,* Article 1361410. https://doi.org/10.3389/feduc.2024.1361410

Stansfield, C., Dickson, K., & Bangpan, M. (2016). Exploring issues in the conduct of website searching and other online sources for systematic reviews: How can we be systematic? *Systematic Reviews*, *5*, Article 191. https://doi.org/10.1186/s13643-016-0371-9

Subki. (2025). A review of green campus management sustainability with a bibliometric approach. *International Journal of Sustainable Development and Planning*, *20*(6), 2717–2729. https://doi.org/10.18280/ijsdp.200638

Sukjairungwattana, P., Hu, H., Liu, R., & Huang, J. (2025). From local to global: systematically reviewing higher education internationalization in Asia. *Frontiers in Education, 9*, Article 1473820. http://doi.org/10.3389/feduc.2024.1473820

Tang, X., Rousse-Malpat, A., & Duarte, J. (2024). Implementing translanguaging strategies in the English writing classroom in higher education: A systematic review. *AILA Review*. http://doi.org/10.1075/aila.23018.tan

Tang, M., Maat, S. M., & Azman, N. (2025). Languages and the internationalization of higher education in China: A review from a sociolinguistic perspective. *Forum for Linguistic Studies*, *7*(5), 98–113. https://doi.org/10.30564/fls.v7i5.9309

Tareke, T. G., Oo, T. Z., & Jozsa, K. (2025). Bridging theoretical gaps to improve students' academic success in higher education in the digital era: A systematic literature review. *International Journal of Educational Research Open, 9*, Article 100510. https://doi.org/10.1016/j.ijeoro.2025.100510

Tareke, T. G., Woreta, G. T., Zewude, G. T., Amukune, S., Oo, T. Z., & Józsa, K. (2024). Overview of Ethiopian public higher education: Trends, system, challenges, and quality issues. *Education Sciences, 14*(10), Article 1065. https://doi.org/10.3390/educsci14101065

Templier, M., & Paré, G. A. (2015). Framework for Guiding and Evaluating Literature Reviews. *Communications of the Association for Information Systems*, 37. https://doi.org/10.17705/1CAIS.03706

Teng, Y., & Cosier, M. E. (2024). Influences of cultural capital and internationalization on global competence in higher education: A systematic literature review. *Frontiers in Education*, *9*, Article 1397642. https://doi.org/10.3389/feduc.2024.1397642

Thiedig, C., & Wegner, A. (2024). Evidence use in higher education decision-making and policy: A scoping review of empirical studies from 2010 to 2022. *London Review of Education*, *22*(1), 36. https://doi.org/10.14324/LRE.22.1.36

Tikhonova, E., & Raitskaya, L. (2024). The culture of research: A systematic scoping review. *Journal of Language and Education*, *10*(1), 5-24. https://doi.org/10.17323/jle.2024.21526

Tillmanns, T., Salomão Filho, A., Rudra, S., Weber, P., Dawitz, J., Wiersma, E., Dudenaite, D., & Reynolds, S. (2025). Mapping tomorrow's teaching and learning spaces: A systematic review on GenAI in higher education. *Trends in Higher Education*, *4*(1), Article 2. https://doi.org/10.3390/higheredu4010002

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, 4(3), 356–367. https://doi.org/10.1177/1534484305278283

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed Management knowledge by means of systematic review. *British Journal of Management*, *14*, 207–222. http://doi.org/10.1111/1467-8551.00375

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., Straus, S.E. (2018). PRISMA extension for scoping reviews (PRISMA- ScR): Checklist and explanation. *Annals of Internal Medicine , 169*(7), 467–473. https://doi.org/10.7326/M18-0850

Trujillo-Juárez, S.-I., Chaparro-Sánchez, R., Morita-Alexander, A., Escudero-Nahón, A., & Delgado-González, A. (2025). Strengthening teacher digital competence in higher education through micro-courses: A systematic literature review. *Discover Education*, *4*(1), 247. https://doi.org/10.1007/s44217-025-00687-0

Turnbull, D., Chugh, R., & Luck, J. (2022). Systematic narrative hybrid literature review: A strategy for integrating a concise methodology into a manuscript. *Social Sciences & Humanities Open*, *7*(1), 1–4. https://doi.org/10.1016/j.ssaho.2022.100381

Urrútia, G., & Bonfill, X. (2010). Declaración PRISMA: Una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis [PRISMA declaration: A proposal to improve the publication of systematic reviews and meta-analyses]. *Medicina Clinica, 135*, 507–511. https://doi.org/10.1016/j.medcli.2010.01.015

Valencia Quecano, L. I., Guzmán Rincón, A., & Barragán Moreno, S. (2024). Dropout in postgraduate programs: A underexplored phenomenon – A scoping review. *Cogent Education, 11*(1), Article 2326705. https://doi.org/10.1080/2331186X.2024.2326705

van der Wee, M. L. E., Tassone, V. C., Wals, A. E. J., & Troxler, P. (2024). Characteristics and challenges of teaching and learning in sustainability-oriented Living Labs within higher education: a literature review. *International Journal of Sustainability in Higher Education*, *25*(9), 255–277. https://doi.org/10.1108/IJSHE-10-2023-0465

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Veres, C., Tänase, M., Bacos, I. B., & Kardos, M. (2025). Sustainable universities: A bibliometric and thematic analysis in higher education. *Sustainability*, *17*(5), 1817. https://doi.org/10.3390/su17051817

Vinueza-Morales, M., Rodas-Silva, J., Vidal-Silva, C., Cordova-Morán, J., & Cevallos-Ayón, E. (2025). Teaching programming in higher education: A bibliometric analysis of trends, technologies, and pedagogical approaches. *Frontiers in Education, 10*, Article 1525917. https://doi.org/10.3389/feduc.2025.1525917

Vuorjainen, A., Rikala, P., Heilala, V., Lehesvuori, S., Öz, S., Kettunen, L., & Hämäläinen, R. (2025). The six C's of successful higher education-industry collaboration in engineering education: A systematic literature review. *European Journal of Engineering Education, 50*(1), 26–50. https://doi.org/10.1080/03043797.2024.2432440

Wang, D., & Ishak, Z. (2025). Unveiling the implementation of social-emotional learning among college students: A systematic literature review. *International Journal of Learning, Teaching and Educational Research*, *24*(5), 376–399. https://doi.org/10.26803/ijlter.24.5.20

Wang, Q., & Abdullah, A. H. (2024). Enhancing students' critical thinking through mathematics in higher education: A systemic review. *SAGE Open*. Advance online publication. https://doi.org/10.1177/21582440241245678

Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. *Journal of Advanced Nursing*, *52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.

Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J. & Pawson, R. (2013). RAMESESpublication standards: Realist syntheses. *BMC Medicine*, *11*, 1, 2-14. http://doi.org/10.1186/1741-7015-11-21.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, *39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

You, C., Awang, S. R., & Wu, Y. (2024). Bibliometric analysis of global research trends on higher education leadership development using Scopus database from 2013–2023. *Discover Sustainability*, *5*(1), Article 246. https://doi.org/10.1007/s43621-024-00432-x

Zhao, D., & Selvaratnam, D. P. (2024). A systematic literature review on the reform of vocational education in China. *Cogent Education, 11*(1), Article 2343525. https://doi.org/10.1080/2331186X.2024.2343525

APPENDIX 1

Coding Review Themes

Review reference	Theme	Code
Abdallah et al., 2025	ChatGPT applications in higher education	1
Adauto Medina et al., 2024	ChatGPT in university academic writing	1
Ahmad & Khurizan, 2024	Data Envelopment Analysis	2
Alonso et al., 2025	Blended learning in higher education for the development of intrinsic motivation	5
Alotaibi, 2025	AI and LMS Integration in higher education	1
Alvarodo, 2025	Design thinking as a teaching methodology	8
Álvarez Martinez et al., 2025	Mobile-assisted language learning (MALL) for improving English language teaching and learning in higher education	2
Amavasi & Zimmerman, 2024	Infection prevention and control continuous education and training in pre-registration nursing programmes	3
Andrade-Girón et al., 2025	Generative artificial intelligence in higher education learning	1
Balalee, 2025	Virtual reality, augmented reality, and mixed reality in higher education	2
Banarjee et al., 2024	Inconsistencies in the inclusion of subjects in research reports on STEM workforce skills in theUK	8
Bannigan et al., 2025	Whole-university approaches to embedding well-being in the curriculum	3
Barikzai et al., 2024	E-learning adoption in emerging economies	5
Barua & Lockee, 2024	Strategies to incorporate flexibility in higher education course designs	2
Basheer et al., 2024	Sustainability Assessment Practices in Higher Education	6
Bayly-Castaneda et al., 2024	Crafting personalized learning paths with AI for lifelong learning	1
Benson et aL., 2024	Quality assurance interventions in blended learning design	5
Bonilla-Jurado et al., 2024	Benefits of Education for Sustainable Development at Universities	6
Brauer et al., 2024	Creativity-Fostering Teacher Behaviors in Higher Education	2
Buele & Llerena-Aguirre, 2025	Transformations in academic work and faculty perceptions of artificial intelligence in higher education	1, 2
Burton et al., 2024	Online team-based learning approaches in health professional education	5
Bustamante-Mona et al., 2025	Inclusive Pedagogical Models in STEM	3
Cabrera-Duffaut et al., 2024	Immersive learning platforms in higher education	2
Cadiz, 2024	Technology integration in Phillippine higher education	2
Castillo et al., 2024	AI Impact on Immersive Learning in Higher Education	2
Castillo-Martinez et al., 2024	AI in Higher Education	1
Chashmyazdan et al., 2024	Challenges of Universities and Higher Education Institutions in the Knowledge-Based Economy	1
Chen et al., 2024	Factors Influencing Lecturers' Organizational Commitment in Higher Education	2
Chugh et al., 2025	Generative Artificial Intelligence as a Learning Assistant in ICT Education	1

Review reference	Theme	Code
Coleman et al., 2025	The career prospects of migrant, female academics from minority ethnic backgrounds in the UK's higher education sector	3
Cui & Alias, 2024	Opportunities and challenges in higher education arising from AI	1
Din Eak & Annamalai, 2024	The impact of screencast feedback on student learning outcomes	2
Dukes III et al., 2024	Disability-focus aspects of higher education	3
Edvardsen Tonheim et al., 2024	Online group work in higher education within health sciences	3
El Aatik et al., 2024	Students Psychological Needs at the Level of Higher Education Institutions in Morocco	3
Eltaiba et al., 2025	Technology-enhanced learning applications in higher education in Middle East and North Africa	2
Esteban, 2025	Language use in study Abroad, student mobility	7
Fengye et al., 2025	Research interest in medical students: an international and Chinese perspective	3
Fernandez-Batanero et al., 2024	Adoption of the Internet of Things in higher education	1, 2
Ferreira Santos, 2024	Internationalisation in the digital transformation of higher education	7
Frez-Pulgar et al., 2025	Institutionalization of Bioethics in Higher Education Institutions	7
Galdames-Calderón et al., 2024	Challenge-Based Learning and Teaching Practices in Higher Education	2
Ghazian & Lortie, 2024	Incorporating the UN Sustainable Development Goals (SDGs) into Environmental and Natural Science Courses	6
Gomez et al., 2025	Factors for the Development of Critical Thinking in Higher Education	4
Goncalves et al., 2024	Undergraduate Student Retention Activities	5
Gudoniene et al., 2025	Hybrid Teaching and Learning in Higher Education	5
Harmon et al., 2024	Interdisciplinary reflection by higher education academics using teaching squares	2
Heintalu et al., 2025	The conceptualisation of goal setting and goal orientation in higher education	8
Henry et al., 2024	Pre-Perception of Gamification and Game-Based Learning in Higher Education Students	2
Hidayat et al., 2024	Online game-based learning in mathematics education among Generation Z	2
Ishmuradova et al., 2024	Blended learning in higher education	5
Isiaku et al., 2024	Leveraging ChatGPT for teaching, learning and evaluation	1
Jaxin et al., 2024	Global competence in higher education	7
Junaštíkova, 2024	Self-regulation of learning in the context of modern technology	2
Kalim et al., 2025	Barriers to AI adoption for women in higher education	1
Kalocsányiová et al., 2024	Facilitating displaced and refugee-background students' access and participation in European higher education	3
Karimi & Khawaja, 2025	Digital Competence among Higher Education Teachers	2
Kaymakcioglu & Thomas, 2024	Gender inequalities and academic leadership by countries	3,7
Klimova & Chen, 2024	Impact of AI on Enhancing Students' Intercultural Communication Competence at the University Level	1
Kovari, 2025	AI-powered collaborative learning in higher education	1

Review reference	Theme	Code
Liang et al., 2025	Enhancing EFL Writing Through Online Peer Feedback	2
Lim & Lee, 2024	Utilising Pervasive Tools in Higher Education	2, 3
Liu et al., 2025	Student Entrepreneurship Competence and Its Contribution to Sustainable Development	2, 6
Lopez-Chila et al., 2024	Artificial Intelligence in Higher Education	1
López-Nuñez et al., 2024	Digital Competence Evaluation in Higher Education	2
Lou & Zhou, 2024	Effectiveness of self-regulated learning strategies in higher education blended learning	2, 5
Ma, 2025	ChatGPT in higher education	1
Marano et al., 2024	Student experience of remote proctoring	5
Mare & Mutezo, 2025	Community of inquiry, shared metacognition, and student engagement in online learning	4, 5
Marques et al., 2024	Academic identities and higher education change	7
Masalimova et al., 2024	Trends in online learning in higher education in the BRICS countries	5
Mashingaidze & Mayayise, 2025	Relevance in information systems curriculum	2
Mireku & Bervell, 2024	Quality assurance in higher education (QAiHE) within subSaharan Africa	7
Mireku et al., 2024a	Plagiarism in Higher Education within Sub-Saharan Africa	3, 8
Mireku et al., 2024b	Examination malpractice behaviours in Higher Education in sub-Saharan Africa	3
Mohammed & Ozdamli, 2024	Soft Skills in Information Technology Education	4
Marano et al., 2024	Student experience of remote proctoring	5
Mursalzade et al., 2025	Factors Associated with Psychological Flexibility in Higher Education Students	3
Nalweyiso et al., 2025	Educational interventions for teaching and learning evidence-based practice knowledge, skills, attitudes, and behaviours	8
Ncube & Ngulube, 2024	Ethical Imperatives of Data Privacy in Sustainable Educational Data Analytics in Postgraduate Programmes	3, 6
Ndibalema, 2025	Digital literacy gaps in promoting 21st century skills among students in higher education institutions in Sub-Saharan Africa	2
Nieminen et al., 2024	Assessment experiences of students with disabilities in higher education	3
Ogunleye et al., 2024	Generative AI for Teaching and Learning Practice	1
Olivares-De la fuente et al., 2025	Twitter and YouTube as digital tools in higher education	2
Ortega-Ruiperez & Correa-Gorospe, 2024	Peer assessment to promote self-regulated learning with technology in higher education	2
Oulamine et al., 2025	Barriers Affecting e-Learning in Higher Education	2, 5
Pallaris et al., 2024	Mapping the landscape of Makerspaces in higher education	2
Parambil et al., 2024	Integrating AI-based and conventional cybersecurity measures into online higher education settings	1
ParedesCanencio et al., 2024	Carbon footprint of higher education institutions	6
Peláez-Sánchez et al., 2024a	The impact of large language models on higher education: exploring the connection between AI and Education 4.0	1
Pelaez-Sanchez et al., 2024b	Digital competencies in higher education for Industry 5.0	2

Review reference	Theme	Code
Punch et al., 2025	Experiences and challenges of students with disability in Australian universities	3
Rahajeng et al., 2024	Strategies among Students with Disabilities in Indonesia	3
Raitskaya & Tikhonova, 2024	Appliances of Generative AI-Powered Language Tools in Academic Writing	1
Raitskaya & Tikhonova, 2025	Enhancing Critical Thinking Skills in ChatGPT-Human Interaction	1, 4
Renfors, 2024	Education for the circular economy in higher education	2
Rosales-Ricardo & Caceres-Manzano, 2024	Effects of Physical Exercise on Academic Performance in University Students	3
Rosario & Raimundo, 2024	Sustainable Entrepreneurship Education	6
Roy et al., 2025	Metaverse-based education for sustainable development and improving the performance	2, 6
Saez-Zevallos & Montalvo-Apolín, 2025	Strategies for learning English in higher education	2
Sahar & Munawaroh, 2025	Artificial intelligence in higher education	1
Schei et al., 2024	Perceptions and Use of AI Chatbots among Students in Higher Education	1
Sergeeva et al., 2024	Gamified learning in higher education	2
Setiamurti & Kurniawati, 2024	Fostering Creativity in Higher Education Institution	4
Shahjahan & Seinn, 2025	'Decolonizing' engineering education	2
Singun, 2025	Unveiling the barriers to digital transformation in higher education institutions	2
Solis-Garcia et al., 2024	Mental Health of University Students with Disabilities	3
Solis-Garcia et al., 2025	Inclusion of students with disabilities in the European Higher Education Area (EHEA)	3
Stamou et al., 2024	Counseling for non-traditional students in formal higher education	3
Subki, 2025	Green Campus Management Sustainability	6
Sukjairungwattana et al., 2024	Higher Education Internationalization in Asia	7
Tang et al., 2024	Translanguaging strategies in the English writing classroom in higher education	2
Tang et al., 2025	Languages and the Internationalization of Higher Education in China	7
Tareke et al., 2024	Ethiopian Public Higher Education	7
Tareke et al., 2025	Students' academic success in higher education in the digital era	5
Teng & Cosier, 2024	Influences of cultural capital and internationalization on global competence in higher education	7
Thiedig & Wegner, 2024	Evidence use in higher education decision-making and policy	8
Tikhonova & Raitskaya, 2024	Culture of Research	3, 7, 8
Tillmanns et al., 2025	GenAI in Higher Education	1
Trujillo-Juárez et al., 2025	Teacher digital competence in higher education	2
Valencia Quecano et al., 2024	Dropout in postgraduate programs	5
van der Wee et al., 2024	Teaching and learning in sustainability-oriented Living Labs within higher education	6
Veres et al., 2025	Sustainable Universities	6
Vinueza-Morales et al., 2025	Teaching programming in higher education	7

Review reference	Theme	Code
Vuoriainen et al., 2024	Successful higher education-industry collaboration in engineering education	2
Wang & Abdullah, 2024	Enhancing Students' Critical Thinking Through Mathematics in Higher Education	4
Wang & Ishak, 2025	Social-Emotional Learning among College Students	3
You et al., 2024	Global research trends on higher education leadership development	7, 8
Zhao & Selvaratnam, 2024	Reform of vocational education in China	2

APPENDIX 2

Types of reviews and application of Frameworks & Protocols

Review Reference	Type of the review as stated in the title or in the text	PICO(T)-PPC & PRISMA PAGER
Abdallah et al., 2025	systematic	PRISMA
Adauto Medina et al., 2024	bibliometric	PRISMA
Ahmad & Khurizan, 2024	systematic & bibliometric	-
Alonso et al., 2025	systematic	PRISMA
Alotaibi, 2025	not identified	PRISMA
Alvado, 2025	systematic	PRISMA
Álvarez-Martinez et al., 2025	systematic	PRISMA
Amavasi & Zimmerman, 2024	not identified in the title, integrative	PRISMA
Andrade-Girón et al., 2025	not identified in the title, systematic	PRISMA
Balalee, 2025	systematic	PRISMA, PICO
Banarjee et al., 2024	systematic	PRISMA
Bannigan et al., 2025	scoping	PCC, PRISMA
Barikzai et al., 2024	scoping	PRISMA
Barua & Lockee, 2024	exploratory literature review	-
Basheer et al., 2024	comprehensive and bibliometric review	PRISMA
Bayly-Castaneda et al., 2024	systematic	PRISMA
Benson et aL., 2024	systematic	PRISMA
Bonilla-Jurado et al., 2024	not identified in the title	PRISMA
Brauer et al., 2024	systematic	-
Buele & Llerena-Aguirre, 2025	not identified in the title, narrative	-
Burton et al., 2024	systematic	PICO, PRISMA
Bustamante-Mona et al., 2025	not identified in the title, systematic mapping	-
Cabrera-Duffaut et al., 2024	systematic	PRISMA
Cadiz, 2024	bibliometric	-
Castillo et al., 20244	bibliometric	-
Castillo-Martinez et al., 2024	systematic	PRISMA
Chashmyazdan et al., 2024	literature review	-
Chen et al., 2024	systematic	PRISMA
Chugh et al., 2025	literature review, systematic narrative literature review	-
Coleman et al., 2025	integrative	PRISMA
Cui & Alias, 2024	systematic	PRISMA
Din Eak & Annamalai, 2024	systematic	-
Dukes III et al., 2024	not identified in the title, systematic	PRISMA
Edvardsen Tonheim et al., 2024	scoping	PRISMA
El Aatik et al., 2024	systematic	PRISMA, PICO
Eltaiba et al., 2025	systematic	PRISMA

Review Reference	Type of the review as stated in the title or in the text	PICO(T)-PPC & PRISMA PAGER
Esteban, 2025	bibliometric	PRISMA
Fengye et al., 2025	bibliometric	-
Fernandez-Batanero et al., 2024	not identified in the title	PICO, PRISMA
Ferreira Santos, 2024	scoping	PRISMA
Frez-Pulgar et al., 2025	systematic mapping	PICO
Galdames-Calderón et al., 2024	systematic	PRISMA, PICO
Ghazian & Lortie, 2024		-
Gomez et al., 2025	not identified in the title, systematic narrative review	-
Goncalves et al., 2024	not identified in the title	-
Gudoniene et al., 2025	systematic	PRISMA
Harmon et al., 2024	scoping	PRISMA, PAGER, PCC
Heintalu et al., 2025	systematic	PRISMA
Henry et al., 2024	systematic mapping study	-
Hidayat et al., 2024	systematic	PRISMA
Ishmuradova et al., 2024	bibliometric	-
Isiaku et al., 2024	not identified in the title, comprehensive literature review	PRISMA
Jaxin et al., 2024	systematic	PRISMA
Junaštíkova, 2024	review of empirical studies	-
Kalim et al., 2025	systematic	PRISMA
Kalocsányiová et al., 2024	multilingual systematic review	PRISMA
Karimi & Khawaja, 2025	systematic	PRISMA
Kaymakcioglu & Thomas, 2024	systematic	PRISMA
Klimova & Chen, 2024	not identified in the title, systematic	PRISMA
Kovari, 2025	systematic	PRISMA
Liang et al., 2025	systematic	PRISMA
Lim & Lee, 2024	systematic	-
Liu et al., 2025	systematic	PRISMA
Lopez-Chila et al., 2024	bibliometric	-
López-Nuñez et al., 2024	systematic	PRISMA
Lou & Zhou, 2024	systematic	PRISMA
Ma, 2025	systematic	PRISMA
Marano et al., 2024	pragmatic scoping	PRISMA
Mare & Mutezo, 2025	systematic	PRISMA
Marques et al., 2024	not identified in the title, systematic	PRISMA
Masalimova et al., 2024	bibliometric	PRISMA
Mashingaidze & Mayayise, 2025	systematic	PRISMA
Mireku & Bervell, 2024	systematic	PRISMA
Mireku et al., 2024a	systematic	PRISMA
Mireku et al., 2024b	systematic	PRISMA

Review Reference	Type of the review as stated in the title or in the text	PICO(T)-PPC & PRISMA PAGER
Mohammed & Ozdamli, 2024	systematic	PRISMA
Mursalzade et al., 2025	systematic	PRISMA
Nalweyiso et al., 2025	systematic	PRISMA, PICOT
Ncube & Ngulube, 2024	systematic	PRISMA, PICO
Ndibalema, 2025	systematic	not specified as 'PRISMA', though a PRISMA flowchart included
Nieminen et al., 2024	meta-ethnographic review	-
Ogunleye et al., 2024	systematic	PRISMA
Olivares-De la fuente et al., 2025	systematic	-
Ortega-Ruiperez & Correa-Gorospe, 2024	systematic	PRISMA
Oulamine et al., 2025	systematic	PRISMA
Pallaris et al., 2024	systematic concept mapping	-
Parambil et al., 2024	not identified in the title	-
ParedesCanencio et al., 2024	systematic	PRISMA
Peláez-Sánchez et al., 2024	not identified in the title, systematic	-
Pelaez-Sanchez et al., 2024	not identified in the title, systematic	PRISMA
Punch et al., 2025	scoping	PRISMA
Rahajeng et al., 2024	not identified in the title	PRISMA
Raitskaya & Tikhonova, 2024	scoping	PRISMA, PPC
Raitskaya & Tikhonova, 2025	scoping	PRISMA, PPC
Renfors, 2024	overview of the current state	-
Rosales-Ricardo & Caceres-Manzano, 2024	systematic	PRISMA
Rosario & Raimundo, 2024	bibliometric systematic	-
Roy et al., 2025	not identified in the title	PRISMA
Saez-Zevallos & Montalvo-Apolín, 2025	mapping	PRISMA
Sahar & Munawaroh, 2025	bibliometric	PRISMA
Schei et al., 2024	scoping	PRISMA
Sergeeva et al., 2024	bibliometric	-
Setiamurti & Kurniawati, 2024	systematic	PRISMA
Shahjahan & Seinn, 2025	integrative review	-
Singun, 2025	systematic	PRISMA
Solis-Garcia et al., 2024	systematic	PRISMA
Solis-Garcia et al., 2025	systematic	PRISMA
Stamou et al., 2024	scoping	PRISMA
Subki, 2025	bibliometric	-
Sukjairungwattana et al., 2024	systematic	PRISMA
Tang et al., 2024	systematic	PRISMA
Tang et al., 2025	not identified in the title	PRISMA
Tareke et al., 2024	not identified in the title	-

Review Reference	Type of the review as stated in the title or in the text	PICO(T)-PPC & PRISMA PAGER
Tareke et al., 2025	systematic	PRISMA
Teng & Cosier, 2024	systematic	PRISMA
Thiedig & Wegner, 2024	scoping	PRISMA
Tikhonova & Raitskaya, 2024	systematic scoping	PPC, PRISMA
Tillmanns et al., 2025	systematic	PRISMA
Trujillo-Juárez et al., 2025	systematic	PRISMA, PICO
Valencia Quecano et al., 2024	scoping	PRISMA
van der Wee et al., 2024	not identified in the title, realist review	PRISMA
Veres et al., 2025	bibliometric	PRISMA
Vinueza-Morales et al., 2025	bibliometric	-
Vuoriainen et al., 2024	systematic	PRISMA
Wang & Abdullah, 2024	systematic	PRISMA
Wang & Ishak, 2025	systematic	-
You et al., 2024	bibliometric	PRISMA
Zhao & Selvaratnam, 2024	systematic	PRISMA

APPENDIX 3

Extracted Data on Further Research from the Reviews

Review Reference	Further Research as Stated in the Review
Abdallah et al., 2025	Future research should include longitudinal studies to assess long-term impacts on learning and critical thinking. Greater focus is also needed on underrepresented regions, emotional well-being, and discipline-specific applications. Robust, transparent frameworks are essential to ensure responsible use that prioritizes academic integrity and inclusivity
Adauto Medina et al., 2024	Future research should focus on systematic literature reviews that delve into key areas such as the impact of ChatGPT on the development of specific academic writing skills, risks related to academic integrity and ethics, and its role in educational assessment
Ahmad & Khurizan, 2024	Bibliometric investigations. future researchers might consider evaluating works that were not included in the current analysis. could benefit from creating new techniques and utilizing new keywords to incorporate additional relevant publications for more comprehensive metadata analysis
Alonso et al., 2025	It would be convenient to analyze BL interventions in countries where these problems exist
Alotaibi, 2025	Future research should focus on quantifying the long-term sustainability impacts of AI–LMS integration and developing implementation frameworks that balance technological advancement with sustainable development objectives
Alvado, 2025	As a recommendation, it is essential to continue promoting the research and implementation of the DT in university education. This involves developing training programs for educators, promoting interdisciplinary collaboration between faculties and departments, and fostering pedagogical innovation based on this methodological approach. In addition, it is suggested to continuously evaluate the impact of design thinking on students' learning and development in order to adjust and improve its application in the university environment
Álvarez-Martinez et al., 2025	Future research endeavors should dedicate attention to assessing the viability of MALL to acquire new languages, while also exploring any potential variations in its efficacy across different languages
Amavasi & Zimmerman, 2024	This integrated review also found opportunities for future research addressing newly graduated nurses' IPC core competency throughout their induction or probation period to comprehend the challenges and strategies needed during the transition phase. An interventional control study of a group of newly graduated nurses on IPC core competency and their reflective learning during the early stages of their nursing career is recommended to monitor how nurses adapt and adopt IPC standards
Andrade-Girón et al., 2025	This research highlights the need for further study and future analysis to address variations in results based on different prompts or words used with ChatGPT and the potential impact on student satisfaction and effectiveness. Likewise, it opens opportunities for future exploration and improvement in designing and implementing AI-assisted learning systems, ensuring their optimal use and addressing concerns and difficulties that students may face. Therefore, it is recommended that experimental research be carried out with more rigorous criteria in the selection of the sample and in the application of measurement instruments to guarantee the validity and reliability of the results obtained in the research
Balalee, 2025	Future research can be conducted to determine the long-term applications of VR, AR, and MR in education. This type o research will help elucidate the immense impact of technology on education in the long run... more research is needed to determine how VR, AR, and MR applications are impacting traditional classroom settings in the digital domain. Research in VR, AR, and MR must explore applications across subject domains such as science, technology, engineering, and mathematics, as well as the humanities and health. Continuous research will be needed to stay up to date with new developments in VR, AR, and MR...
Banarjee et al., 2024	Not identified
Bannigan et al., 2025	It is alsowhole-university approach is one that considers all actors within HE. This is something that future research could explore. important to address the role of educators both in delivering well- being content but also in relation to the well-being of HE staff themselves. For educators to effectively impart well-being knowledge on students, they must be able to take care of their own well-being first (Querstret, 2019). It may also be worth considering the roles of other staff members and actors within the HE ecosystem
Barikzai et al., 2024	Future research should focus on exploring adaptive strategies to address these growing challenges, assessing the long-term impact of e-learning on academic outcomes, and examining the effectiveness of specific interventions in overcoming infrastructure-related barriers in various emerging contexts

Review Reference	Further Research as Stated in the Review
Barua & Lockee, 2024	Future research in this area can look more closely into the optimal degree of flexibility across the dimensions of flexible learning and develop a better understanding of how much choice is beneficial without causing confusion for learners and instructors. Further research is also needed to investigate the constraints of flexible course design and delivery including but not limited to access, connectivity, institutional policies, legal regulations, and the logistical and administrative barriers that can hinder the design and implementation of flexible courses in higher education
Basheer et al., 2024	Increasing research on student acceptance of the sustainability initiatives and possible changes that can be implemented to speed up the process
Bayly-Castaneda et al., 2024	Future lines of research are proposed among which are the determination of the factors that influence the effectiveness of personalization of learning as well as the ethical implications of this development to ensure equity and non-discrimination in access to these solutions in order to promote the democratization of learning
Benson et aL., 2024	Future research could focus on multi-faceted interventions that combine these approaches to upskill academics and lift the quality of BL across higher education institutions
Bonilla-Jurado et al., 2024	Future research should explore the effectiveness of ESD in diverse populations, including adolescents and older adults, to understand its applicability and benefits in a broader context. ... it is essential to continue to investigate methodological variations and demographic factors that may influence educational outcomes, as well as the impacts of different educational protocols and participant characteristics on observed effects. In addition, it is crucial to increase research on ESD implementation in the Global South, identifying strategies to overcome economic and infrastructural constraints and adapting successful approaches from the Global North to local contexts
Brauer et al., 2024	We thus encourage future researchers to zoom in on the interplay of our proposed themes to study how teachers apply different creativity-fostering behaviors interactively. The meta-creativity model could provide the structure for such a comprehensive measurement tool. We hope to spark researchers' ideas about studying and applying transdisciplinary approaches to fostering creativity within domain-specific interdependencies
Buele & Llerena-Aguirre, 2025	Future reviews should consider broader inclusion criteria and adopt dynamic frameworks that respond to the evolving nature of AI in education
Burton et al., 2024	Further research is required to better understand how online Team Based Learning approaches, now adopted as mainstream, impact knowledge, learning processes, academic outcomes, and the perceived experiences of students in health professional courses
Bustamante-Mona et al., 2025	Building on this conceptual mapping, future studies are encouraged to empirically assess how emotional intelligence, resilience, and motivation influence learning outcomes in various STEM environments. More research is needed on the best practices for training educators in inclusive pedagogical methods, particularly those that incorporate socio-emotional learning and gender-sensitive approaches. Investigating the long-term effects of inclusive STEM education on career choices and workforce participation would provide valuable insights into the effectiveness of these models
Cabrera-Duffaut et al., 2024	Expansion of Applications in Diverse Disciplines: More research is needed to explore the use of VR beyond health sciences, opening potential for its application in areas such as the exact and social sciences, arts, and humanities. Long-Term Studies: More longitudinal studies that assess the effects of implementing VR in higher education over time would be beneficial, allowing a better understanding of its long-termimpact and effectiveness. Impact Assessment on Digital Competence: Investigating how the implementation of VR affects the digital competence of students and educators could provide valuable insights for the development of training programs and technological adaptation in educational institutions
Cadiz, 2024	Not identified
Castillo et al., 20244	Future research should be directed towards a deeper understanding of the effectiveness of these environments in developing specific skills such as teamwork, leadership, communication, as well as creating strategies that promote inclusive accessibility. This will not only maximize the value of AI and metaversesin higher education, but will ensure that their impact contributes equitably to theadvancement of learning in an increasingly digitalized world
Castillo-Martinez et al., 2024	Not identified
Chashmyazdan et al., 2024	Not identified
Chen et al., 2024	future research could explore lecturers' organizational commitment in specific cultural or regional contexts to provide more specific insights.

Review Reference	Further Research as Stated in the Review
Chugh et al., 2025	The use of GenAI to evaluate programming code efficiency as a potential study; future research could explore leveraging generative AI (GenAI) to deliver interactive, conversational feedback that emulates the guidance typically provided by human teaching assistants
Coleman et al., 2025	Care is needed to ensure that such studies provide insight into the various ways in which such barriers impact women from different contexts, together with the efficacy of alternative strategies in addressing these
Cui & Alias, 2024	Future research should focus on developing and establishing guidelines for the ethical and effective use of AI technology
Din Eak & Annamalai, 2024	Research should focus on factors like student engagement, the effectiveness of feedback across different learning styles and the impact of digital literacy on the feedback process
Dukes III et al., 2024	Scholars interested in intervention or service delivery might conduct longitudinal studies to examine the evolution of service programs, following faculty, staff and students over time to determine access, effectiveness, and development of activities, programs, and policies intended to benefit students with disabilities. Furthermore, scholars might examine intended and unintended consequences of activities, programs, and policies for the general university population on students with disabilities
Edvardsen Tonheim et al., 2024	Further research is needed to investigate the different facilitators and barriers in more detail including the educators' perspective
El Aatik et al., 2024	Comparative research with international institutions would further help identify which interventions are most effective in this context
Eltaiba et al., 2025	Suggestion for future studies may include areas such as: expanding or narrowing down geographical scope to provide insights into how TEL can be customized to address diverse regional or individual-specific countries barriers. Future research could investigate TEL's impact in secondary and basic education settings, conduct longitudinal studies and multi-faceted impact studies
Esteban, 2025	Future research should broaden its inclusion criteria by extending the time period, considering the addition of more keywords relevant to the focus of the study, and utilizing additional databases to capture a more comprehensive dataset. Future studies may also consider the inclusion of other manuscripts, such as articles conference proceedings and books, to gain more valuable insights about the field
Fengye et al., 2025	Future research will focus on three main areas: optimizing AI teaching systems to guide research, developing interactive learning modules based on clinical problems; establishing cross-regional medical education and research collaboration platforms to regenerate high-quality resources via cloud-based teaching research communities; and designing dynamic evaluation systems to accurately track the effectiveness of research interest cultivation
Fernandez-Batanero et ak., 2024	Among the future lines of research, it is intended to continue investigating how the so-called emerging technologies, including the IoTs, Blockchain or Big Data, can transform higher education institutions
Ferreira Santos, 2024	Not specified
Frez0Pulgar et al., 2025	A systematic and collaborative approach to future research could significantly improve the implementation and impact of bioethics in higher education
Galdames-Calderón et al., 2024	Subsequent studies should focus on elaborating and detailing the specific techniques that can facilitate the successful application of these identified practices. By providing a more detailed understanding of these methods, future research can contribute significantly to the refinement of CBL pedagogy, thereby enhancing the learning experiences and outcomes of students engaged in this innovative educational approach
Ghazian & Lortie, 2024	Future research should focus on evaluating the long-term impact of embedding the SDGs in course curricula. Longitudinal studies could assess how student engagement with SDGs translates into sustainable practices in their personal and professional lives post-graduation. Furthermore, emerging digital tools and platforms for teaching, such as virtual reality and artificial intelligence, could enhance experiential learning in sustainability. Future research could examine how these tools can be leveraged to simulate real-world challenges and promote active problem-solving related to the SDGs
Gomez et al., 2025	Less attention paid to physiological factors...a niche for future research is identified here
Goncalves et al., 2024	Further research is needed to understand contextual issues. It would be interesting to replicate the study in different contexts and perform meta-analysis and meta-synthesis to better understand the effects. The researchers also verified complementary keywords and used a method to mitigate other possible limitations
Gudoniene et al., 2025	Not identified
Harmon et al., 2024	Further research is also required on the experiences and perceptions of reflection for this cohort of HE academics

Review Reference	Further Research as Stated in the Review
Heintalu et al., 2025	Future studies could test and use the proposed IGSO (integrated goal setting and orientation) theory to interpret students' goal setting and goal-pursuit processes, and to develop interventions and predict and test their mechanisms.
Henry et al., 2024	With the trend of moving to playful and experiential learning in education in a post pandemic era, gamification bears significant influence in shaping the future of education
Hidayat et al., 2024	Not identified
Ishmuradova et al., 2024	Not identified
Isiaku et al., 2024	It encourages interdisciplinary collaboration among researchers from various elds, such as education, computer science and ethics, to holistically explore the implications of ChatGPT adoption and foster innovation
Jaxin et al., 2024	Future research should address these limitations by conducting longitudinal studies and considering larger sample sizes for experimental participants
Junaštíkova, 2024	Not identified
Kalim et al., 2025	Future research needs to examine the relationship between women's AI adoption and their age and economic situation, as literature is very limited in this regard in Asia. Future studies and reviews can also include published reports of AI adoption among women in Asia from government and non-government organizations to integrate their views and remarks to get more in depth insights about the women's situation in the Asian context
Kalocsányiová et al., 2024	Future e orts should establish the impact of the discussed measures and recommendations for HE enrolment and study completion, while continuing to consider the voices, perceptions and input of refugee-background students and other stakeholders. There is also a need for more robust research into support mechanisms in later stages of the HE study cycle. Most research to date has failed to look beyond welcome and preparatory programmes, even though many of the disadvantages refugee-background students face are likely to persist throughout their HE journey
Karimi & Khawaja, 2025	Future studies should adopt mixed methods approaches and explore longitudinal interventions to examine how sustained training programmes influence digital skill development over time
Kaymakcioglu & Thomas, 2024	It is recommended that future research shift its focus towards examining the implementation of such strategies, exploring success stories, and understanding the characteristics of resistance and failures in achieving objectives. Additionally, future research could broaden its scope by including more LMICs from Sub-Saharan Africa.
Klimova & Chen, 2024	One important direction for future research is the exploration of the long-term impact of AI on ICC. Exploring how AI can be integrated into broader educational frameworks to support holistic cultural education
Kovari, 2025	Future research should investigate how AI can better foster deeper collaboration in diverse groups where challenges may be at odds with equitable participation. Future research should address how AI can ensure cohesion in the group, equal participation from the members, and effective collaboration collectively in order to realize its full potential in higher education institutions. Symbolic AI and hybrid systems, which merge approaches based on reasoning, therefore have immense potential for further development along this direction. For that reason, future research should be directed into probing these underrepresented AI methodologies so that application of AI technologies in education will be holistic. However, with a more expanded notion of AI for educators and research scholars, the true potential of AI to augment collaborative learning and ameliorate educational outcomes may be seen
Liang et al., 2025	Future studies should explore diverse OPF configurations, investigate long-term effects on EFL writing development, and examine the integration of emerging technologies like AI and VR in multicultural learning environments
Lim & Lee, 2024	Future research might, for example, develop a pervasive tool for students in higher education and investigate students' acceptance and intention to use it in the learning process. In addition, future research can consider investigating students' use of pervasive tools at the primary school level. Studies relating to the population at the primary school level were excluded from this study
Liu et al., 2025	Future studies should adopt more longitudinal research methods to track the development process of college students' entrepreneurial competence and combine experimental research with follow-up surveys to more closely analyze the long-term impact of different educational interventions. Future research should further combine multi-variate data, longitudinal analysis, mixed-method research, and a comparative analysis of different education systems to enhance the comprehensive understanding of the development path of students' entrepreneurial competence and to promote more targeted policy optimization and education reform practices

Review Reference	Further Research as Stated in the Review
Lopez-Chila et al., 2024	In future research, the authors proposed to undertake an in-depth bibliographic analysis in order to explore the results of the research and methodologies used to measure the specific impact of AI in higher education and to consider the influence of global events and international collaborations
López-Nuñez et al., 2024	On the one hand is the international comparison of the digital competences of teachers in higher education institutions with the same instrument, being able to make correlations according to different characteristics. On the other hand, is the generation of practical and training experiences for trainee teachers on digital competences by trying to establish practical tasks to demonstrate the competences obtained
Lou & Zhou, 2024	Future research could enhance understanding of individual and contextual factors influencing self-regulated learning strategy application through longitudinal studies (Green et al., 2022), as well as case studies and interviews
Ma, 2025	Research should focus on exploring interdisciplinary collaboration to address the multifaceted challenges posed by ChatGPT … This includes addressing its inherent deficiencies, mitigating the ethical implications, and harnessing potential of ChatGPT to revolutionize teaching and learning practice. Research in this realm should carry out more empirical experiments that based on effective research methods, in an effort to examine student learning outcomes and offer insights into measures to enhance the integration of AI into higher education
Marano et al., 2024	Some further research on these variables could be helpful for education providers when implementing remote proctoring (e.g. privacy laws and proctoring)
Mare & Mutezo, 2025	It is recommended that future studies focus on the impact of social regulation within the community of inquiry to enhance student engagement. More studies on social regulation and how it links with self- and co-regulation are required
Marques et al., 2024	First, we recommend longitudinal and comparative studies to help understand better how academic identities develop over time and vary across different spatial dimensions. Second, we think that studies that explore the links between identity differences and responses to managerialist university contexts (e.g. considering differences in disciplinary group traditions, generational perspectives, and career stages) would be useful. There is a need, too, for further research into academic identity and higher education change from an intersectional perspective, exploring specific local, regional and cultural factors, including gender, ethnicity, and religion. Finally, given that most of the studies we included used qualitative methodology, the generation of studies using mixed method designs might further enrich investigations in this field
Masalimova et al., 2024	Future bibliometric studies should focus on conducting detailed analyses of the areas that have shown significant relevance for online learning post pandemic… Additionally, content analysis may provide a more detailed understanding of scientific discourse by exploring topics and contexts in bibliometric studies. Moreover, conducting future analyses and comparisons would be beneficial in exploring the differences between countries with different social and educational contexts…Longitudinal studies could also help understand the development of online learning trends
Mashingaidze & Mayayise, 2025	Future research can also consider the perspectives of stakeholders other than industry and students
Mireku & Bervell, 2024	Experts in the field of quality assurance promotion should increase their efforts in searching and publishing on the subject and make recommendations for growing the educational institutions in the sub-region…Many more researches need to be conducted to predict possible challenges that may, in the future, disturb the systems for achieving quality assurance in higher education
Mireku et al., 2024a	We suggest that future studies look into these aspects of the phenomenon in higher education in SSA in order to fill this gap and inform policy and practice
Mireku et al., 2024b	Educational institutions should embrace more research to uncover how advanced technologies, such as CCTV cameras, biometrics, and online assessment practices, could be used to supervise examinations…It is suggested that future studies could also delve into how the use of artificial intelligence is promoting academic dishonesty among students and faculty
Mohammed & Ozdamli, 2024	Researchers should develop instructional designs that align with the results identified for sustainable education
Mursalzade et al., 2025	Future research and interventions aimed at enhancing psychological flexibility may thus serve as strategic levers for sustainable development in educational contexts
Nalweyiso et al., 2025	Future researchers may focus on conducting quality research with an emphasis on the long-term educational effects on students. Future research on teaching EBP to undergraduate healthcare students in developing countries need to concentrate on studying the impact of EBP education on long-term EBP knowledge, skills, attitudes, and behaviours utilising rigorous techniques and assessment tools

Review Reference	Further Research as Stated in the Review
Ncube & Ngulube, 2024	Future research could delve deeper into the efficacy of specific techniques within the context of student privacy and data sustainability. Longitudinal studies could track the impact of a particular technique (e.g., learning analytics dashboards, social network analysis) on student learning over time, examining both academic outcomes and potential privacy concerns. Additionally, comparative studies could explore the relative effectiveness of different techniques for achieving specific learning objectives within different postgraduate programmes, while also considering their long-term data management requirements and impact on student trust
Ndibalema, 2025	Future research should investigate long-term trends and evaluate the effectiveness of digital literacy initiatives across varied educational contexts
Nieminen et al., 2024	We call for future research and practice that focuses on examining how assessment shapes student identities, and how it may create 'othered' student identities
Ogunleye et al., 2024	Future research should be focused on interdisciplinary studies to develop guidelines for GenAI usage in HE. Experimental comparisons of advanced GenAI toolslike Gemini and the performance of AI content detectors in plagiarism systems will be explored. Comparative studies should be conducted to assess the effectiveness of GenAI tools in educational settings, accurately
Olivares-De la fuente et al., 2025	A potential future line of research would be to incorporate qualitative studies that explore in depth the experiences of students and teachers with the use of Twitter and YouTube as educational tools, allowing for a more enriching and contextualized analysis
Ortega-Ruiperez & Correa-Gorospe, 2024	It is therefore recommended that future literature reviews be conducted with a more open search
Oulamine et al., 2025	Longitudinal studies would be advantageous for observing evolutions and changes in attitudes over time, thus providing more robust data and more precise recommendations for the future of e-learning in higher education...Future studies would benefit from using interdisciplinary approaches and mixed methodologies to understand these resistances better, taking into account the diversity of actors and contexts
Pallaris et al., 2024	More future research is needed in the areas of the advancement of making culture in academic Makerspaces to enhance active and wide students and teaching personnel participation; Makerspace's academic effects through design learning experiences by using Makerspaces in study programs curriculum in the form of making co-curricular assignments and group projects; and Academic Makerspaces effectiveness on student's professional career prospect through longitudinal studies
Parambil et al., 2024	Future research directions include developing benchmark datasets specific to cybersecurity in online education, promoting interdisciplinary collaboration, and exploring emerging technologies such as GNNs and blockchain
ParedesCanencio et al., 2024	Future studies could focus on assessing the carbon footprint of specific emission sources and the specific scope of HEIs, reporting the factors that may make it difficult to compare the results of their study with results from other studies, so that consistent comparisons can be made between different institutions around the world, and inspire more institutions to report their carbon footprints while establishing a global standard. The results of upcoming carbon footprint studies may further provide the information to develop such a standard
Peláez-Sánchez et al., 2024	Future research should focus on exploring how advanced technologies can facilitate the development of specific competencies such as emotional, collaborative, complex problem-solving skills, and critical thinking
Pelaez-Sanchez et al., 2024	Future research aiming for a more global scope should include studies in various languages
Punch et al., 2025	Further research should explore the long-term influence of these recommendations on student retention and success, investigate the effectiveness of specific assistive technologies, and examine the benefits of peer support programmes. Additionally, studies must address the barriers faced by students with disabilities in disclosing their disabilities and accessing support services, providing further insights for improving university policies and practices
Rahajeng et al., 2024	Students with mental and intellectual disabilities may not exhibit overt or visible signs of disability, which could contribute to their under-recognition and marginalization in research on SwD in Indonesia. Consequently, the findings of this review might not fully capture the experiences and challenges faced by students with mental and intellectual disabilities. This gap underscores a significant area for future research
Raitskaya & Tikhonova, 2024	Further and regular (at least yearly) reviews are essential for the field that is evolving fast New knowledge is added monthly with an unclear though impressive perspective of the GenAI appliances in the long run

Review Reference	Further Research as Stated in the Review
Raitskaya & Tikhonova, 2025	Future research should address these gaps in several directions. First, longitudinal studies are needed to investigate the enduring effects of GenAI-supported instruction on critical thinking across diverse educational contexts and learner populations. Second, there is a clear need to develop more robust theoretical models that explicitly connect the dimensions of critical thinking, as defined in Bloom's taxonomy, with the specific learning mechanisms that are activated through the use of generative AI. Third, further work should focus on evaluating how different forms of instructional support, including scaffolded AI use and targeted AI literacy training, influence students' cognitive outcomes and their ability to engage with AI critically and productively
Renfors, 2024	Further studies with a focus on these matters are suggested because more research is needed to explore the connection between competencies and teaching and learning approaches, as well as a moreanalytical approach to critically examine the effectiveness of using these specific approaches to improve their use
Rosales-Ricardo & Caceres-Manzano, 2024	Not identified
Rosario & Raimundo, 2024	Research might focus on the effectiveness of interdisciplinary curricula and teaching methods and investigate methodologies for assessing the impact of sustainable entrepreneurship education in broader geographical areas
Roy et al., 2025	Future research should empirically test the framework in different settings while solving these practical problems for its inclusive and sustainable implementation in education.
Saez-Zevallos & Montalvo-Apolín, 2025	In light of the critical role of English proficiency in higher education, it is recommended to conduct new research on the construct using SMS and systematic reviews. Additionally, evaluations of the design and implementation of strategies aimed at enhancing the competence in writing and publishing scientific articles in English are suggested…In light of the critical role of English proficiency in higher education, it is recommended to conduct new research on the construct using SMS and systematic reviews
Sahar & Munawaroh, 2025	Future research should incorporate multiple databases to provide a more comprehensive perspective on AI adoption in higher education. Future studies should include multilingual sources to capture a more diverse global understanding of AI's impact on higher education. Future research should complement these methods with qualitative approaches, such as case studies, surveys, or interviews, to explore faculty perceptions, institutional challenges, and student experiences with AI. Future studies should consider a broader range of academic and institutional sources to gain deeper insights into AI applications in education.
Schei et al., 2024	We also have limited knowledge about which disciplines, countries, and research methods are prevalent in this field of research. Gaining such knowledge holds critical value not only for policymakers, educational institutions, educators, and students but also for guiding future research
Sergeeva et al., 2024	Future research has to find the impact of instructors' competence and how they decide to use a gamified learning style as they will determine the effectiveness of this style
Setiamurti & Kurniawati, 2024	Researchers are encouraged to investigate and incorporate studies from a broader range of cultural and regional contexts to achieve a comprehensive and nuanced understanding of fostering creativity in HE worldwide
Shahjahan & Seinn, 2025	Future efforts should prioritize exploring how DEE is understood in diverse geographical and transnational contexts to include voices from historically neglected regions. Addressing these gaps will be crucial in ensuring that DEE is both inclusive and globally relevant, and in broadening the conversation beyond the current limitations of the field
Singun, 2025	Future research will aim to confirm the findings of this review through studies conducted in HEIs. To advance this research, it would be beneficial to consider how specific legislative frameworks, and the governance structures of public sector entities influence barriers to DT
Solis-Garcia et al., 2024	Not identified
Solis-Garcia et al., 2025	Future research should support current findings and implement necessary actions to improve inclusion
Stamou et al., 2024	Future studies could focus on answering review-related questions by examining more databases and broadening their focus to incorporate different types of publications, such as book chapters, dissertations, and conference proceedings
Subki, 2025	Future research should address key gaps identified in the current literature. Specifically, there is a need for deeper investigation of the behavioral, cultural, and policy-related dimensions of green campus management. Comparative studies across regions and time periods could provide valuable, context-specific insights, while integrating qualitative approaches with bibliometric methods may offer a more comprehensive understanding of sustainability practices.

Review Reference	Further Research as Stated in the Review
Sukjairungwattana et al., 2024	Future research should further explore how to strike a balance between internationalization and local culture to promote more inclusive and sustainable educational development. Additionally, the research should focus on the unique challenges and opportunities faced by di□erent countries and regions in the internationalization process, in order to provide more targeted recommendations and solutions for policymakers and educators
Tang et al., 2024	It is also hoped that future research will explore and help students' English writing, such as whether there are effective translanguaging teaching methods provided for teachers to help students learn English writing, how students use translanguaging strategies in their writing process, whether translanguaging helps students improve English writing proficiency, and what are the attitudes of students towards translanguaging
Tang et al., 2025	To address these gaps, future research should incorporate institutional case studies that analyse internal language policy documents and their alignment—or misalignment—with national directives. Ethnographic or interview-based research involving international students, faculty members, and university administrators could provide rich insight into how language policies are interpreted, negotiated, and resisted in everyday academic life. Additionally, studies comparing Chinese institutions with those in other non-Anglophone contexts would illuminate how different systems manage the balance between English dominance and national language promotion. Finally, longitudinal research tracking the implementation and outcomes of bilingual or multilingual policies over time could offer deeper understanding of their long-term implications for equity, identity, and international engagement
Tareke et al., 2024	Further investigations need to be conducted on the higher education system in the country to focus on integrating technology and innovation in higher education, as well as to what extent those emerging initiatives are working and contributing to the alignment of public university missions with sustainable development
Tareke et al., 2025	Further investigations are needed using diverse methodologies and theoretical frameworks to more deeply examine the dynamic between technology, education, and academic success. Researchers need to balance the broader theoretical frameworks with variable-specific frameworks and bridge the "generational gaps" through theoretical development or reevaluating the foundational academic success theories in the digital context
Teng & Cosier, 2024	Not identified
Thiedig & Wegner, 2024	Studies would benefit from making more explicit whether they intend to study the extent of evidence use, qualitative aspects of evidence use, or its influence/impact, and how this is defined and measured…A more systematic and comparative investigation of sources of evidence would enable it to derive conclusions about how the choice of sources might impact evidence use, and whether this is determined by the (un)availability of evidence, or rather by the explicit comparison and assessment of different sources…More deliberate and independent comparative studies including a larger number of organisations (or organisational units) are needed to prevent bias resulting from the selection of good practice examples as empirical cases. In doing so, the specific organisational characteristics of HERI, and their effects on the use context, should be reflected and, ideally, varied in a systematic manner. Country-specific factors, such as the regulatory environment or incentive and funding structures, should be noted, and their implications for the generalisability of the findings discussed. At the same time, the high number of local or national studies currently leaves potential for international comparative analyses and collaboration unused
Tikhonova & Raitskaya, 2024	A review of publications dated back to the 1990s and later may add to the general understanding of the field. Although the Scopus database is comparatively comprehensive, other bases may broaden today's views of the problem field
Tillmanns et al., 2025	While the need for ethical guidelines is widely recognised, there is limited research on the long-term effects of GenAI integration on student learning outcomes and faculty performance. Furthermore, the potential negative consequences of GenAI, such as technostress and over-reliance on AI tools, remain underexplored, particularly in the context of higher education. Future research should address these gaps, focusing on the development of scalable, adaptable policy frameworks and effective training models that can be implemented across diverse educational settings
Trujillo-Juárez et al., 2025	Future research should investigate the potential of integrating learning analytics and AI-driven personalization into micro-course design. Such technologies can adapt content in real time based on individual teacher needs, engagement patterns, and performance, thereby enhancing the relevance and effectiveness of professional development
Valencia Quecano et al., 2024	Regarding future research directions and academic work opportunities, this research suggests several areas of focus that could further enrich the understanding of dropout at the postgraduate level. It becomes evident that there is a need for longitudinal research that analyzes student life throughout their entire academic journey, which would allow the identification of changes in the intensity with which variables influence dropout over time

Review Reference	Further Research as Stated in the Review
van der Wee et al., 2024	Further research could enrich and deepen the ndings of the current work by empirically examining when, how and why certain characteristics of Living Labs, as identified in this study, could support transformative teaching and learning toward sustainability and empower higher education students to nd pathways toward a sustainable future…Hence, future studies should empirically examine characteristics of and experiences with education in Living Labs by focusing on the interplay between the Living Lab as a hybrid learning space, the approaches for teaching and learning used therein, the practices that foster collaboration and co-learning with societal stakeholders and the policies and structures of the higher education institute in which the Living Lab is embedded
Veres et al., 2025	We aim to focus our research efforts on refining sustainability indicators, exploring the role of digital platforms in fostering collaboration and examining the long- term impacts of interdisciplinary education on sustainability outcomes
Vinueza-Morales et al., 2025	Longitudinal studies on programming education outcomes: more research is needed to understand how different teaching methodologies influence students' career progression and long-term skills development. future research should explore how integrating programming with other fields (e.g., cognitive science, psychology, and educational technology) can enhance learning outcomes. Future studies can address the challenges and research gaps outlined in this discussion and contribute to the development of more inclusive, e□ective, and technologically enhanced programming education models. These insights will be invaluable for educators, curriculum designers, and policymakers seeking to optimize programming instruction for the next generation of learners
Vuoriainen et al., 2024	Further research can help by examining individual factors and combinations of factors. Since we did not limit our study to just one or two continents or countries, future research could further explore global variations in collaboration and potentially apply the proposed framework within different contexts
Wang & Abdullah, 2024	This study is expected to provide researchers and educators worldwide with a broad horizon and a clear reference point for further research on the develop-ment of students' CT, especially in China. ..further investigation into enhancing students' CT through intervention strategies is warranted
Wang & Ishak, 2025	Future research could further explore the longitudinal effects of SEL interventions to assess long-term outcomes, such as students' career readiness and psychological well-being. Studies should also investigate how SEL frameworks can be tailored to specific cultural contexts, ensuring the development of inclusive, adaptable SEL programmes that resonate across diverse student populations
You et al., 2024	Not identified
Zhao & Selvaratnam, 2024	Our research will help to provide future scholars with basic and more comprehensive theoretical references, to provide a theoretical overview of relevant studies on reform in this field, and to contribute to the reform and development of vocational education in China

APPENDIX 4

Raw Data on the Journal Rankings, Numbers, Populations and Settings of Primary Research

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Abdallah et al., 2025	Social Sciences & Humanities Open	NA	Social Sciences (miscellaneous)	39	NA	HEI	international
Medina et al., 2024	International Journal of Learning, Teaching and Educational Research	Q2	NA	50	NA	HEI	international
Ahmad & Khurizan, 2024	Sage Open	NA	General Social Sciences	75	NA	HEI	international
Alonso et al., 2025	Cogent Education	Q2	NA	15	Students	HEI	international
Alotaibi, 2025	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development	60	NA	HEI	international
Alvado, 2025	Frontiers in Education	Q2	NA	28	Students, Teachers	HEI	international
Martinez et al., 2025	Knowledge Management & E-Learning	Q2	NA	51	Students, teachers	HEI	international
Amavasi & Zimmerman, 2024	Nurse Education Today	Q1	NA	15	Student nurses	Nurse education institution	international
Andrade-Girón et al., 2025	Iberoamerican Journal of Science Measurement and Communication	NA	Social Sciences (miscellaneous)	28	Students	HEI	global, mainly Asian
Balalee, 2025	Discover Education	Q4	NA	22	Students	HEI	international
Banarjee et al., 2024	Cogent Education	Q2	NA	19	NA	HEI	national
Bannigan et al., 2025	Frontiers in Education	Q2	NA	72	Students	HEI	international
Barikzai et al., 2024	Cogent Education	Q2	NA	84	NA	University in an emerging economy	international
Barua & Lockee, 2024	Discover Education	Q4	NA	Not identified	Not identified	HEI	international

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Basheer et al., 2024	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development	83	Not identified	HEI	international
Bayly-Castaneda et al., 2024	Frontiers in Education	Q2	NA	78	Students, graduates	HEI	international
Benson et aL., 2024	Australasian Journal of Educational Technology	Q1	NA	15	NA	HEI	international
Bonilla-Jurado et al., 2024	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development	50	Students	HEI	international
Brauer et al., 2024	Review of Educational Research	Q1	NA	58	Teachers	HEI	international
Buele & Llerena-Aguirre, 2025	Frontiers in Education	Q2	NA	9	Teachers, faculty members	HEI	international
Burton et al., 2024	Nurse Education Today	Q1	NA	14	Tertiary health profession related students, Tertiary health		
course teachers/ professors/faculty	Nurse education institution	international					
Bustamante-Mona et al., 2025	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development	51	STEM students	STEM departments of universities	international
Cabrera-Duffaut et al., 2024	Frontiers in Education	Q2	NA	27	Students	HEI	international
Cadiz, 2024	Jurnal Ilmiah Ilmu Terapan Universitas Jambi	Q3	NA	61	NA	HEI	national
Castillo et al., 20244	International Journal of Learning, Teaching and Educational Research	Q2	NA	42	Students	HEI	international
Castillo-Martinez et al., 2024	Frontiers in Education	Q2	NA	85	HE stakeholders	HEI	international
Chashmyazdan et al., 2024	Journal of						

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Higher Education Policy							
And							
Leadership Studies	Q3	NA	48	Not identified	HEI	international	
Chen et al., 2024	Journal of Curriculum and Teaching	Q3	NA	30	Lecturers	HEI	international
Chugh et al., 2025	Computer Applications in Engineering Education	Q1	NA	26	NA	HEI	international
Coleman et al., 2025	Educational Review	Q1	NA	14	migrant, female academics from		
minority ethnic backgrounds	HEI	national					
Cui & Alias, 2024	Journal of Infrastructure, Policy and Development	NA	Social Sciences (miscellaneous)	16	NA	HEI	international
Din Eak & Annamalai, 2024	Asian Association of Open Universities Journal	Q1	NA	25	Students	HEI	International
Dukes III et al., 2024	Disabilities	NA	Social Sciences (miscellaneous)	1,479	Students	HEI	International
Edvardsen Tonheim et al., 2024	Medical education online	Q1	NA	39	Students, Teachers	HEI	international
El Aatik et al., 2024	The Open Public Health Journal	NA	Social Sciences (Health)	11	Students	HEI	national
Eltaiba et al., 2025	Global Transitions	NA	Social Sciences (Health)	85	Students	HEI	regional
Esteban, 2025	Cogent Education	Q2	NA	149	Students	HEI	international
Fengye et al., 2025	Cogent Education	Q2	NA	553	Medical students	Medical Inoversity	national, international
Fernandez-Batanero et ak., 2024	Interactive Technology and Smart						
Education	Q1	NA	11	NA	HEI	international	
Ferreira Santos, 2024	Higher Education Quarterly	Q1	NA	45	NA	HEI	international

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Frez-Pulgar et al., 2025	Journal of Bioethical Inquiry	NA	Arts & Humanities (Philosophy); Socail Sciences (Health)	444	NA	HEI	international
Galdames-Calderón et al., 2024	Education Sciences	Q1	NA	20	NA	HEI	international
Ghazian & Lortie, 2024	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development)	940	NA	HEI	international
Gomez et al., 2025	Journal of Intelligence	Q1	NA	83	NA	HEI	international
Goncalves et al., 2024	SAGE Open	NA	General Social Sciences	125	Students	HEI	international
Gudoniene et al., 2025	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development)	27	HEI teachers	HEI	international
Harmon et al., 2024	Nurse Education Today	Q1	NA	13	HE academics	HEI	international
Heintalu et al., 2025	Educational Research Review	Q1	NA	198 and 48	NA	HEI	international
Henry et al., 2024	Simulation & Gaming	NA	General Social Sciences	46	Students	HEI	international
Hidayat et al., 2024	International Electronic Journal of Mathematics Education	Q4	NA	20	Gen Z students of mathematics	HEI	international
Ishmuradova et al., 2024	Contemporary Educational Technology	Q1	NA	107	NA	HEI	international
Isiaku et al., 2024	Quality Education for All	Q4	NA	29	NA	HEI	international
Jaxin et al., 2024	Frontiers in Education	Q2	NA	24	Not identified	HEI	international
Junaštíkova, 2024	Interactive Technology and Smart						
Education	Q1	NA	36	Students	HEI	international	
Kalim et al., 2025	Smart Learning Environments	Q1	NA	17	Female students	HEI	regional

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Kalocsányiová et al., 2024	Educational Review	Q1	NA	44	Students	HEI	regional
Karimi & Khawaja, 2025	International Journal of Learning, Teaching and Educational Research	Q2	NA	22	HEI teachers	HEI	international
Kaymakcioglu & Thomas, 2024	Social Sciences & Humanities Open	NA	Social Sciences (miscellaneous)	37	Researchers	HEI	national
Klimova & Chen, 2024	Language Teaching Research Quarterly	Q3	NA	11	Students	HEI	international
Kovari, 2025	Social Sciences & Humanities Open	NA	Social Sciences (miscellaneous)	27	NA	HEI	international
Liang et al., 2025	Journal of Curriculum and Teaching	Q3	NA	24	Students	HEI	international
Lim & Lee, 2024	Journal of Educators Online	Q3	NA	30	NA	HEI	international
Liu et al., 2025	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development)	11	Students	HEI	national
Lopez-Chila et al., 2024	Education Sciences	Q1	NA	870	NA	HEI	international
López-Nuñez et al., 2024	Education Sciences	Q1	NA	47	Students, teachers	HEI	international
Lou & Zhou, 2024	Journal of Computer Assisted Learning	Q1	NA	15	NA	HEI	international
Ma, 2025	Computera and Education: Artificial Intelligence	Q1	NA	67	NA	HEI	international
Marano et al., 2024	Higher Education Quarterly	Q1	NA	21	Students	HEI	international
Mare & Mutezo, 2025	Interactive Laerning Environments	Q1	NA	21	Students	HEI	international
Marques et al., 2024	Educational Research	Q2	NA	44	Academics	HEI	international
Masalimova et al., 2024	Frontiers in Education	Q2	NA	971	NA	HEI	regional

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Mashingaidze & Mayayise, 2025	Cogent Education	Q2	NA	45	NA	HEI	international
Mireku & Bervell, 2024	Higher Education	Q1	NA	143	NA	HEI	regional
Mireku et al., 2024a	Research Ethics	Q1	NA	171	NA	HEI	regional
Mireku et al., 2024b	International Journal of Educational Development	Q1	NA	96	NA	HEI	regional
Mohammed & Ozdamli, 2024	Behavioral Sciences	NA	Social Sciences (Development)	69	IT students	HEI	international
Mursalzade et al., 2025	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development)	31	Students	HEI	international
Nalweyiso et al., 2025	Cogent Education	Q2	NA	8	Undergraduate healthcare students	HEI	regional, developing countries
Ncube & Ngulube, 2024	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development)	14	Postgraduate students	HEI	international
Ndibalema, 2025	Cogent Education	Q2	NA	14	Students	HEI	regional
Nieminen et al., 2024	Educational Research Review	Q1	NA	42	Students with disabilities	HEI	international
Ogunleye et al., 2024	Education Sciences	Q1	NA	355	NA	HEI	international
Olivares-De la fuente et al., 2025	Frontiers in Education	Q2	NA	27	NA	HEI	international
Ortega-Ruiperez & Correa-Gorospe, 2024	Frontiers in Education	Q2	NA	15	Students	HEI	international
Oulamine et al., 2025	Educational Process. International Journal	Q1	NA	77	NA	HEI	international
Pallaris et al., 2024	Interactive Technology and Smart						
Education	Q1	NA	183	NA	HEI	international	
Parambil et al., 2024	Computers and Education: Artificial Intelligene	Q1	NA	92	NA	HEI	international

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Paredes Canencio et al., 2024	Environment, Development and Sustainability	NA	Social Sciences (Geography, Planning and Development)	50	NA	HEI	international
Peláez-Sánchez et al., 2024	Frontiers in Education	Q2	NA	182	NA	HEI	international
Pelaez-Sanchez et al., 2024	Frontiers in Education	Q2	NA	46	NA	HEI	international
Punch et al., 2025	International Journal of Inclusive Education	Q1	NA	17	NA	HEI	national
Rahajeng et al., 2024	Disabilities	NA	Social Sciences (miscellaneous)	17	Students with disabilities	HEI	national
Raitskaya & Tikhonova, 2024	Journal of Language and Education	Q2	NA	44	NA	HEI	international
Raitskaya & Tihkonova, 2025	Journal of Language and Education	Q2	NA	30	Students	HEI	international
Renfors, 2024	International Jurnal of Sustainability in Higher Education	Q1	NA	22	NA	HEI	international
Rosales-Ricardo & Caceres-Manzano, 2024	Health Professions Education	Q4	NA	11	Students	HEI	international
Rosario & Raimundo, 2024	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development)	61	NA	Entrepreneurship HEI	international
Roy et al., 2025	Sustainable Futures	NA	Social Sciences (Sociology and Political Science)	105	NA	HEI	international
Saez-Zevallos & Montalvo-Apolín, 2025	Frontiers in Education	Q2	NA	143	NA	HEI	international
Sahar & Munawaroh, 2025	Discover Sustainability	NA	Social Sciences (Geography, Planning and Development)	275	NA	HEI	international
Schei et al., 2024	Education Sciences	Q1	NA	24	Students	HEI	international
Sergeeva et al., 2024	Contemporary Educational Technology	Q1	NA	67	NA	HEI	international
Setiamurti & Kurniawati, 2024	Open Education Studies	Q3	NA	28	NA	HEI	international

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Shahjahan & Seinn, 2025	Cogent Education	Q2	NA	47	NA	Engineering Faculties	international
Singun, 2025	Discover Education	Q4	NA	20	NA	HEI	international
Solis-Garcia et al., 2024	Behavioral Sciences	NA	Social Sciences (Development)	16	Students with disabilities	HEI	international
Solis-Garcia et al., 2025	Cogent Education	Q2	NA	31	Students with disabilities	HEI	regional
Stamou et al., 2024	Frontiers in Education	Q2	NA	25	Non-traditional students	HEI	international
Subki, 2025	International Journal of Sustainable Development and						
Planning	NA	Social Sciences (Geography, Planning and Development)	364	NA	Green campus	international	
Sukjairungwattana et al., 2024	Frontiers in Education	Q2	NA	13	NA	HEI	regional
Tang et al., 2024	Aila Review	NA	Social Sciences (Linguistics and Language)	23	NA	HEI	international
Tang et al., 2025	Forum for Linguistic Studies	Q4	NA	34	NA	HEI	national
Tareke et al., 2024	Education Sciences	Q1	NA	40	NA	HEI	national
Tareke et al., 2025	International Journal of Educational Research Open	Q1	NA	21	Students	HEI	international
Teng & Cosier, 2024	Frontiers in Education	Q2	NA	26	Students	HEI	international
Thiedig & Wegner, 2024	London Review of Education	Q2	NA	77	NA	HEI	international
Tikhonova & Raitskaya, 2024	Journal of Language and Education	Q2	NA	56	Researchers, academics, PhD students, students, HE teachers	HEI	international
Tillmanns et al., 2025	Trends in Higher Education	NA	NA	93	NA	HEI	international
Trujillo-Juárez et al., 2025	Discover Education	Q4	NA	18	HE teachers	HEI	international

Review Reference	Journal	Scopus ranking in Social Sciences (Education)	Scopus subject areas if not included in Social Sciences (Education)	Number of primary research	Populations of primary research	Type of HEI if applicable	Setting of primary research
Quecano et al., 2024	Cogent Education	Q2	NA	40	Dropped out postgraduate students	HEI	international
van der Wee et al., 2024	International Journal of Educational Development	Q1	NA	35	Students	HEI	international
Veres et al., 2025	Sustainability Switzerland	NA	Energy; Computer Science; Environmental Science; Social Sciences (Geography, Planning and Development)	286	NA	HEI	international
Vinueza-Morales et al., 2025	Frontiers in Education	Q2	NA	1,697	NA	HEI	international
Vuoriainen et al., 2024	European Journal of Engineering Education	Q1	NA	36	NA	HEI Engineering Faculty	international
Wang & Abdullah, 2024	SAGE Open	NA	General Social Sciences	15	Students	HEI	international
Wang & Ishak, 2025	International Journal of Learning, Teaching and Educational Research	Q2	NA	26	Students	HEI	international
You et al., 2024	Discover Sustainability	NA	Social Sciences (Geography, Planning and Development)	2,447	NA	HEI	international
Zhao & Selvaratnam, 2024	Cogent Education	Q2	NA	61	NA	HEI	national

https://doi.org/10.17323/jle.2025.27560

Assessing Academic and Disciplinary Literacies: Rubric Validation to Mjeasure Argumentation, Comparison and Source-Based Writing Skills

Roberto Arias-Hermoso, Eneritz Garro Larrañaga, Ainara Imaz Agirre

Mondragon Unibertsitatea, Spain

ABSTRACT

Introduction: Over the last two decades, research has suggested that academic and disciplinary literacies (ADLs) are key to integrating content and language, as language is used to express content knowledge, often through Cognitive Discourse Functions, discourse patterns that respond to cognitive actions in formal education contexts. Nevertheless, systematic assessment tools are required to assess student production of ADLs.

Purpose: This validation study focuses on developing an analytic rubric to measure three ADL skills, consisting of three dimensions. Two of these dimensions are CDFs: students' skills to *argue* and *compare.* A third rubric measures an additional academic skill, students' ability to write from sources. The rubrics are designed to capture cross-disciplinary and multilingual productions of these skills, therefore being applicable for various disciplines (history, science, mathematics, among others...) and languages.

Method: A five-step validation process based on expert judgement was used, involving 13 international experts. They quantitatively and qualitatively evaluated the proposed rubrics based on pertinence, conceptual clarity, coherence, and relevance across two iterative rounds. Quantitative descriptive statistics and agreement indices were used, in addition to thematic analysis of qualitative feedback.

Results: The quality of the rubric showed clear progression between validation rounds. In the first version, several issues were raised by the experts–most criteria showed weak agreement and low means. After revisions, there was a substantial improvement in the second version, with 83 % of the criteria reaching strong or acceptable agreement. Qualitative feedback highlighted the need for precise and multidimensional operationalisations in each ADL dimension. To illustrate the application of the rubrics, multilingual student samples are provided.

Conclusion: The rubrics offer the first steps towards systematising ADL assessment by combining qualitative and quantitative feedback from 13 experts, underscoring the importance of expert input in advancing assessment practices. The study has theoretical and practical contributions: it highlights the multidimensional nature of ADLs and provides an adaptable rubric that can be used across disciplines, languages and educational contexts. This study focused on the validation process; therefore, empirical use of the rubric is still required. Future research should apply the rubrics to large-scale corpora and complement the expert-based validation with psychometric approaches.

KEYWORDS

cognitive discourse functions; writing assessment; disciplinary literacies; source-based writing

Citation: Arias-Hermoso, R., Garro Larrañaga, E., & Imaz Agirre, A. (2025). Assessing academic and disciplinary literacies: Rubric validation to measure argumentation, comparison and source-based writing skills. *Journal of Language and Education, 11*(3), 60-75. https://doi.org/10.17323/jle.2025.27560

Correspondence:
Roberto Arias-Hermoso,
rarias@mondragon.edu

Received: July 03, 2025
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

Mastering academic language has been defined as a critical factor for educational success (Lorenzo, 2023), as it is through language that learners construct, communicate and are assessed on school-related knowledge (Schleppegrell, 2004). In fact, every aspect of the teaching and learning process entails some type of language: building new knowledge through classroom interaction (Nikula et al., 2024), answering exam prompts (Breeze & Dafouz, 2017), or teaching

materials (Lersundi, 2022). All of these require students to *do* something with language, shaping their thoughts and knowledge. It is, therefore, essential to understand that language processes cannot be isolated from content learning. Consequently, implementing and considering educational interventions that foster academic language skills in integration with content is necessary (Imaz Agirre et al., 2025).

The disciplinary dimension of academic literacies has been highlighted, since academic literacy is heavily influenced by "the shared cultural values and communicative repertoires within the disciplines" (Li, 2022: 2). In this regard, recent research on content and language integration learning (CLIL) and bilingual education has shifted its attention towards the disciplinary aspect of academic language, i.e., disciplinary or subject-specific literacies (Dalton-Puffer et al., 2024). Despite nuances in its definitions and operationalisations, disciplinary literacy has been defined as the set of conceptual and linguistic skills needed to construct and communicate subject-specific knowledge (Moje, 2015; Shanahan & Shanahan, 2008; Dalton-Puffer et al., 2024). Disciplinary literacies thus extend academic language skills (Shanahan & Shanahan, 2008) and provide a pedagogical recontextualization of professional expert domains for school purposes (Fang & Coatoam, 2013). In this article, we focus on Academic and Disciplinary Literacies (ADLs), which include both aspects related to more than one discipline, and, therefore, cross-disciplinary, as well as specific disciplinary practices.

As a way of connecting the discourse practices that emerge in multiple disciplines, but at the same time considering specific disciplinary aspects, Dalton-Puffer (2013) proposed the construct of Cognitive Discourse Functions (CDFs hereafter), which has been suggested as central to understanding ADLs, since it perfectly captures the interconnectedness between content, language and literacies (Morton, 2020). CDFs represent verbalisations of cognitive patterns common to educational contexts, such as *defining, exploring* or *explaining*. These patterns require specific discourse and lexicogrammatical schemata for their realisation (Dalton-Puffer & Bauer-Marschallinger 2019), and are grouped into seven major categories: categorise, define, describe, explain, explore, evaluate and report. The CDF construct has been gaining international attention since its conception, and several studies have used it to examine classroom interaction (Lersundi, 2022), written exam responses (Breeze & Dafouz, 2017) or pedagogical interventions (Rieder-Marschallinger, 2024). A central aspect of CDFs is their elicitation in CDF-tasks (Gerns, 2023a), in fact, learners are often asked to perform CDF-tasks, defined as processing subject contents and their verbalisation through corresponding language forms (Gerns, 2023a). Most disciplinary productive tasks elicit some kind of CDF, such as defining what the ozone layer is, evaluating how cheques and bills affected trade or classifying triangles according to their angles (all examples from Lorenzo et al., 2024).

CDFs, therefore, represent both the cognitive (content) and discourse (language) resources needed by students to perform certain ADL tasks. In these tasks, students often respond based on their previous knowledge of the content matter, normally acquired through school settings. Nevertheless, many disciplines have emphasised the importance of a crucial academic skill: using discipline-related sources to construct knowledge. For example, Sendur et al. (2021) or Steiss et al. (2024) mention that using and corroborating sources is crucial for historical literacy. Research on writing from sources, or source-based writing (SBW), has provided ample evidence of the types of tasks that can be performed after using sources, with argumentative and summary tasks being the most frequently used ones (Chan & May, 2022). These fall within the scope of CDFs, as *arguing* and *summarising* are indeed part of the CDF construct. Although the connection between CDFs and SBW has yet to be well established in the literature, both represent important academic literacy skills that are activated in educational contexts.

The usefulness of the CDF construct for research and teaching has been emphasised, and Lorenzo et al. (2024) claim that the construct offers great opportunities for ADL assessment. Nevertheless, while various approaches have been proposed to assess ADLs, existing tools may be limited and fail to capture the multidimensional nature of each CDF. In particular, there is a lack of validated analytical–not holistic–rubrics that are adaptable across languages and disciplines. This article, therefore, addresses this gap by reporting the validation process of an analytic rubric focusing on two specific CDFs (*argue* and *compare*) and the ability to write from sources. The following sections will summarise research on ADL assessment, as well as provide contextualisation on how these three skills have previously been operationalised for assessment purposes, therefore providing evidence of the most important components of the three dimensions of the rubric presented here.

LITERATURE REVIEW

Assessing Academic and Disciplinary Literacies

Despite the fact that ADL assessment still needs to be further explored and developed to create systematic assessment tools, previous research has used diverse approaches. These can be grouped into three broad categories: studies using component-based assessment, those using comparative judgement and those using rubrics or scales.

Component-based assessment of ADLs has focused on identifying certain features of ADLs and CDFs in students' production by, normally, employing conceptual maps or Systemic Functional Linguistics. For instance, the work by Connolly (2019) analysed secondary education students' scientific explanations. Her component-based assessment

model included conceptual sophistication, addressing reasoning and validity, and communicative sophistication, addressing global and local cohesion and subject specificity. Similarly, other studies have proposed conceptual frameworks to analyse specific CDFs, such as *defining* (Nashaat-Sobhy & Llinares, 2023), *arguing* (Garro Larrañaga et al., 2024), *reporting* (Roca de Larios et al., 2023), *comparing* (Gerns, 2023; Evnitskaya & Dalton-Puffer, 2023, 2024) or *evaluating* (Whittaker & McCabe, 2023), most of them based on Systemic Functional Linguistics (see Halliday & Matthiessen 2014). The studies, overall, suggest that conceptual frameworks and maps can be beneficial to understanding how learners perform in CDF-tasks (Gerns, 2023). Nevertheless, while valuable for research purposes, these frameworks can often be too complex for classroom use.

A second group of studies has used comparative judgement techniques (Lesterhuis et al., 2017) to explore the main assessment criteria highlighted by teachers when assessing learners' ADL production. Comparative judgement uses algorithms that, after a series of pair comparisons of texts by a number of judges, calculate scores based on the number of wins and losses. Llinares et al. (2024) found that when reflecting on comparative judgement procedures to assess ADLs, content and language teachers' language awareness was fostered, and it helped them articulate their thoughts on CLIL. Similarly, Nashaat-Sobhy and Morton (2024) and Morton (2022) identified the most important characteristics when teachers evaluated learner writing, and proposed that content quantity and quality, language function and form, and the integration of all of these played an important role in assessing ADL writing. Similarly, the study by Meneses et al. (2023) found that the comparative judgement scores given by the teachers were highly correlated with rubric and analytic scores. Although it provides promising opportunities, comparative judgement with teachers has yet to yield systematic descriptors for specific CDFs, following Llinares and Morton's (2024) work with the CDF *explore*, moving from teacher awareness to specific tools.

A final group of studies focusing on ADL has used rubrics, sometimes in combination with the other two strategies mentioned above, such as the studies by Gerns (2023a) and Connolly (2019) using component-based assessment and a rubric, or Meneses et al. (2023), using both comparative judgement and an analytical rubric. The usefulness of rubrics and scales has been highlighted in the literature, as they have frequently been used in educational research and/or language testing (Kuiken & Vedder, 2020). Rubrics, in fact, can provide a detailed account of a learner's performance levels in certain tasks, as they normally provide descriptors or indicators reflecting students' performance. In the field of CDF research, the use of rubrics can prove useful for both teachers and researchers to capture performance in students' writing (Granados & Lorenzo, 2024; Lorenzo et al., 2024). Different studies examining CDF productions have proposed specific rubrics for specific CDF-tasks, such as de-Boer (2020) for online interaction contexts, del Pozo and Llinares (2021) and del Pozo (2024) for history writing, or Bauer-Marschallinger (2022; Rieder-Marschallinger, 2024) for historical literacy. Although not necessarily from a CDF perspective, other rubrics have also been used to assess the skills under study in this article, i.e., rubrics to measure argumentation (e.g., Qin & Karabacak, 2010) or SBW skills (Uludag & McDonough, 2022).

Many of these rubrics, whether CDF-related or not, however, normally consider the skill as a whole. Lorenzo et al. (2024), for example, provide a list of CEFR-based descriptors for B1 and B2 levels for the 7 CDFs in history, mathematics and science. While the disciplinary differences capture important nuances of each subject, the descriptors measure each CDF as a whole unit holistically (e.g., *defining* at B1 in mathematics) and may fall short when observing specific aspects that can lead to nuanced interpretations. This is a critical gap in CDF assessment for research, as ADLs are inherently multilayered and multidimensional (Nikula et al., 2024; Dalton-Puffer et al., 2024). Therefore, reducing CDF assessment to a single descriptor may fall short on capturing nuances of student performance. Consequently, research has called (del Pozo, 2024) for the use of analytical scales that consider various aspects of the same dimension, which improve construct validity (Paltridge & Phakiti, 2015). In the following paragraphs, we define and operationalise how the three main dimensions under study (the CDF *argue*, CDF *compare* and source-based writing) have been assessed previously in the literature.

The CDF Argue

Arguing is a sub-form of the CDF category *evaluate*, which Dalton-Puffer (2013: 234) defined as "I tell you what my position is vis a vis X". Arguing goes beyond expressing one's position, as it focuses on defending that position using evidence (Dalton-Puffer, 2016; Toulmin, 1958). Argumentation could thus be understood as the ability to use data or evidence to support one's opinion (Jiménez-Aleixandre & Erduran, 2007), and has been highlighted as a central element of education overall (Pylonitis & Meyer, 2024), and both in science (Polias, 2016; Jiménez-Aleixandre & Erduran, 2007) and history education (Lorenzo, 2017; Coffin, 2009; Sendur et al., 2021).

According to Crossley et al. (2022), learner argumentation has traditionally been analysed using Toulmin's model (1958). The model includes basic and complementary elements: the claim (the writer's opinion), data supporting the claim, warrants justifying how the data supports the claim, or rebuttals, refuting exceptions in which the claim might not be true. Toulmin's model, however, has received certain criticism (see for example Liu & Stapleton, 2014), as it does not include counterarguments, which are an essential part of the argumentative structure. Because of that, adaptations and modifications based on Toulmin's model have

been proposed (Sampson & Clark, 2008; Qin & Karabacak, 2010; Liu & Stapleton, 2014; Crossley et al. 2022). Using Toulmin's model as a starting point, Qin and Karabacak (2010) provide a holistic rubric from 1-5 measuring a writer's ability as a whole. Other proposals (e.g., Stapleton and Wu, 2015; Allagui, 2019; Jackson, 2024) specified different levels for each of the argumentation elements in Toulmin's model. Similarly, Sendur et al. (2021) proposed a scale for CLIL History, including different levels for the claim and evidence. All in all, these studies point out that successful argumentation quality is characterised by a clear claim supported by empirical data, with explicit justifications and rebuttals. These rubrics, however, often focus on argumentation skills at the textual level in higher education, when students typically have a greater understanding of argumentative writing. In addition, they frequently rely on vague descriptors (e.g., 'the claim is clear' vs. 'very clear'), which lacks specificity and operational precision.

CDF Compare

Comparing falls within the scope of the CDF category *categorise*, defined as "I tell you how we can cut up the world according to certain ideas" (Dalton-Puffer, 2013: 234). Comparing is a crucial step to categorise – it is fundamental to establish how two elements are equal and/or different according to certain criteria to be organised into categories (Evnitskaya & Dalton-Puffer, 2023). Comparing, therefore, is much about identifying similarities and differences between two or more elements, e.g., how are carnivores and herbivores different from each other? (Gerns, 2023a, 2023b). Despite the importance of comparing to build new knowledge by creating (dis)connections between concepts (Chen & Zhou, 2022), not much research has focused on learners' comparison skills. Nevertheless, it has been proposed as a central function in science (Gerns, 2023), and important for history too (Lorenzo et al., 2024).

The identification of a good comparison, however, is not as straightforward as argumentation, since comparison-and-contrast has normally been addressed as a genre move in scientific articles (see for example Chen & Zhou, 2022). In the domain of ADLs, however, research on comparing has recently emerged (Gerns, 2023a, 2023b; Evnitskaya & Dalton-Puffer, 2023, 2024). Based on previous proposals (for example, Trimble 1985 for science), the studies suggest that three main elements constitute the structure of comparison: what is being compared (the comparative item), the criteria used to establish similarity/difference, and the textual structure of the comparison. To our knowledge, only Gerns (2023a) has proposed a rubric to assess students' comparison skills. The rubric consists of three quality criteria: completeness, precision and explicitness. According to her rubric, students' comparisons receive higher quality scores when, in general, they write more balanced and complete comparisons (including both similarities and differences in a balanced way), the criteria are explicit and well-justified, and the structure is parallel and clear. However, research on *compare* as a CDF has recently begun to provide operational instruments (Gerns, 2023a) and do often not focus on the quality of specific elements, i.e., comparative items or criteria.

Source-Based Writing

The final aspect of interest is source-based writing (SBW), which is defined as the ability to use input sources to produce written compositions (Cheong et al., 2021). Although the connection between SBW and CDFs has not been explored in depth, they can easily be linked. In fact, students are often asked to produce a CDF-based composition in a SBW task. According to Chan and May (2022), arguing and summarising (falling under *report*) would be the most frequent functions activated in SBW tasks. In SBW tasks, learners use diverse inputs to construct compositions. These inputs, or sources, could be multimodal resources, which has been highlighted as one of the pillars of ADLs (Nikula et al., 2024), as students are exposed to various semiotic modes in the learning process, such as texts, audiovisual materials or graphs, some of them subject-specific (a timeline in History, a formula in Science or Mathematics).

Different aspects of SBW have been studied in the literature, particularly the ability of students to 1) use information from the sources, and 2) their ability to transform the source using their own words (e.g., Arias-Hermoso et al., 2024; Kato, 2018; Plakans & Gebril, 2013; van Weijen et al., 2019). Rubrics that consider different aspects of SBW have been previously proposed in the literature. The previously mentioned scale by Sendur et al. (2021) for CLIL History, in addition to the two items focused on argumentation, also measured students' ability to evaluate the sources, corroborate information and use the sources to contextualise the historical period. The rubric proposed by Uludag and McDonough (2022), on the other hand, focused on content (addressing the argumentative prompt correctly), organisation and language use, with an additional specific item to assess *source use*. This item considered both students' comprehension and paraphrasing of the source, as well as citing (Harsh et al., 2024). While some rubrics for SBW exist, they tend to focus on content and organisation–general aspects of writing–and citation practices. These overlook other crucial aspects of SBW, such as transforming sources or integrating students' own knowledge alongside information from the sources. The latter has been shown a critical aspect of SBW quality, closely related to creativity (Arias-Hermoso et al., 2024; Cheong et al., 2021), but has not been systematically included in previous rubrics.

Despite the valuable contributions of previous studies assessing ADLs, as has been mentioned, existing tools that focus on CDF assessment remain limited. Consequently, previous research (e.g., del Pozo, 2024) has called for more systematic and multidimensional approaches that can cap-

ture nuanced patterns in ADL student productions. Furthermore, many of the existing rubrics or component-based frameworks are tailored to productions written in English, ignoring the multilingual nature of ADLs. Together, these limitations underscore the need for validated and multidimensional rubrics that can assess specific CDFs and other ADL-related skills across languages and disciplines. Therefore, this study has a main research objective (RO) and two sub-objectives:

RO1: To validate an analytic rubric to assess students written production of the CDF *argue*, the CDF *compare* and their source-based writing skills.
RO1.1: To refine the rubric based on ADL expert judgement, ensuring clarity and coherence.
RO1.2: To illustrate the rubrics' applicability across languages and disciplines.

METHOD

Context and Intended Corpus

The rubric was designed to be applied in a multilingual disciplinary source-based writing corpus to better understand students' ADL development. As has been said, the present paper focuses exclusively on the validation of the rubric through expert judgement. While it does not report empirical analyses of the corpus, it serves two purposes. First, it aims to guide the design of rubric components to ensure applicability across languages and disciplines, and second, it aims to provide illustrative examples from the corpus of how the final rubric can be used. The empirical findings related to students' development of ADLs using the rubrics presented here are reported in Arias-Hermoso et al. (2025a) and Arias-Hermoso et al. (2025b), with a focus on either the language of instruction or the three languages.

As background, the intended corpus consists of texts written by 535 students across secondary education (ages 12-16) in a Basque-immersion model[1]. In this model, students typically receive 26 hours a week of exposure to Basque (in language and content classes) and 3 hours to both Spanish and English. In these models, the majority of students have Basque and/or Spanish as their first language, with the other being a second language, and English being a third language. Due to extensive schooling and to their status as the languages of the context, students in secondary education tend to have a good command of both Basque and Spanish (around B2), while English proficiency is usually lower (around A2).

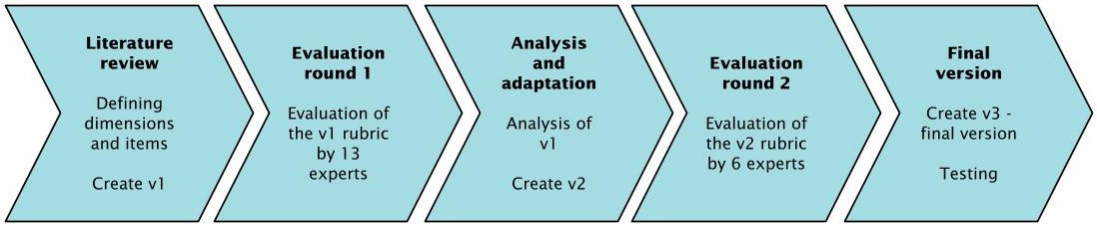
The intended corpus includes students' texts in their three languages of the curriculum (Basque, Spanish and English) and two disciplines (Science and History). The corpus is based on a SBW task in which students have to argue in favour of one of two subject-specific elements by comparing them, using information from the given sources. In the field of Science, the topic is nutrition, and learners have to compare the nutritional values and characteristics of two foods (chicken and chocolate), to establish which is healthier. In History, students are expected to compare the Middle Ages and today's society, using information about politics, economy, education and health. The compositions are written as argumentative blog entries for the school blog and were expected to be around 150 words. All participants and their legal guardians were informed about the nature of the research, and signed informed consent was collected before participation.

It is important to mention that, because of the nature of the task (cross-disciplinary and multilingual), the rubrics were created to be applied to different languages and disciplines. Furthermore, the rubric was planned to measure the three abilities in this particular type of task. Nevertheless, the three dimensions could also be used independently from each other, by, for instance, assessing source-based writing skills alone in a non-argumentative or non-comparative task.

The Validation Process

Various techniques have been used in educational research to address construct and content validity. For the purposes of this study, an expert judgment procedure was implemented. This procedure has been previously used in educational research (e.g., Lázaro Cantabrana & Gisbert Cervera, 2015) and entails the collection of experts' evaluations and feed-

Figure 1
Validation Process



[1] For a detailed description of the participants, the data collection procedures and empirical results, the reader is encouraged to read Arias-Hermoso et al. (2025a, 2025b), which present the systematic and empirical aspect of the developmental study.

back on the instruments, and then modifying them accordingly until quality standards are met (Landeta & Lertxundi, 2023). As has been mentioned, the main objective of the rubric of this study is to capture quality-related differences in three dimensions (argumentation skills, comparison skills and SBW) by considering multiple components of each. As can be seen in **Figure 1,** the validation process was carried out in five differentiated steps, which will be described in detail in this section.

Step 1: Literature Review and Creating V1

In the first step of the validation process, a literature review on each of the dimensions was conducted. Special focus was placed on identifying the key components or elements essential for successful performance within each dimension. In other words, we reviewed the literature focusing on answering the following questions: 'What are the essential features of *argumentation, comparison* and *writing from sources*? Which of these features signal better or more advanced performance?'. To answer these questions, we also examined already existing rubrics that measured the same dimensions (see Literature Review above). Due to the multidimensional nature of ADLs (Dalton-Puffer et al., 2024; Nikula et al., 2024), it is essential to consider how students perform at different components of the same dimension, as strongly highlighted by del Pozo (2024).

Following the suggestions by Vercellotti and McCormick (2021), the first step then consisted of identifying separate categories within each dimension. Therefore, after the literature review, the main components of each dimension were identified, and the first version of the rubric (V1) was created. The dimensions were called as follows: argumentation skills (D1), comparison skills (D2) and source-based writing (D3). Each dimension was then divided into items, representing the different elements or components expected for successful performance in said dimensions. Both D1 and D2 were composed of 4 items and D3 of 2. Then, the performance levels for each item (from 0, the lowest, to 3, the highest) were described based on the literature review.

Step 2: Expert Evaluation: Round 1

The second step consisted of creating the evaluation matrix and recruiting the experts for the evaluation process. An Excel sheet with the instructions and the evaluation matrix for each dimension was designed. Experts were instructed to rate each item from the rubric against 4 assessment criteria using a Likert scale with indicators from 1 to 5, as well as adding possible qualitative suggestions and comments for each of the items, or the dimension in general. The assessment criteria included the following four aspects:

(1) Pertinence (PE): the extent to which the item's formulation corresponds to what it measures.
(2) Conceptual clarity (CC): the item is accurately formulated and easy to comprehend, avoiding confusion or contradictions.
(3) Coherence (CO): the item has a logical connection with the dimension or indicator that it measures.
(4) Relevance (RE): the item is essential or important, i.e., it must be included.

In addition to an explanation of how to use the evaluation matrix, the document included a detailed description of the task that was going to be assessed with the scale. Subsequently, 15 international expert raters were contacted, out of which 13 accepted doing the evaluation process. All 13 experts held a PhD, had ample experience in language education and/or applied linguistics, and were familiar with the CDF construct and ADLs. In fact, most experts had previously published articles on CDFs and/or ADLs in peer-reviewed journals, and some also had extensive experience with either Science or History education. This decision was based on the importance of including both content and language experts' insights to improve the scale. They were given 2 months to complete the evaluation. All experts were informed about the nature of the validation process and written informed consent was received.

Step 3: Feedback Integration and Adaptation (V2)

After receiving the ratings from the expert judges, an analysis of their responses was conducted, both quantitatively and qualitatively. In a quantitative approach, descriptive statistics (means, standard deviations (SD), mode, minimum and maximum values) were calculated for each item and criterion. While Cronbach's alpha is suitable for within-dimension consistency, and ICC assume multiple targets (e.g., texts, for example), these are not suitable for this design. Other models (i.e., Rasch measurements) are often employed to validate rubrics, but they require multiple raters scoring multiple texts. In this case, raters did not *use* the rubrics but rather evaluated them (meta-assessment). Therefore, inter-rater agreement was calculated with dispersion measures: strong agreement was considered when the SD was below 0.5, acceptable between 0.51 and 0.8, and weak below 0.81. These measures were used in combination with the qualitative feedback to improve the rubric (see Results). Despite most items reaching an average score above 4 (M of all items and criteria = 4.32, SD = 0.51), especially for D1 and D3, some important issues with the items were raised in the qualitative feedback. The qualitative feedback analysed using thematic analysis, i.e., identifying common patterns in the comments. The common patterns were grouped for each item. This step was crucial to improve the rubric, and, accordingly, some important changes were made. Some items were removed, others were added, and some were divided into two different items. Most were reformulated following the feedback. All of the changes were summarised in a written report that compared V1 with the modified second version (V2).

Step 4: Expert Evaluation: Round 2

The report was sent back to the 13 raters, and they were asked to complete a second round of evaluation to assess whether the adaptations were sufficient. Out of the 13 original judges, 6 completed a second round of evaluation. The procedure was identical to the first round of evaluation: raters had to complete an evaluation matrix for each item assessing the 4 criteria. After this round, the feedback was considered to be sufficiently positive, with the vast majority of items reaching average values above 4.5, and all above 4. Some minimal changes were made after this round of suggestions.

Step 5: Feedback Integration and Adaptation (V3)

The final feedback was integrated and a final (V3) version was created. This version was piloted with 12 texts by the three authors until reaching a total agreement.

RESULTS

In this section, we present the results divided into two main sections: first, we focus on the development and validation process of the rubric by including a summary of the quantitative and qualitative feedback from the experts. Subsequently, we will highlight the main changes to each dimension. The second section will illustrate the final version of the rubric, supported by real student examples in English to facilitate reading the examples to an international audience, as well as instructions for research and practice.

Integration of Expert Judges' Evaluation

As can be seen in Table 1, there was a quantitative improvement from V1 to V2, particularly in D2, which presented the lowest scores. Quantitative results are based on the expert rating matrix, in which each item was judged on the four criteria using a 5-point Likert scale. In V1, following the SD threshold to calculate inter-rater agreement (i.e., SD ≤ 0.50 strong, 0.51-0.80 acceptable, > 0.81 weak), only 8 criteria (18.6 %) showed strong agreement, 9 (20.9%) acceptable, and 26 (60.5%) weak. In particular, items related to conceptual clarity tended to generate higher variability across judges (e.g., stance-taking [1.09]). After the revisions, not only did the averages improve, but also agreement: 36 criteria (76.6 %) reached strong agreement, 3 (6.4 %) acceptable and only 8 (17 %) weak. All items with weak agreement had a SD between 0.82-0.85, therefore being almost acceptable (see Supplementary Materials for complete descriptive and interrater agreement statistics[2]).

Nevertheless, the progression demonstrates that the iterative validation process, guided by expert feedback, significantly improved the rubric's clarity and coherence. Some general changes were made throughout all items of the rubric, such as the formulation patterns. V1 included different formulations for various items, such as 'the writer has presented' or 'the text presents', with occasional use of the

Table 1

Means for each Item in V1 and V2, Divided by the Assessment Criteria

Version 1 (13 raters)	PE	CC	CO	RE	Version 2 (6 raters)	PE	CL	CC	RE
D1. Argumentation skills	Sufficiency = 100 %				**D1. Argumentation skills**	Sufficiency = 100 %			
Stance-taking	4.69	4.23	4.62	5	Stance-taking	5	4.67	5	5
Evidencing	4.54	4.15	4.62	4.85	Reasoning	5	4.67	5	5
Opposition	4.69	4.31	4.46	5	Counterargumentation and rebuttal	5	5	5	5
Use of argumentative language	4.46	4.23	4.5	4.42	Language for arguing	5	4.5	5	5
D2. Comparison skills	Sufficiency = 90 %				**D2. Comparison skills**	Sufficiency = 100 %			
Comparison	4.45	3	4.55	4.82	*Comparative structure*	*4.67*	*4.83*	*5*	*5*
Comparees	3.85	2.85	3.85	3.54	Comparative items	4.83	4.67	4.83	5
Criterion	4.5	3.17	4.17	4.75	Comparative criteria	4.33	4.5	4.83	4.83
Use of comparative language	4.33	3.92	4.42	4.67	Language for comparing	5	4.6	5	5
D3. Source-based writing	Sufficiency = 70 %				**D3. Source-based writing**	Sufficiency = 100 %			
Use of source information	4.23	4.15	4.31	4.86	Use and comprehension of sources	5	4.60	5	5
Sourcing (degree of transformation)	4.54	3.69	4.62	5	Degree of transformation	5	4.40	5	5
					Use of original ideas	*5*	*4.60*	*5*	*5*

Note. Italics represent items that were added after the feedback, as they were not included in V1.

[2] Supplementary Materials available here: https://doi.org/10.6084/m9.figshare.30069613.v1

passive voice. In order to avoid misunderstanding and to ensure consistency, all items were reformulated to 'The writer presents/elaborates…'. An important issue present in D1 and D2 was raised by some of the judges – the fourth item within these dimensions referred to *the use of argumentative/ comparative language*. As has been said in the Introduction, it is true that language (form) cannot be separated from the production of discourse-functional patterns or elements related to the item. For example, stance-taking requires using certain lexicogrammar choices such as *I think* or *in my opinion* (Liu & Stapleton, 2014). Some judges suggested, therefore, merging this category into the corresponding item, instead of a single dimension. However, as also suggested by other judges, students might perform well at a functional level (be able to reason their claim with arguments) but fail to use accurate, appropriate or varied language forms, an issue also noted by Bauer-Marschallinger (2022), particularly in the context of additional (L2/L3) language learners. Therefore, we understand these two items as the lexicogrammatical coating of CDF production. Consequently, the first three items in D1 and D2 measure functional-structural aspects of CDFs, while the formal-linguistic aspects are measured by the fourth items.

Changes to D1: Argumentation Skills

The main changes carried out to D1 after the feedback are summarised here. V1 considered restating the claim as the highest value, but it was deemed unnecessary due to text length. In addition, inconsistent claims were considered part of 1, but they were moved to 0. The other three items' names were also changed: *evidencing* became *reasoning*, a common term used in both scientific and historical literacy research (e.g., Coffin, 2009). In this item, the previous formulation used 'evidence' to refer to both data and warrants, but following main research on argumentation (Toulmin, 1958; Crossley et al., 2022; Liu & Stapleton, 2014; Qin & Karabacak, 2010), the term 'evidence' was erased, and only data and warrants were included. *Opposition* was considered too broad and not entirely consistent with what the item intended to measure, and thus it was renamed to *counterargumentation and rebuttal*, following research on argumentation (Toulmin, 1958; Crossley et al., 2022; Liu & Stapleton, 2014; Qin & Karabacak, 2010). V1 did not establish very clear differences between levels 2 and 3, and it was modified to include a more explicit use of rebuttals. *Use of argumentative language* was changed to *language for arguing*, so as to include a wide array of lexicogrammatical resources used to produce arguments, including morphosyntactic, lexical, grammatical and textual level resources.

Changes to D2: Comparison Skills

This dimension experienced the greatest changes from V1 to the final version. The initial version included an item called *comparison*, which purposefully included the comparative items and the criteria, but a split version was recommended by all of the judges. Therefore, this item was erased, and the items were considered separately: *comparees* was changed to *comparative items* (following Gerns, 2023; Chen & Zhou, 2022), and *criterion* to *comparative criteria*. One of the experts suggested that "equally elaborating on differences and similarities among the two compared items is a crucial aspect of comparison", therefore, this was added as the highest level of *comparative items*. V1 included the basis of classification (following Trimble, 1985) as the main comparative criterion, however, this was reformulated in the definition to 'basis of comparison' in order to better capture the item's focus (Gerns, 2023; Evnitskaya & Dalton-Puffer, 2023, 2024).

As suggested by some judges, an additional dimension was added to examine the structure of the comparisons, *comparative structure*, as using a parallel structure has been highlighted as a crucial move in comparing (Gerns, 2023; Smith, 2019). The structures include the point-by-point method (at sentence or intrasentential levels) or the block method (at the text/paragraph level). Finally, the fourth item (*language for comparing*) was reformulated and unified with the parallel item in D1.

Changes to D3: Source-based Writing (SBW)

Finally, some important changes were also made in this dimension. While most raters recommended the inclusion of a dimension focusing on referencing and citation, which is an important aspect of SBW (Pecorari, 2013; Crossley et al., 2021), students in the sample were not instructed to reference or cite the sources, as it is not a common practice in Secondary Education in the context of the study. Regarding major changes, the first item considered the 'use and comprehension of sources', and the importance of doing so accurately and relevantly has been highlighted. More importantly, the second item was divided into two, as the version in V1 considered the use of original ideas as part of the highest level, but as recommended by the majority of the experts, the degree of transformation and the use of original ideas were kept separate. The *degree of transformation* item, therefore, considered lexical and/or structural changes – paraphrasing information in a way relevant to the task (Uludag & McDonough, 2022; Cheong et al., 2021). The new item created, *use of original ideas*, focused on students' ability to integrate their previous knowledge accurately and relevantly to the task.

Using the Rubric: Instructions and Examples

This section presents the final version of the rubric (V3) divided by dimensions and provides examples from the learner corpus gathered for this research. As has been mentioned above, the raters recommended that some important terms, such as *accurate* or *relevant,* be defined to properly use the rubric. As including them in the rubric itself would make it very dense, the scale is accompanied by the following definitions:

(1) Accurate: producing correct language forms following the linguistic norms of the target language, i.e., no errors (Wolfe-Quintero et al., 1998).
(2) Appropriate: the correct use of a linguistic resource according to its function within a context (Nikula et al., 2016).
(3) Complex/simple: related to the sophistication (depth) of the linguistic structures (syntactic and lexical) employed by the writer (Lu, 2017).
(4) Varied: related to the size (diversity) of the linguistic structures (syntactic and lexical) employed by the student (Lu, 2017), i.e., lack of repetition of structures.
(5) Relevant: use of information or ideas that are connected to the main purpose of the text (Leki & Carson, 1997; Macagno, 2016).

Tables 3, 4 and 5 below provide the final version of the rubric for each dimension[3]. To illustrate the rubric performance categories and their pedagogical relevance, three student examples are presented in Table 2. These excerpts are included for illustrative purposes only, as the systematic empirical analysis of the full corpus is reported elsewhere (see Arias-Hermoso et al., 2025a, 2025b). Two examples come from science essays (Ex1 and Ex3) and one from history (Ex2). Ex1 and Ex2 show texts from Year 7 students (12 years old), and Ex3 from a Year 10 student (16 years old).

Focusing on the argumentation dimension, we can observe important differences between the three texts above. While all three texts clearly present the writer's opinion explicitly (3 in D1_ST), their reasoning is very different. Ex1 uses some data to support their claim, but does not elaborate or link to the stance directly (level 1 D1_R), while the other two texts present more complete reasoning structures, with Ex3 including warrants (level 3 in D1_R), such as 'the chocolate cause adiction and this produce obesity, if we don't control it'. None of the texts includes explicit rebuttals, however, they give potential relevant counterarguments (level 2 in D1_CR). Regarding the *language for arguing* item, there is a great difference between Ex1 and Ex2/3. The first example does not use explicit argumentation markers, apart from 'because' twice, whereas the other two texts include more varied and complex lexicogrammar, such as 'in my opinion', 'I don't think that…', 'first of all' or 'apart of their components'. However, Ex2 and Ex3 did not reach the maximum level of this item, as some of the lexicogrammatical resources were not used appropriately or accurately ('apart of', 'becouse').

Regarding comparison skills, we can also see great differences across texts. The first text barely compares the two concepts, while the second and third texts present more clear comparisons between the two concepts by mentioning some differences. However, they are not 'equally elaborated on', a requirement to achieve the highest level in the rubric. Regarding the comparative criteria, three different levels can be seen in the texts, with Ex1 not displaying a clear criterion for the comparisons, Ex2 making comparisons based on implicit criteria (aspects related to *education* and *health*), and Ex3

Table 2

Three Unaltered Examples from the Corpus

Ex 1 (012_E1)	Hello, today I'm compare chicken and chocolate with milk. And wich is healthier between the two foods. The chicken is more healthier because the chocolate with milk haves more calories, more fat and more sugar. The chicken it is used in many healthy diets, it helps the nervous system and the digestive system and it gives energy to the body and the chocolate with milk only it helps to control our cholesterol and it affect our state of mind and because then the chicken is much healthier.
Ex 2 (006_E1)	In my opinion, I prefer to live in todays society, because, now, the economy is based on the service sector, there are many service, and becouse we live more safe. I don't think that people will live better on the middle ages, because, the local work and life were promoted, as well, as the help among neighbours, and, the childrens had to work every day! That is the most sadest thing in the world. And another interesant thing is that only reach people had access to education!! And now, now it is necessary to have studies! Or you can't be nothin, on your life. In the middle ages, people was killed for common illnesses like, flu, diharrea, and more things like that, and now, that doesn't affect us! And common illnesses on our live. Now you live, with 85 years or 90 years old. We are living in a pandemic, an there are more and more and more and… more, new illnesses, live covid 19 and illnesses like that. It was so hard to leave in the midde ages, that wasn't many doctors, and you can die because anithing can kill you.
Ex 3 (098_E4)	I'm going to explain why 100 g of chicken are healthier than 100 g of milk chocolate. First of all, I'm going to compare their components. The milk chocolate has aproximated % 500 more calories than the chicken, the chocolate has 531 kcal and the other only 99,2 kcal. And refering to the total fats the chocolate has literally % 2900 more the proportion is 1 g / 29 g. Looking to other many factors the proportions stay up like in sugar and calcium. But in other factors, The chiekn has more like in proteins and salt. Apart of their components we can see what effects they have in our sistem. The chocolate help us to mantain stable the coresterol and increasing our memory. But the chicken help our nervous sistem and that is very important. One bad thing that have the chicken is that if we eat it raw We can intoxicate with it. The chocolate cause adiction and this produce obesity, if we don't control it. Is true that the chocolate don't envolve animals but yes persons that are exploted in their jobes.

[3] For the complete version of the rubric, access institutional repository: https://hdl.handle.net/20.500.11984/6860

Table 3

D1: Argumentation Skills Rubric, with 0-3 Indicators

D1: ARGUE					
ITEM	**DEFINITION**	**0 - Inexistent, poor**	**1 - Fair**	**2 - Good**	**3 - Excellent**
Stance-taking (D1_ST)	*This is defined as the writer's ability to express their claim or position towards the topic and take a stance.*	The writer's stance towards the topic is not stated, or it is not consistent throughout the text.	The writer's stance towards the topic is difficult to discern, as it is implied.	The writer's stance towards the topic is presented explicitly and is mostly clear.	The writer's stance towards the topic is presented explicitly, clearly and accurately.
Reasoning (D1_R)	*This is defined as the writer's ability to provide supporting data and warrants to justify their stance.*	The writer does not produce data to support their stance or the data is irrelevant, incorrect and/or contradictory.	The writer produces accurate data to support their stance, but it is not elaborated on or not directly linked to their stance.	The writer produces accurate and relevant data to support their stance, and it is elaborated on and/or explicitly linked to their stance.	The writer elaborates on accurate and relevant data to support their stance, including warrants and/or examples for justification purposes.
Counterargumentation and rebuttal (D1_CR)	*This is defined as the writer's ability to acknowledge and refute opposing views to their claim, by providing counterarguments and rebuttals based on data.*	The writer does not mention, acknowledge, or recognise opposing views.	The writer acknowledges that opposing views exist but does not produce counterarguments based on data.	The writer produces accurate and relevant counterarguments, but they are not rebutted or critiqued.	The writer produces accurate and relevant counterarguments, explicitly rebutting and/or critiquing opposing views.
Language for arguing (D1_LA)	*This is defined as the writer's ability to use appropriate, accurate and varied lexicogrammatical resources to produce argumentation elements, including connectives, words to express cause and effect, evaluative language, modalisation...*	The writer does not use lexicogrammatical resources to produce argumentation elements or uses them incorrectly.	The writer uses minimal lexicogrammatical resources to produce argumentation elements, and/or their use might not be completely accurate and/or appropriate.	The writer uses lexicogrammatical resources to produce argumentation elements appropriately and accurately, but they are mostly simple and not varied.	The writer appropriately and accurately uses a variety of simple and complex lexicogrammatical resources to produce argumentation elements.

explicitly stating the criteria (*components* and *effects in our system*). Concerning the comparative structure, Ex1 is somewhat logically structured at the sentence level, but is not completely clear at the paragraph or textual levels. Ex2 and Ex3, in contrast, present a clearer structure, with paragraphs related to a central element and using a point-by-point structure. The structure within the paragraphs, however, does not present comparisons systematically. Finally, students' use of *language for comparing* also differs to a great extent. Ex1 incorrectly uses the comparative form with adjectives (more healthier) once, but uses correct comparative noun forms, 'more calories'. Performing better, Ex2 and Ex3 use more diverse lexicogrammatical resources by including, for example, the words 'prefer', 'looking to other factors', or 'referring to'. They also present some errors, such as 'the most sadest', and, therefore, their use of comparative language is not completely accurate.

The final dimension of the rubric considered SBW skills. Ex1 includes some information from the source, the information is not well used, therefore, introducing certain irrelevant aspects. In contrast, Ex2 and Ex3 present a good amount of information relevant to the task and there are no inaccuracies, even accurately including details from the sources. Regarding the degree of transformation, Ex1 only made minimal linguistic changes, with a great deal of copying. Conversely, Ex 2 and Ex3 both use the information from the source but put it in their own words by introducing some lexical changes. In fact, for instance, the idea that 'The chocolate help us to mantain stable the coresterol and increasing our memory' was originally presented as two ideas: 'It helps us control our cholesterol' and 'it improves our memory'. The student in Ex3 was able to make variations in the wording and introduce the same ideas with their own words. Finally, the first text does not use any original idea not present in the source, while Ex2 and Ex3 do so to a certain extent. Ex2 includes the idea of doctors not being common in the Middle Ages, and Ex3 mentions people being exploited to create chocolate. These ideas are mostly well integrated with the information from the sources, therefore, the texts received a 2.

DISCUSSION

The main objective of the present article has been to report on the validation process of a rubric aimed at assessing three dimensions of ADLs through an expert judgment procedure. Unlike our other empirical studies (Arias-Hermoso et al., 2025a, 2025b) that apply the rubric to corpus data, this article has focused specifically on the validation process, highlighting how expert evaluations can refine and strengthen an educational assessment tool.

Table 4

D2: Comparison Skills Rubric, with 0-3 Indicators

D2: COMPARE					
ITEM	**DEFINITION**	**0 - Inexistent, poor**	**1 - Fair**	**2 - Good**	**3 - Excellent**
Comparative items *(D2_CI)*	*This is defined as the writer's ability to produce comparisons between the two concepts, by including differences and/or similarities, and elaborating on them.*	The writer does not compare the two concepts.	The writer partly compares the two concepts by mentioning some similarities and/or differences.	The writer clearly compares the two concepts by mentioning similarities and/or differences.	The writer clearly compares the two concepts by equally elaborating on several similarities and/or differences among the concepts.
Comparative criteria *(D2_CC)*	*This is defined as the writer's ability to explicitly express the comparative criteria, i.e., the bases of comparison (the criterion according to which the similarities/differences are being made).*	The writer does not produce a clear criterion for comparisons.	The writer produces comparisons based on a criterion, but it is not explicitly stated.	The writer explicitly states the criteria for comparisons.	The writer explicitly states and elaborates on relevant criteria for comparisons.
Comparative structure *(D2_CS)*	*This is defined as the writer's ability to organise the comparisons logically, at sentence, paragraph and text levels.*	The writer does not present structured comparisons throughout the text.	The writer presents comparisons that are somewhat logically structured.	The writer presents point-by-point comparisons that are logically and appropriately structured.	The writer presents point-by-point comparisons that are logically structured at sentence, paragraph and text levels.
Language for comparing *(D2_LC)*	*This is defined as the writer's ability to use appropriate, accurate and varied lexicogrammatical resources to produce comparisons, including compare/contrast connectives, lexical and grammatical forms to express difference/similarity, juxtapositions…*	The writer does not use lexicogrammatical resources to produce comparisons or uses them incorrectly.	The writer uses minimal lexicogrammatical resources to produce comparisons, and/or their use might not be completely accurate and/or appropriate.	The writer uses some lexicogrammatical resources to produce comparisons appropriately and accurately, but they are mostly simple and not varied.	The writer appropriately and accurately uses a variety of simple and complex lexicogrammatical resources to produce comparisons.

Table 5

D3: Source-Based Writing Skills Rubric, with 0-3 Indicators

D3: SBW					
ITEM	**DEFINITION**	**0 - Inexistent, poor**	**1 - Fair**	**2 - Good**	**3 - Excellent**
Use and comprehension of sources *(D3_UC)*	*This is defined as the writer's ability to convey their comprehension of the sources through the use of sources accurately in a relevant way.*	The writer's text presents indicators of serious comprehension problems or does not use the source at all.	The writer's text presents (little) information from sources somewhat relevant to the task, but there are major inaccuracies in conveying information from the sources.	The writer's text presents some information from sources relevant to the task, but there are minor inaccuracies in conveying details from the sources.	The writer's text presents substantial information from sources always relevant to the task, and details are accurately presented.
Degree of transformation *(D3_DT)*	*This is defined as the writer's ability to transform the information from the sources via paraphrasing (lexical and structural changes).*	The writer copies most information directly from the source and makes minimal or no linguistic changes.	The writer copies most information directly from the source, but there is some linguistic change.	The writer integrates information from the source by mostly paraphrasing with lexical or structural changes.	The writer accurately integrates information from the source by transforming it with both lexical and structural changes.
Inclusion of original ideas *(D3_O)*	*This is defined as the writer's ability to include ideas not present in the sources, (original ideas from the writer's background) by appropriately integrating external knowledge into the text.*	The writer does not include external information.	The writer includes some external information, but it is not appropriately integrated with source information.	The writer includes some external information in the text, and it is mostly well integrated with information from the sources.	The writer includes some external information in the text, and it is always appropriately integrated with information from the sources.

The first key finding is related to the validation process per se and how expert judgement proved crucial in developing a validated and multi-dimensional rubric. With the input from 13 leading scholars in the field of ADLs, the iterative process made it possible to identify ambiguous descriptors, particularly in clarity, and to reformulate them into more operational items. As highlighted by previous research (Lázaro Cantabrana & Gisbert Cervera, 2015; Uludag & McDonough, 2022), the design of assessment tools guided by experts' views provides a systematic foundation to create reliable and sound rubrics, a crucial aspect to ensure quality (Landeta & Lertxundi, 2023). Just as comparative judgement studies have been successful in drawing teachers' and researchers' insights to refine ADL-assessment (e.g., Llinares et al., 2024), the present study contributes to the growing body of research demonstrating the usefulness of systematic expert involvement for rubric development. This can be seen in the expert consensus across items, which increased from 18.6 % agreement in V1 to 76.6 % in V2, which represent a crucial improvement in rubric quality and construct validity.

A second important finding concerns the nature of *quality* in each of the three dimensions. As ADL-related skills are often complex and multi-componential (del Pozo, 2024; Nikula et al., 2024), it is crucial that tools include multiple aspects or components that clearly operationalise what constitutes a *great* argument, comparison or use of sources. The inclusion of multiple items underscores this multidimensional nature and improves construct validity (Paltridge & Phakiti, 2015). In particular, experts' consensus converged on the need for complete descriptors with various components. For instance, experts highlighted that high-quality performance in argumentation involves not only presenting clear claims but also justifying evidence and presenting rebuttals. This resonates with previous research on both rubric-based and component-based assessment of ADLs, regarding the integration of complete argument structures (Jiménez-Aleixandre & Erduran, 2007; Polias, 2016; Sendur et al., 2021), a structured account of comparative items and criteria (Gerns, 2023a; Chen & Zhou, 2022), and the role of source use and integration (Chau et al., 2022; Cheong et al., 2021). The inclusion of different items per dimension helps track student performance at various levels or ages to capture more nuanced patterns (see Arias-Hermoso et al., 2025b). In fact, as shown by empirical research on ADL development, different items within the same dimension may develop differently, for example, *reasoning* over *counterargumentation* (Arias-Hermoso et al., 2025a). The rubric, supported by the 13 experts' voices, therefore, emphasises the relevance of multiple components or items to address ADLs, and helps to define both conceptually and operationally the three key dimensions under study.

CONCLUSION

The present study aimed to validate a multidimensional rubric to assess three key dimensions of ADLs, i.e., argumentation skills, comparison skills and source-based writing skills. Through an expert-judgement procedure that engaged 13 scholars in two rounds, we demonstrated that expert feedback is crucial to refine ADL descriptors. The rubrics advance work on previous descriptors or rubrics by confirming and emphasising the multidimensional nature of ADLs, as well as by incorporating SBW into the construct. Methodologically, the study shows that expert-based feedback can produce rubrics that are both theoretically grounded and usable across languages and disciplines. Our study, therefore, combining research from different perspectives and disciplines, provides a tool that can foster the integration of content and language for assessment and teaching purposes.

The study, therefore, has important implications for both theory and practice. At a theoretical level, the rubric confirms the multidimensional nature of ADLs and CDFs. In addition to that, to our knowledge, no study has yet focused on SBW within ADLs, and this paper offers interesting points for their connection that still require further research. At a more practical level, this study offers great opportunities for research and classroom practices. The rubrics validated here are of great benefit for researchers interested in CDFs, ADLs and SBW, and could be used and adapted to different linguistic, educational and disciplinary contexts. The rubrics can be applied in various educational levels and contexts, and can be of great benefit for primary, secondary and higher education teachers, as well as for material designers, curriculum developers and other researchers in academic writing. Furthermore, the rubrics and the accompanying illustrative examples could be beneficial for language and content teachers for classroom instruction and assessment. Based on the rubrics, for example, content (e.g., History) teachers could design lessons focusing on comparing two historical phenomena, instructing them on how to highlight differences and similarities.

Despite the implications, some limitations to this research have to be addressed. One limitation is that it did not include empirical analyses of student performance, as the focus of the paper was on validating the rubric through expert judgement. Another limitation is related to the lack of psychometric modelling (for example, many-facet Rasch measurements) beyond descriptive statistics, which future research could combine with expert judges' qualitative input. An aspect that is both a contribution and a limitation is that the rubrics are not language- and discipline-specific. While this allows for their use in multiple languages and disciplines, they still may not fully capture subject-specific characteristics of CDF production, since what constitutes defining in science might be different from what is expected in history. CDF-based assessments could be adapted to the needs of specific disciplines. This is in line with previously mentioned CEFR-style assessment grids based on CDFs for mathematics, science and history. These grids, however, do not consider various elements within each CDF. Building on this, future studies could bring these perspectives together by building

subject-specific rubrics that consider multiple components for each CDF on top of the two presented in this study. Ideally, future research could try to create similar rubrics for the 7 CDF categories for multiple disciplines, which would also highlight the similarities and differences across disciplines, therefore proving the cross-disciplinary nature of the CDF construct. In addition, we believe that this process could also benefit from not only expert academics' judgement, but also from classroom-based teachers' ratings to create these rubrics.

ACKNOWLEDGEMENTS

We would like to express our gratitude to the 13 experts who generously contributed to the validation of this rubric, for their time, effort and expertise. Thank you for your support and collaboration.

This research received funding from the Basque Government under Grants [IT1664-22 & PRE_2021_1_0001] and by the Spanish Ministry of Science, Innovation and Universities [PID2019-111655RA-I00].

ETHICS STATEMENT

This research received the approval of the Ethics Committee of Mondragon University (ID: IEB20231218). Expert participants (the 13 judges) were informed about the nature of the task and informed consent was given to participate. Student participants from the intended corpus were also informed about the nature of the research, and their parents and/or guardians signed an informed consent for their participation. Data was treated solely for research purposes and anonymity was ensured during the whole process.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTIONS

Roberto Arias-Hermoso: conceptualisation; methodology; validation; analysis; writing (original draft); funding acquisition.

Eneritz Garro Larrañaga: conceptualisation; methodology; validation; analysis; reviewing and editing; funding acquisition.

Ainara Imaz Agirre: conceptualisation; methodology; validation; analysis; writing; funding acquisition.

REFERENCES

Allagui, B. (2019). Investigating the quality of argument structure in first-year university writing. In S. Hidri (Ed.), *English language teaching research in the Middle East and North Africa* (pp. 179–199). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-98533-6_9

Arias-Hermoso, R., Imaz Agirre, A., & Garro Larrañaga, E. (2024). A comparison between input modalities and languages in source-based multilingual argumentative writing. *Assessing Writing, 60,* 100813. https://doi.org/10.1016/j.asw.2024.100813

Arias-Hermoso, R., Garro Larrañaga, E., Imaz Agirre, A., & Dalton-Puffer, C. (2025a). Producing cognitive discourse functions in disciplinary Basque writing: developmental patterns across secondary education and L1 profiles. *International Journal of Bilingual Education and Bilingualism, 28*(7), 744–763. https://doi.org/10.1080/13670050.2025.2488736

Arias-Hermoso, R., Imaz Agirre, A., & Garro Larrañaga, E. (2025b). Multilingual disciplinary literacies: exploring developmental patterns of science writing across secondary education. *Journal of Multilingual and Multicultural Development,* 1–19. https://doi.org/10.1080/01434632.2025.2481200

Bauer-Marschallinger. (2022). *CLIL with a capital I: Using cognitive discourse functions to integrate content and language learning in CLIL history education* [Unpublished doctoral thesis]. University of Vienna.

Breeze, R., & Dafouz, E. (2017). Constructing complex Cognitive Discourse Functions in higher education: An exploratory study of exam answers in Spanish- and English-medium instruction settings. *System, 70*, 81-91. https://doi.org/10.1016/j.system.2017.09.024

Chan, S., & May, L. (2022). Towards more valid scoring criteria for integrated reading-writing and listening-writing summary tasks. *Language Testing, 40*(2), 410-439. https://doi.org/10.1177/02655322221135025

Chau, L.T., Leijten, M., Bernolet, S. & Vangehuchten, L. (2022). Envisioning multilingualism in source-based writing in L1, L2, and L3: The relation between source use and text quality. *Frontiers in Psychology, 13*, 01-20. https://doi.org/10.3389/fpsyg.2022.914125

Chen, M., & Zhou, H. (2022). Comparison-and-contrast in research articles of applied linguistics: A frame-based analysis. *Lingua, 276*, 103387. https://doi.org/10.1016/j.lingua.2022.103387

Cheong, C. M., Zhu, X. & Liao, X. (2018). Differences between the relationship of L1 learners' performance in integrated writing with both independent listening and independent reading cognitive skills. *Reading and Writing, 31,* 779–811. https://doi.org/10.1007/s11145-017-9811-8

Cheong, C. M., Zhu, X. & Xu, W. (2021). Source-based argumentation as a form of sustainable academic skill: An exploratory study comparing secondary school students' L1 and L2 writing. *Sustainability, 132*, 2869. https://doi.org/10.3390/su132212869

Coffin, C. (2009). *Historical discourse: The language of time, cause and evaluation*. Bloomsbury.

Connolly, T. (2019). *Die Förderung vertiefter Lernprozesse durch Sachfachliteralität: Eine vergleichende Studie zum expliziten Scaffolding kognitiver Diskursfunktionen im bilingualen Chemieunterricht am Beispiel des Erklärens* [] [Unpublished doctoral dissertation]. Johannes Gutenberg-Universität Mainz.

Crossley, S. A., Tian, Y., & Wan, Q. (2022). Argumentation features and essay quality: Exploring relationships and incidence counts. *Journal of Writing Research, 14*(1), 1–34. https://doi.org/10.17239/jowr-2022.14.01.01

Dalton-Puffer, C., & Bauer-Marschallinger, S. (2019). Cognitive discourse functions meet historical competences: Towards an integrated pedagogy in CLIL history education. *Journal of Immersion and Content-Based Language Education, 7*(1), 30–60. https://doi.org/10.1075/jicb.17017.dal

Dalton-Puffer, C., Bauer-Marschallinger, S., Brückl-Mackey, K., Hofmann, V., Hopf, J., Kröss, L., & Lechner, L. (2018). Cognitive discourse functions in Austrian CLIL lessons: Towards an empirical validation of the CDF construct. *European Journal of Applied Linguistics, 6*(1), 5–29. https://doi.org/10.1515/eujal-2017-0028

Dalton-Puffer, C., Hüttner, J., & Nikula, T. (2024). The conceptualisation of disciplinary literacies in CLIL. In J. Hüttner & C. Dalton-Puffer (Eds.), *Building disciplinary literacies in content and language integrated learning* (pp. 1-25). Routledge.

Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics, 1*(2)*,* 216–253. https://doi.org/10.1515/eujal-2013-0011

Dalton-Puffer, C. (2016). Cognitive discourse functions: Specifying an integrative interdisciplinary construct. In T. Nikula, E. Dafouz, P. Moore, & U. Smit. (Eds.), *Conceptualising Integration in CLIL and Multilingual Education* (pp. 29–54). Multilingual Matters.

deBoer, M. (2020). Teacher-based assessment of learner-led interactions in CLIL: The power of cognitive discourse functions. In M. deBoer & D. Leontjev (Eds.), *Assessment and learning in content and language integrated learning (CLIL) classrooms* (pp. 203–223). Springer. https://doi.org/10.1007/978-3-030-54128-6_10

del Pozo, E., & Llinares, A. (2021). Assessing students' learning of history content in Spanish CLIL programmes: A content and language integrated perspective, in C. Hemmi, & D. L. Banegas (Eds.), *International Perspectives on CLIL. International perspectives on English language teaching.* Palgrave MacMillan. https://doi.org/10.1007/978-3-030-70095-9_3

del Pozo, E. (2024). Assessment in CLIL: the pending subject in bilingual education? A case study. *Revista De Educación, 1*(403), 231–257. https://doi.org/10.4438/1988-592X-RE-2024-403-605

Evnitskaya, N., & Dalton-Puffer, C. (2024). CLIL learners' categorizations: Writing about history in English across three grade levels in Spanish bilingual schools. In J. Hüttner & C. Dalton-Puffer (Eds.), *Building disciplinary literacies in content and language integrated learning* (pp. 60-79). Routledge.

Evnitskaya, N., & Dalton-Pufrer, C. (2023). Cognitive discourse functions in CLIL classrooms: Eliciting and analysing students' oral categorizations in science and history. *International Journal of Bilingual Education and Bilingualism, 26*(3), 311–330. https://doi.org/10.1080/13670050.2020.1804824

Fang, Z. & Coatoam, S. (2013). Disciplinary literacy: what you want to know about it. *Journal of Adolescent and Adult Literacy, 56* (8), 627-632. https://doi.org/10.1002/JAAL.190

Gerns, P. (2023a). Building scientific knowledge in English: Integrating content, cognition and communication in secondary school CLIL biology. *Journal of Language and Education, 9*(3), 52–78. https://doi.org/10.17323/jle.2023.17569

Gerns, P. (2023b). Qualitative insights and a first evaluation tool for teaching with cognitive discourse function: "Comparing" in the CLIL science classroom. *Porta Linguarum, 40*, 161–179. https://doi.org/10.30827/portalin.vi40.26619

Granados, A., & Lorenzo, F. (2024). A functional description of disciplinary literacy in history: Applications of the Common European Framework of Reference for Languages to content and language integrated learning courses. In J. Hüttner & C. Dalton-Puffer (Eds.), *Building disciplinary literacies in content and language integrated learning* (pp. 83–100). Routledge.

Halliday, M. A. K., & Matthiessen, C. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). Routledge.

Harsch, C., Koval, V., Kanistra, P. V., & Delgado-Osorio, X. (2024). Validating an integrated reading-into-writing scale with trained university students. *Assessing Writing*, *62*, 100894. https://doi.org/10.1016/j.asw.2024.100894

Imaz Agirre, A., Arias-Hermoso, R., & Ipiña, N. (2025). The effect of an intervention focused on academic language on CAF measures in the multilingual writing of secondary students. *International Review of Applied Linguistics in Language Teaching, 63*(2), 1373–1396. https://doi.org/10.1515/iral-2023-0137

Jackson, D. O. (2024). The longitudinal development of argumentative writing in an English for academic purposes course in Japan. *System*, 103482. https://doi.org/10.1016/j.system.2024.103482

Jiménez-Aleixandre, M. P., & Erduran, S. (2007) Argumentation in science education: An overview. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education* (pp. 3–27). Springer.

Kato, M. (2018). Exploring the Transfer Relationship of Summarizing Skills in L1 and L2. *English Language Teaching Archives, 11*(10). https://doi.org/10.5539/elt.v11n10p75

Kuiken, F., & Vedder, I. (2020). Scoring approaches: Scales/rubrics. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (1st ed., pp. 10). Routledge. https://doi.org/10.4324/9781351034784

Landeta, J., & Lertxundi, A. (2024). Quality indicators for Delphi studies. *Futures & Foresight Science, 6*(1), e172. https://doi.org/10.1002/ffo2.172

Lázaro Cantabrana, J. L., & Gisbert Cervera, M. (2015). El desarrollo de la competencia digital docente a partir de una experiencia piloto de formación en alternancia en el Grado de Educación [Developing teachers' digital competence through a pilot alternating training experience in the Bachelor's degree in education]. *EDUCAR, 51*(2), 321-348. http://dx.doi.org/10.5565/rev/educar.725

Lersundi, A. (2022). *Analysis of subject-specific literacies in a multidisciplinary project in upper-secondary education. Case Study.* [Unpublished Doctoral thesis]. Mondragon Unibertsitatea. https://hdl.handle.net/20.500.11984/5964

Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2017). Comparative judgement as a promising alternative to score competences. In E. Cano & G. Ion (Eds.), *Innovative practices for higher education assessment and measurement* (pp. 119-138). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-5225-0531-0.ch007

Li, D. (2022). A review of academic literacy research development: From 2002 to 2019. *Asian-Pacific Journal of Second and Foreign Language Education, 7*(1), 1-22. https://doi.org/10.1186/s40862-022-00130-z

Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System, 45*, 117-128. https://doi.org/10.1016/j.system.2014.05.005

Llinares, A., Morton, T., & Whittaker, R. (2024). Fostering language awareness for integration through teacher-researcher collaboration in a Spanish bilingual education context. *Language Awareness*, 1–21. https://doi.org/10.1080/09658416.2024.2385766

Llinares, A., & Morton, T. (2024). Assessing CLIL students' expression of Explore across languages and school disciplines: An interdisciplinary approach, in J. Hüttner & C. Dalton-Puffer (Eds.), *Building Disciplinary literacies in content and language integrated learning* (pp. 101-118). Routledge. https://doi.org/10.4324/9781003403685

Lorenzo, F., Lorenzo, F., Cvikic, L., Llinares, A., De Boer, M., Adadan, E., Arias-Hermoso, R., Ćaleta, M., Demirkol Orak, S., Evnitskaya, N., Glasnović Gracin, D., Granados, A., Guzmán-Alcón, I., Kasprzak, M., Lehesvuori, S., Miloshevska, L., Özdemir, H., Piacentini, V., del Pozo, E., Roquet, H., Bagalová, D., and Ting, T. (2024). *Assessing disciplinary literacy with CEFR descriptors: History, Mathematics and Science. A paper by CLILNetLE Working Group 2*. V1. PHAIDRA repository, University of Vienna. https://doi.org/10.25365/phaidra.532_30

Lorenzo, F. (2017). Historical literacy in bilingual settings: Cognitive academic language in CLIL history narratives. *Linguistics and Education, 37*, 32-41. https://doi.org/10.1016/j.linged.2016.11.002

Lorenzo, F. (2023). Academic language as linguistic capital – A window to social justice: A commentary on "Midadolescents' language learning at school: Toward more just and scientifically rigorous practices in research and education." *Language Learning, 73*(1), 136-143. https://doi.org/10.1111/lang.12560

Meneses, A., Montenegro, M., Acevedo, D., Figueroa, J., & Hugo, E. (2023). Cross-disciplinary language changes in 4th graders as a predictor of the quality of written scientific explanation. *Journal of Writing Research, 15*(1), 105–132. https://doi.org/10.17239/jowr-2023.15.01.05

Moje, E. B. (2015). Doing and teaching disciplinary literacy with adolescent learners: A social and cultural enterprise. *Harvard Educational Review, 85*(2), 254–278. https://doi.org/10.17763/0017-8055.85.2.254

Morton, T., & Nashaat-Sobhy, n. (2024). exploring bases of achievement in content and language integrated assessment in a bilingual education program. *TESOL Quarterly, 58*(1), 5-31. https://doi.org/10.1002/tesq.3207

Morton, T. (2020). Cognitive discourse functions: A bridge between content, literacy and language for teaching and assessment in CLIL. *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural Education, 3*(1), 7-17. https://doi.org/10.5565/rev/clil.33

Morton, T. (2022). Using cognitive discourse functions and comparative judgement to build teachers' knowledge of content and language integration for assessment in a bilingual education program. *Journal of Immersion and Content-Based Language Education, 10*(2), 302-322. https://doi.org/10.1075/jicb.21017.mor

Nashaat-Sobhy, N., & Llinares, A. (2023). CLIL students' definitions of historical terms. *International Journal of Bilingual Education and Bilingualism, 9*, 1–14. https://doi.org/10.1080/13670050.2020.1798868

Nikula, T., Nashaat-Sobhy, N., Minardi, S., Gülle, T., Yalçın, S., Duman, S.K., Bozbıyık, M., Delibegović Džanić, N., Ellison, M., Gerns, P., Gómez, E., Hanušová, S., Kováčiková, E., Kääntä, L., Lin, A., Llinares, A.; Yi Lo, Y., Lopriore, L., Meyer, O., Morton,T., Neville, C., Rannu, K., Sağlamel, H., Sula, G., Sulkunen, S., Pavičić Takač, V., Tiermas, A., Ting, T., Tsagari, D., Yüksel, G., & Žmavc, J. (2024). *Towards an initial operationalisation of disciplinary literacies: A paper by CLILNetLE Working Group 1*. PHAIDRA repository, University of Vienna. https://hdl.handle.net/11353/10.2050621

Paltridge, B., & Phakiti, A. (2015). *Research Methods in Applied Linguistics: A practical resource*. Bloomsbury.

Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing, 22*(3), 217-230. https://doi.org/10.1016/j.jslw.2013.02.003

Polias, J. (2016). *Apprenticing students into science: doing, talking & writing scientifically*. Lexis Education.

Pylonitis, C., & Meyer, O. (2024). Arguing for global citizenship: Mapping deeper learning in the language-as-discipline classroom. In S. Greco & L. Cinganotto (Eds.), *Innovation in education for deeper learning* (pp. 55–78). INDIRE-IUL Press.

Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System, 38* (3), 444-456. https://doi.org/10.1016/j.system.2010.06.012

Rieder-Marschallinger, S. (2024). Cognitive discourse functions in history CLIL education: Insights from a design-based research study on conceptual links. In J. Hüttner & C. Dalton-Puffer (Eds.), *Building disciplinary literacies in content and language integrated learning* (pp. 195–215). Routledge.

Roca de Larios, J., Coyle, Y., & Garcia, V. (2023). The effects of using cognitive discourse functions to instruct 4th-year children on report writing in a CLIL science class. *Studies in Second Language Learning and Teaching, 12*(4), 597–622. https://doi.org/10.14746/ssllt.2022.12.4.4

Sampson, V., & Clark, D. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education, 92*(3), 447–472. https://doi.org/10.1002/sce.20276

Schleppegrell, M. (2004). *The language of schooling: a functional linguistic perspective*. Mahwah: Erlbaum.

Sendur, K. A., van Drie, J., van Boxtel, C., & Kan, K. J. (2020). Historical reasoning in an undergraduate CLIL course: Students' progression and the role of language proficiency. *International Journal of Bilingual Education and Bilingualism, 25*(6), 2058–2074. https://doi.org/10.1080/13670050.2020.1844136

Shanahan, T. & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40-59. https://psycnet.apa.org/doi/10.17763/haer.78.1.v62444321p602101

Smith, S. (2019). *Academic writing genres: Essays, reports & other genres*. Evident Press.

Stapleton, P., & Wu, Y. (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes, 17*, 12-23. https://doi.org/10.1016/j.jeap.2014.11.006

Steiss, J., Wang, J., Kim, S. G., & Olson, C. B. (2024). U.S. Secondary Students' Source-Based Argument Writing in History. *Written Communication*. https://doi.org/10.1177/07410883241263549

Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.

Trimble, L. (1985). *English for science and technology: A discourse approach*. Cambridge University Press.

Uludag, P., & McDonough, K. (2022). Validating a rubric for assessing integrated writing in an EAP context. *Assessing Writing, 52*, 100609. https://doi.org/10.1016/j.asw.2022.100609

van Weijen, D., Rijlaarsdam, G. & van den Bergh, H. (2019) Source use and argumentation behavior in L1 and L2 writing: A within-writer comparison. *Reading and Writing, 32,* 1635–1655. https://doi.org/10.1007/s11145-018-9842-9

Vercellotti, M. L., & McCormick, D. E. (2021). Constructing analytic rubrics for assessing open-ended tasks in the language classroom. *TESL-EJ, 24*(4).

Whittaker, R., & McCabe, A. (2023). Expressing evaluation across disciplines in primary and secondary CLIL writing: A longitudinal study. *International Journal of Bilingual Education and Bilingualism, 5*(2), 1–18. https://doi.org/10.1080/13670050.2020.1798869

https://doi.org/10.17323/jle.2025.28343

Stylistic Deviations in Linguistics Introductions: A Move-Step Analysis of Wordiness, Redundancy, and Communicative Impact

Tatiana Golechkova [1], Nadezhda Arupova [2], Elena Golubovskaya [3]

[1] New Economic School, Moscow, Russia
[2] MGIMO University, Moscow, Russia
[3] RUDN University, Moscow, Russia

ABSTRACT

Introduction: Stylistic deviations such as wordiness and redundancy undermine clarity and precision in academic writing. Their frequency and communicative impact, however, are likely to vary across disciplinary traditions. Earlier research has examined these phenomena in education-related corpora and revealed patterned distributions of redundancy in justificatory passages. By contrast, the ways in which such deviations manifest in linguistics research articles remain underexplored.

Purpose: To investigate the distribution, functional localization, and communicative impact of stylistic deviations in linguistics Introductions, with a particular focus on how wordiness and redundancy interact with rhetorical structure.

Method: A corpus of 40 linguistics Introductions (388 sentence-level fragments) was compiled and annotated. Each fragment was coded by rhetorical Move and Step according to the CARS model and further categorized by deviation class (wordiness or redundancy), subcategory, and communicative impact. The taxonomy of deviations previously validated in education corpora was applied in full. Statistical analysis included descriptive profiling of classes and subcategories, chi-square tests of association, and effect size estimation.

Results: Wordiness predominated across the corpus, accounting for 70.6% of all deviations, while redundancy accounted for 29.4%. Class balance was stable across Moves and Steps, but severity was functionally localized: high-impact deviations clustered in M3_S3, M3_S1, M2_S2, and M1_S3. Almost all high-impact cases were linked to wordiness, with syntactic complexity alone responsible for 54 of 61 instances. Redundancy, although frequent in structural and lexical repetition, rarely reached high severity.

Conclusion: These findings show that in linguistics Introductions in English the primary stylistic risk lies not in repetition but in syntactic overload at rhetorically dense points of the text. The results extend previous applications of the taxonomy by demonstrating a discipline-specific pattern of risk concentration. The study highlights the value of combining rhetorical segmentation with fine-grained stylistic annotation and suggests that pedagogical efforts should focus on reducing syntactic complexity in high-pressure rhetorical contexts. Limitations include the modest corpus size and the absence of cross-disciplinary comparison, which future research should address to refine understanding of stylistic risk across fields.

KEYWORDS

stylistic deviations, wordiness; redundancy; syntactic complexity; communicative impact; Move–Step analysis; rhetorical segmentation; academic writing pedagogy

Citation: Golechkova, T., & Arupova, N., & Golubovskaya, E. (2025). Stylistic deviations in linguistics introductions: A move-step analysis of wordiness, redundancy, and communicative impact. *Journal of Language and Education, 11*(3), 76-85. https://doi.org/10.17323/jle.2025.28343

Correspondence:
Tatiana Golechkova,
tgolechkova@nes.ru

Received: June 13, 2025
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

Inadequate management of rhetorical pressures in research article Introductions gives rise to two recurrent types of stylistic deviation. Wordiness reduces informational density by introducing low-yield formulae, excessive epistemic modification, and nominalizations that weaken verbal force (Tikhonova et al., 2024).

Redundancy repeats meanings already expressed through lexical tautology and structural duplication (Degen et al. 2020). Both phenomena increase processing costs for readers and undermine clarity. Their presence has been widely documented across disciplines and genres, including medicine, law, and applied linguistics (Salager-Meyer, 1994; Biber & Gray, 2010; Goonaratna, 2002a, 2002b; Williams & Bizup, 2017; Flowerdew & Forest, 2015; Leufkens, 2023; Kravtchenko & Demberg, 2022).

These stylistic risks are not evenly distributed across the rhetorical structure of Introductions. In Move 1, where authors establish territory, writers often resort to generalizations and formulaic openings. In Move 2, devoted to identifying a research gap, repetition and vague evaluative language are common (Farneste, 2015). In Move 3, where aims and contributions are presented, templated statements can echo earlier content rather than add new information. Such tendencies have been documented in discipline-specific corpora as well as in cross-disciplinary comparisons of Introduction structures (Samraj, 2002; Kanoksilapatham, 2005; Hyland, 2008; Gong & Barlow, 2022). In the Russian academic context, these problems are amplified by the coexistence of different publication cultures and the continuing prominence of Russian as a language of research dissemination. This situation complicates transfer to and from English and sustains local stylistic traditions that privilege elaboration over economy (Raitskaya & Tikhonova, 2019; Raitskaya & Tikhonova, 2020; Smirnova, Lillis, & Hultgren, 2021).

Within this broader literature, Tikhonova et al. (2025) provide a systematic account of how wordiness and redundancy distribute across the Move and Step structure of Russian language Introductions in education. The association between rhetorical function and deviation type was statistically robust, with a nontrivial share of high impact cases that plausibly distort the uptake of aims and arguments. The study operationalized deviation categories by adapting typologies of wordiness and redundancy previously systematized in scoping reviews on academic Russian, which supports conceptual transparency and reproducibility (Tikhonova & Mezentseva, 2024).

We adopt the methodology of Tikhonova et al. (2025) and apply it to English language Introductions in linguistics for four reasons that are both theoretical and practical. First, it aligns analysis with the CARS model, which is the dominant account of Introduction organization and which enables function sensitive evaluation rather than surface level counting (Swales, 1990; Swales, 2004). Second, the taxonomy distinguishes wordiness and redundancy into analytically tractable subtypes that are grounded in prior reviews and have face validity for Russian academic discourse. This yields observable tags for manual annotation that can be audited and reused across corpora. Third, the IMPACT scale adds a communicative lens to the analysis and allows editors and instructors to prioritize interventions where stylistic deviations most threaten interpretability. Fourth, the protocol closes a known gap, where detailed, rhetorically anchored, corpus-based diagnostics of Introduction writing remain scarce beyond single disciplines. Together these properties make the approach suitable for replication and extension on new data, while ensuring comparability with prior findings (Tikhonova et al., 2025; Tikhonova & Mezentseva, 2024).

At the same time, the original study acknowledged limits. Its corpus was monodisciplinary and exclusively Russian language, which constrains generalization beyond education and beyond one academic culture. The assignment of IMPACT levels relied on expert judgment and therefore calls for inter-rater validation and triangulation with independent indicators of reader difficulty. These constraints point to the need for replication on new material and for designs that separate structural rhetorical pressures from discipline specific habits and period effects. Our study responds to this need by applying the same annotation and inference procedures to a new author compiled corpus and by testing whether the previously observed localization of overload persists. We also reassess the relative burden of wordiness and redundancy under different rhetorical demands and re-estimate the distribution of IMPACT across functionally critical segments (Tikhonova et al., 2025).

The present study pursues three research questions. RQ1 asks how wordiness and redundancy are distributed across Moves and Steps in the new corpus and whether the earlier clusters recur in justifying significance, gap indication, and aim presentation. RQ2 asks how the overall frequency and the subcategory profile of wordiness and redundancy shift across rhetorical functions and whether the balance between them changes under different communicative pressures. RQ3 asks how IMPACT levels distribute across Moves and Steps and whether high impact deviations concentrate in the same segments as before. Our hypotheses are that stylistic overload remains structurally motivated by rhetorical function, that wordiness remains more frequent overall while redundancy accounts for a disproportionate share of high impact cases, and that the most vulnerable segments coincide with those identified previously. These hypotheses are directly anchored in the baseline study and in the wider literature on genre organization and concision (Swales, 1990; Hyland, 2005; Williams & Bizup, 2017; Biber & Gray, 2010; Salager-Meyer, 1994).

Methodologically, we retain the core elements of the protocol. We segment Introductions by Move and Step, annotate stylistic deviations using the validated typologies for wordiness and redundancy, assign IMPACT levels to each marked fragment, and test associations with chi square. This preserves comparability with the baseline study and allows us to attribute any differences to corpus composition rather than to measurement change. The design supports actionable implications for editorial practice and instruction since the IMPACT profiles identify where targeted revision is most

likely to yield large rhetorical gains. In this sense, our contribution is both empirical and translational. We offer an evidence based map of risk in English language Introductions and a function sensitive framework for intervention that can be integrated into local training and review practices (Tikhonova et al., 2025; Raitskaya & Tikhonova, 2020).

METHOD

Design

This study is a corpus based replication and extension of the protocol introduced by Tikhonova et al. (2025) for diagnosing stylistic overload in research article Introductions. We adopt their function sensitive approach that aligns annotation with the CARS Move–Step model and distinguishes two superordinate classes of deviation, wordiness and redundancy, complemented by a three level communicative impact scale. This design preserves comparability with the baseline study while allowing discipline specific interpretation on the present dataset.

Corpus and Inclusion Criteria

The dataset consists of Introductions from forty research articles in linguistics published in 2024–2025. The corpus includes Q1 Scopus English-medium publications on a range of linguistics sub-fields, including applied linguistics, computational linguistics, typology, phonology, linguistics research methods. Table 1 summarises the journals included in the corpus. The selection of papers was performed using a continuous sampling method based on the four criteria: 1. The paper presents empirical research; 2. There is a separate Introduction section; 3. It is published in English; 4. The Introduction is publicly available.

The unit of analysis is the Introduction section. The working file is a single spreadsheet that records fragment level annotations with the following fields: ID, TEXT, CATEGORY, TAG, IMPACT, and MOVE_STEP. Article level identity is encoded in ID as the substring before the fragment counter. No article level bibliographic metadata are included in the file. The dataset contains 388 fragments. The number of annotated fragments per article ranges from 3 to 23 with a median of 8.

Fragments are distributed across the nine admissible Move–Step labels used in the baseline protocol. The inventory and the interpretation adopted here are as follows.

M1_S1 establishing territory.
M1_S2 justifying significance or centrality.
M1_S3 reviewing the field or delimiting a research direction.
M2_S1 indicating a gap.
M2_S2 problematizing or raising a question.
M2_S3 contrast in approaches or findings.
M3_S1 stating aim or research tasks.
M3_S2 brief orientation to method or data.
M3_S3 signposting contribution or expected results.

Table 1

Journals Demographics

	title	issn	SJR	SJR Quartile	H index	country	publisher
1	Transactions of the Association for Computational Linguistics	2307387X	1,824	Q1	67	United States	MIT Press Journal
2	Journal of Memory and Language	0749596X, 10960821	1,802	Q1	177	United States	Academic Press Inc.
3	Linguistic Typology	14300532, 1613415X	0,854	Q1	45	Germany	De Gruyter Mouton
4	Laboratory Phonology	18686346, 18686354	0,755	Q1	17	Germany	Open Library of Humanities
5	System	23693762	2,205	Q1	115	United Kingdom	JMIR Publications Inc.
6	Research Methods in Applied Linguistics	27727661	2,452	Q1	14	United Kingdom	Elsevier B.V.
7	Linguistics	00243949, 1613396X	0,732	Q1	60	Germany	De Gruyter Mouton
8	Journal of World Languages	21698252, 21698260	0,729	Q1	11	Germany	De Gruyter Mouton

Operational Definitions and Coding Scheme

We follow the deviation taxonomy used by Tikhonova et al. (2025) and apply it at the fragment level. Two superordinate classes are distinguished.

The first class is Wordiness low yield material that depresses informational density. Subcategories present in the dataset are:

(1) WORDINESS_GENERAL formulaic and low information phrasing
(2) WORDINESS_HEDGING excessive or stacked epistemic modification
(3) WORDINESS_COMPLEXITY syntactic overcomplexity and overload
(4) NOMINALIZATION preference for derived nouns where a finite verb would increase clarity
(5) FORMULAIC_PHRASE stock openings or transitions that add no propositional content
(6) EMPTY_REFERENCE vague labels or references that fail to advance the argument.

The second class is Redundancy avoidable repetition of information already present. Subcategories present in the dataset are

(1) REDUNDANCY_LEXICAL repetition of words or near synonyms within a local span
(3) REDUNDANCY_STRUCTURE re stating earlier rhetorical content or re cycling a step
(3) APPOSITIVE_PHRASE appositive expansions that restate the head without adding content.

Each marked fragment also receives an IMPACT label that captures expert judgment of potential interference with readers' uptake of the Introduction. Levels are ordered as LOW, MEDIUM, and HIGH. The scale is identical in purpose and granularity to the one used by Tikhonova et al. (2025).

Annotation Unit and Procedure

The annotation unit is a sentence level fragment. Each fragment is assigned exactly one Move–Step label and one deviation category. When a fragment exhibits more than one problem coders assign the category judged to have the strongest communicative impact following the decision rule reported in the baseline study. The present spreadsheet represents a single pass of primary annotation. Parallel double coding is not available in the file and inter-rater reliability cannot be computed from the supplied data.

Quality Control

Before analysis we verified that all MOVE_STEP values belong to the nine admissible labels, that categorical fields use a consistent case, and that each fragment contains a non-empty TEXT. Records with malformed identifiers or missing labels were flagged and corrected when the intended value was unambiguous from context. The final analytic dataset includes 388 fragments and 40 distinct article identifiers.

Primary outcomes are the counts of deviations by Move and by Step and the counts by superordinate class and subcategory. Secondary outcomes are the distributions of IMPACT across Moves and Steps and across deviation classes. The corpus level composition is reported descriptively to characterize the dataset that underpins inferential tests.

Statistical Analysis

Analyses mirror the baseline inferential strategy. We construct contingency tables for Move by deviation class, Step by deviation class, and Step by IMPACT. We use Pearson's chi square tests to assess associations with alpha set at 0.05 and report Cramér's V as effect size. Where expected cell counts are low, we merge rare subcategories within their superordinate class or compute Monte Carlo *p*-values. Descriptive results are reported as counts and proportions with 95 percent confidence intervals. All computations are performed in Python using pandas for data handling and SciPy for inferential statistics. This pipeline reproduces the logic of the baseline study and ensures that any cross study differences can be attributed to corpus composition rather than to changes in measurement or inference.

Ethical Considerations

All materials used are publicly available scholarly texts. The study does not involve human subjects or personal data and does not require ethics approval. Quotations from the corpus used illustratively in the Results follow fair use norms for research and teaching.

Data Availability

The annotated spreadsheet used in this study is part of the project archive and can be provided to reviewers and readers upon reasonable request, subject to the rights of original publishers of the underlying articles.

RESULTS

This section first describes the corpus and the distribution of fragments across CARS Moves and Steps. It then answers RQ1 by examining how the two deviation classes are distributed across Moves and Steps. It answers RQ2 by quantifying the overall balance between wordiness and redundancy. It answers RQ3 by analyzing communicative impact labels and showing where high impact fragments concentrate. Finally, it drills down to subcategories to identify which linguistic

patterns carry most of the risk and it integrates these patterns with the rhetorical map.

Corpus and Rhetorical Segmentation

The overall size profile of the corpus sets the baseline for interpreting all subsequent findings. As shown in Table 1, the majority of material is concentrated in Move 1, with a substantial closing block in Move 3, while Move 2 is comparatively smaller. This distribution reflects a rhetorical preference for elaborating background and signalling contribution, rather than devoting equal space to the gap-establishing middle section. Step-level patterns refine this picture. Within Move 1, the field-orienting step (M1_S3) accounts for the largest share of that move, highlighting authors' tendency to extend the review of previous work. In Move 2, the dominance of M2_S2 indicates that authors allocate proportionally more space to gap statements than to adjacent contrastive or counterclaim steps. In Move 3, the three steps are more balanced, though the slight expansion of M3_S3 points to the rhetorical weight placed on contribution signalling.

These differences mean that raw counts of stylistic deviations will inevitably be higher in larger steps, particularly in M1_S3 and M3_S3. To prevent misinterpretation, the following results are reported both in absolute numbers and in proportions relative to step totals.

Deviation Classes at Corpus Level and by Step

The third block of results examines the distribution of communicative impact labels across rhetorical steps. The analysis distinguishes three levels of impact: LOW, MEDIUM, and HIGH. Counts and proportions are presented in Table 3, with the "high rate" column indicating the share of high-impact fragments within each step.

At the corpus level, the majority of annotated fragments fall into the LOW and MEDIUM categories (179 and 147 respectively), while HIGH impact accounts for 61 fragments. The global test of association between step and impact does not reach significance, $\chi^2(16) = 11.737$, $p = 0.762$, $V = 0.123$. This result means that overall impact distribution does not differ systematically by step once the size of each step is taken into account. Nevertheless, step-level proportions highlight local concentrations of high impact that are masked in the global test. The highest rates occur in M3_S3 (11 of 47, 23.40%), M3_S1 (8 of 41, 19.50%), M2_S2 (8 of 42, 19.05%), and M1_S3 (13 of 69, 18.84%). By contrast, steps such as M2_S1 (0 of 26, 0.00%) and M2_S3 (1 of 24, 4.17%) exhibit little or no high impact. These figures indicate that risk is not evenly distributed across the rhetorical map: segments associated with announcing contribution, formulating aims, and elaborating research problems show markedly higher vulnerability to high-impact deviations.

Wilson confidence intervals further illustrate this pattern. For M3_S3 the 95% CI for the high rate is 13.60–37.22%, while for M1_S3 it is 11.35–29.61%. These intervals confirm that in both steps, high-impact fragments represent a non-trivial portion of the text, with values well above those observed in smaller steps such as M2_S1. This descriptive profile establishes the "risk landscape" of the corpus: while impact categories are formally present in every step, high impact is concentrated in the steps that carry the heaviest rhetori-

Table 2

Rhetorical segmentation of the corpus by CARS Move and Step

Move	Step	n	% of corpus	% within move
M1	M1_S1	53	13.7%	31.7%
M1	M1_S2	45	11.6%	26.9%
M1	M1_S3	69	17.8%	41.3%
M1	Total	167	43.0%	100.0%
M2	M2_S1	26	6.7%	28.3%
M2	M2_S2	42	10.8%	45.7%
M2	M2_S3	24	6.2%	26.1%
M2	Total	92	23.7%	100.0%
M3	M3_S1	41	10.6%	31.8%
M3	M3_S2	41	10.6%	31.8%
M3	M3_S3	47	12.1%	36.4%
M3	Total	129	33.2%	100.0%

cal load in terms of signalling aims and contribution. These findings provide the rationale for examining whether the observed high-impact deviations are tied to class membership, a question addressed in the following subsection.

Communicative Impact by Deviation Class

The next block of results evaluates how impact levels are distributed across the two superordinate classes of stylistic deviation. The purpose of this comparison is to establish whether the communicative consequences of deviations are class-specific or spread evenly across the corpus. The contingency Table is presented in Table 4.

The results demonstrate a marked imbalance. Out of 274 instances of wordiness, 60 are coded as high impact, representing 21.9% of that class. By contrast, redundancy yields only a single high-impact fragment out of 114 instances, which corresponds to 0.9%. The remaining redundancy cases cluster in the low- and medium-impact categories (69 and 43 respectively), showing that redundancy tends to carry minimal communicative risk even when frequent. Wordiness, on the other hand, is distributed across the full severity range: 104 low, 110 medium, and 60 high. A chi-square test confirms that impact is not independent of class, $\chi^2(2) = 29.979$, $p < 0.001$, Cramér's V = 0.278. The effect size indicates a medium association, meaning that class membership substantially influences the severity profile. Specifically, the high-impact category is overwhelmingly associated with wordiness, while redundancy is almost entirely confined to low- and medium-level effects.

This pattern clarifies the earlier step-level findings. Elevated high-impact rates in certain steps cannot be explained by an uneven balance of classes across the rhetorical map, because class distributions were shown to be stable. Rather, these elevated rates arise from the fact that wordiness, when it occurs, is more likely than redundancy to cause communicative overload. This conclusion sets the stage for the subcategory analysis, which pinpoints the precise forms of wordiness most responsible for high impact

Subcategory Profile and Carriers of High Impact

The final block of results turns from the superordinate classes to the finer subcategory level in order to identify which specific patterns of deviation are most strongly associated with communicative risk. The distribution of counts and high-impact rates across all nine annotated subcategories is presented in Table 5.

The corpus-wide frequency profile shows that several subcategories are well represented. Wordiness-related types are the most frequent, including syntactic complexity (63 fragments), hedging proliferation (59), and general low-

Table 3

Impact Distribution by Step and High-Impact Rates

Step	LOW	MEDIUM	HIGH	n total	High rate
M1_S1	26	20	7	53	0.132
M1_S2	22	18	5	45	0.111
M1_S3	33	23	13	69	0.188
M2_S1	14	12	0	26	0.000
M2_S2	18	16	8	42	0.190
M2_S3	14	9	1	24	0.042
M3_S1	16	17	8	41	0.195
M3_S2	19	16	6	41	0.146
M3_S3	17	19	11	47	0.234

Note. "High rate" equals n HIGH divided by n total per Step.

Table 4

Impact Distribution by Class

Class	LOW	MEDIUM	HIGH	n total	High rate
Wordiness	104	110	60	274	0.219
Redundancy	75	37	1	114	0.009

Table 5

Subcategory Counts and High-Impact Rates

Subcategory	n	n HIGH	High rate
WORDINESS_COMPLEXITY	63	54	0.857
WORDINESS_HEDGING	59	0	0.000
WORDINESS_GENERAL	54	4	0.074
REDUNDANCY_STRUCTURE	49	0	0.000
REDUNDANCY_LEXICAL	47	1	0.021
NOMINALIZATION	42	2	0.048
FORMULAIC_PHRASE	41	0	0.000
APPOSITIVE_PHRASE	18	0	0.000
EMPTY_REFERENCE	15	0	0.000

yield phrasing (54). Among redundancy-related types, structural repetition (49) and lexical tautology (47) are the most common, while appositive duplication is less frequent (18). Smaller categories of wordiness include nominalization (42), formulaic phrases (41), and empty references (15).

The severity distribution highlights a sharp concentration of high impact in one subcategory. Out of 61 high-impact fragments in the corpus, 54 belong to WORDINESS_COMPLEXITY, giving this subcategory a high rate of 0.857. The remaining subcategories contribute little to the high-impact pool: WORDINESS_GENERAL produces four high-impact cases (rate 0.074), NOMINALIZATION two (0.048), REDUNDANCY_LEXICAL one (0.021), and all others none. This concentration means that syntactic overload is by far the most consequential deviation in terms of communicative clarity.

The joint consideration of frequency and severity provides further clarification of the class-level results. Wordiness not only exceeds redundancy in overall prevalence but also encompasses the specific subcategory of syntactic complexity, which accounts for the vast majority of high-impact deviations. By contrast, the remaining subcategories, although regularly represented in the corpus, are predominantly confined to the low and medium levels of communicative severity.

DISCUSSION

The present study applied a taxonomy of stylistic deviations, previously validated in education-related corpora (Tikhonova & Mezentseva, 2024), to a new disciplinary (linguistics) and language (English). The findings not only replicate earlier observations about the predominance of wordiness but also refine the understanding of how communicative risk is distributed across rhetorical functions.

The CARS model has long been used to explain variation in Introductions (Swales, 2004; Samraj, 2002). In the present corpus, it provided a stable baseline against which deviations could be measured. Although the Step by Class analysis showed no significant association indicating that wordiness and redundancy are evenly spread across Moves, the Step by Impact results revealed localized risk zones. Specifically, high-impact deviations occurred in M3_S3, M3_S1, M2_S2, and M1_S3. These are steps with high rhetorical load, where authors articulate aims, contribution, problematization, or field orientation. The observation that severity rises in these zones aligns with previous studies showing that genre-imposed pressures intensify in precisely these rhetorical locations (Kanoksilapatham, 2005; Peacock, 2011; Öztürk, 2007; Pho, 2008; Shehzad, 2008; Lim, 2012). Thus, while class membership itself is not predicted by rhetorical function, severity is conditioned by the communicative demands of specific steps.

The cross-study contrast employed here is heuristic rather than strictly causal. The baseline corpus differs from the present dataset in both discipline and language, which precludes attributing observed divergences to either factor alone. Within these constraints, the present findings nonetheless show that high-impact deviations in linguistics Introductions align with sentence-level packaging at points of maximal rhetorical demand (Halliday & Martin, 2003; Banks, 2008). The results align with and extend the observations of Tikhonova et al. (2025). In their education corpus, redundancy showed clearer clustering in certain rhetorical steps, while in the present linguistics corpus wordiness proved to be a stable majority across the board. This contrast should be interpreted cautiously, because the corpora vary simultaneously by discipline and language. Even so, the present evidence indicates that linguistics Introductions concentrate stylistic load in densely packaged sentence structures, whereas the education corpus reported more visible cluster-

ing of redundancy in justificatory segments (Nwogu, 1997; Lim, 2012).

A balanced 2 × 2 design crossing discipline (linguistics vs. education) and language (English vs. Russian), with harmonized sampling windows and identical annotation, would allow the separate contributions of discipline and language to be estimated and would test whether the concentration of high-impact wordiness in linguistics generalizes beyond the present corpus.

The strong link between high impact and wordiness, and especially syntactic complexity, parallels findings from corpus-based studies of academic prose that document the increasing density of clause embedding and nominal constructions in research writing (Biber & Gray, 2010; Hyland, 2005). Psycholinguistic evidence indicates that heavy embedding and long-distance dependencies increase processing cost, which is consistent with the elevated impact observed here (Gibson, 1998; Levy, 2008). The present results are therefore in line with broader accounts of syntactic compression as a disciplinary stylistic norm that also introduces risks of reduced clarity.

Research Questions in Focus

The cross-reading of Tables 2, 3, and 5 underscores that severity is not an artefact of uneven class distribution. Wordiness is consistently dominant in every step, yet only one subcategory, syntactic complexity, accounts for nearly all high-impact cases (54 of 61). This concentration confirms that communicative overload is not diffused across stylistic features but localized in a single recurrent pattern. The implication is that stylistic guidance aimed at reducing syntactic complexity, particularly in Steps where rhetorical pressure is greatest, would yield the largest gains in interpretability (Öztürk, 2007; Pho, 2008; Shehzad, 2008). This insight complements earlier pedagogical recommendations that foreground clarity and explicitness in academic writing instruction (Williams & Bizup, 2017).

Deviation Classes at Corpus Level and by Step

The analysis permits a structured response to the three research questions. With respect to RQ1, which asked whether deviation classes exhibit a patterned distribution across rhetorical Moves and Steps, the results indicate that no such patterning is present. Wordiness consistently predominates in every step, and chi-square testing revealed no significant association between class and rhetorical function. This stability contrasts with findings from education corpora, where redundancy was reported to cluster in specific justificatory passages (Tikhonova et al., 2025). In the present linguistics corpus, the imbalance in favour of wordiness is a general stylistic tendency rather than a consequence of rhetorical positioning.

Turning to RQ2, which addressed the overall balance between wordiness and redundancy and the internal composition of their subcategories, the results confirm the dominance of wordiness, accounting for 70.6% of all annotated fragments, while redundancy contributed 29.4%. Within wordiness, the most frequent subcategories were syntactic complexity, hedging, and general low-yield phrasing. This profile again diverges from earlier studies, where redundancy-related categories such as structural repetition played a more central role. In linguistics writing, the stylistic burden is thus carried predominantly by expansions and overloaded constructions rather than by repeated formulations.

Finally, RQ3 concerned the localization of high-impact deviations and their association with class membership. High-impact cases were concentrated in four steps: M3_S3, M3_S1, M2_S2, and M1_S3, which are responsible for signalling contribution, stating aims, establishing the gap, and reviewing prior work. Importantly, severity was strongly tied to class, with high-impact deviations almost exclusively associated with wordiness and, within that class, overwhelmingly with syntactic complexity. This outcome reinforces the view that the most consequential risks in linguistics Introductions occur when rhetorical density coincides with syntactic overload. By comparison, redundancy, although present, rarely escalated to high severity and therefore poses less communicative risk in this disciplinary setting.

Taken together, the responses to RQ1–RQ3 reveal a disciplinary contrast. Whereas in education writing redundancy can be tolerated and even rhetorically functional in justificatory passages, linguistics writing carries its stylistic burden in densely packed sentences, making syntactic complexity the critical source of communicative risk. This comparison underscores the value of applying a unified taxonomy across domains, since only through cross-disciplinary analysis can field-specific vulnerabilities be identified and pedagogically addressed.

CONCLUSION

The aim of this study was to examine stylistic deviations in linguistics Introductions using a Move–Step framework combined with a taxonomy of wordiness and redundancy. The analysis produced three consistent findings. First, wordiness predominates across all rhetorical segments, with class balance stable across Moves and Steps. Second, the most frequent subcategories are syntactic complexity, hedging, and low-yield phrasing, while redundancy plays only a minor role. Third, communicative severity is functionally localized: high-impact deviations cluster in steps devoted to reviewing prior work, establishing the research gap, stating aims, and announcing contribution, and are overwhelmingly associated with syntactic complexity.

These results contribute to a clearer understanding of how stylistic risk is distributed in linguistics Introductions. They show that communicative burden arises not from repetition but from sentence structures that compress excessive syntactic material, especially in rhetorically dense segments of the text. The findings also carry practical implications, suggesting that efforts to improve clarity should prioritize strategies for reducing syntactic overload in high-pressure rhetorical contexts. More broadly, the study demonstrates the analytic and pedagogical value of combining rhetorical segmentation with a fine-grained taxonomy of stylistic deviations.

Several limitations must be acknowledged. The dataset was limited to 40 Introductions in linguistics, which constrains the generalizability of the findings. The reliance on categorical coding necessarily abstracts from contextual nuances, and the analysis did not consider author background or journal policies, both of which may influence stylistic choices. Addressing these limitations in future research will require larger, multi-disciplinary corpora, as well as discourse-functional and computational approaches capable of capturing gradient forms of stylistic overload. Such extensions would clarify whether the concentration of risk in syntactic complexity is specific to linguistics or characteristic of humanities writing more broadly, and would refine pedagogical recommendations for discipline-sensitive academic writing instruction.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTIONS

Tatiana Golechkova: conceptualization; data curation; formal analysis; funding acquisition; methodology; project administration; visualization; writing – original draft; writing – review & editing.

Nadezhda Arupova: formal analysis; investigation; methodology; resources; software; supervision; writing – original draft.

Elena Golubovskaya: conceptualization; data curation; formal analysis; funding acquisition; methodology; project administration; visualization; writing – original draft; writing – review & editing.

REFERENCES

Banks, D. (2008). *The development of scientific writing: Linguistic features and historical context.* Equinox.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to "overinformative" referring expressions. *Psychological Review, 127*(4), 591–621. https://doi.org/10.1037/rev0000186

Farneste, M. (2015). Moves in the introduction sections of applied linguistics research articles. B*altic Journal of English Language, Literature and Culture, 5*, 27–40. https://doi.org/10.22364/BJELLC.05.2015.03

Flowerdew, J., & Forest, R. W. (2015). *Signalling nouns in English: A corpus-based discourse approach.* Cambridge University Press. https://doi.org/10.1017/CBO9781139135405

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gong, H. & Barlow, M. (2022). A corpus-based analysis of research article macrostructure patterns. *Journal of English for Academic Purposes, 58*, 101138. https://doi.org/10.1016/j.jeap.2022.101138

Goonaratna, C. (2002a). Writing well (6) Wordiness, alias verbosity. *Ceylon Medical Journal, 47*(1), 1-3. https://doi.org/10.4038/cmj.v47i1.6393

Goonaratna, C. (2002b). Writing well (7) Wordiness alias verbosity, continued. *Ceylon Medical Journal, 47*(3), 79-80. https://doi.org/10.4038/cmj.v47i3.3432

Halliday, M. A. K., & Martin, J. R. (2003). *Writing science: Literacy and discursive power.* Routledge. https://doi.org/10.4324/9780203209936

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing.* Continuum.

Hyland, K. (2008). Genre and academic writing in the disciplines. *Language Teaching, 41*(4), 543–562. https://doi.org/10.1017/S0261444808005235

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes, 24*(3), 269–292. https://doi.org/10.1016/j.esp.2004.08.003

Kravtchenko, E., & Demberg, V. (2022). Informationally redundant utterances elicit pragmatic inferences. *Cognition, 225,* 105159. https://doi.org/10.1016/j.cognition.2022.105159

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lim, J.M.-H. (2012) How do writers establish research niches? A genre-based investigation into management researchers' rhetorical steps and linguistic mechanisms. *Journal of English for Academic Purposes, 11*, 229-245. https://doi.org/10.1016/j.jeap.2012.05.002

Leufkens, S. (2023). Measuring redundancy: The relation between concord and complexity. *Linguistics Vanguard, 9*(s1), 95-106. https://doi.org/10.1515/lingvan-2020-0143

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes, 16*(2), 119–138. https://doi.org/10.1016/S0889-4906(97)85388-4

Öztürk, İ. (2007). The textual organisation of research article introductions in applied linguistics. *Journal of English for Academic Purposes, 26*, 25-38. https://doi.org/10.1016/J.ESP.2005.12.003

Peacock, M. (2011). The structure of the Methods section in research articles across eight disciplines. *The Asian ESP Journal, 7*(2), 97–124.

Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies, 10*, 231-250. https://doi.org/10.1177/1461445607087010

Raitskaya, L. K., & Tikhonova, E. V. (2019). Multilingualism in Russian journals: A controversy of approaches. *European Science Editing, 45*(2), 41. https://doi.org/10.20316/ESE.2019.45.18024

Raitskaya, L., & Tikhonova, E. (2020). Overcoming cultural barriers to scholarly communication in international peer-reviewed journals. *Journal of Language and Education, 6*(2), 4-8. https://doi.org/10.17323/jle.2020.11043

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes, 13*(2), 149–170. https://doi.org/10.1016/0889-4906(94)90013-2

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes, 21*(1), 1–17. https://doi.org/10.1016/S0889-4906(00)00023-5

Shehzad, W. (2008). Move Two: Establishing a niche. *Ibérica, 15*, 25–50.

Smirnova, N. V., Lillis, T., & Hultgren, A. K. (2021). English and/or Russian medium publications? A case study exploring academic research writing in contemporary Russian academia. *Journal of English for Academic Purposes, 53*, Article 101015. https://doi.org/10.1016/j.jeap.2021.101015

Swales, J. M. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

Swales, J. M. (2004). *Research genres: Explorations and applications.* Cambridge University Press.

Tikhonova, E.V., Kosycheva, M.A., & Mezentseva, D.A. (2024). Ineffective strategies in scientific communication: Textual wordiness vs. clarity of thought in thesis conclusion section. *Integration of Education, 28*(2), 249–265. https://doi.org/10.15507/1991-9468.115.028.202402.249-265

Tikhonova, E. V., & Mezentseva, D. A. (2024). Wordiness in academic writing: A systematic scoping review. *Research Result. Theoretical and Applied Linguistics, 10*(1), 133–157. https://doi.org/10.18413/2313-8912-2024-10-1-0-8

Tikhonova, E., Zavolskaya, O., Mekeko, N. (2025). Stylistic redundancy and wordiness in introductions of original empirical studies: Rhetorical risks of academic writing. *Journal of Language and Education, 11*(2), 125-136. https://doi.org/10.17323/jle.2025.27389

Williams, J. M., & Bizup, J. (2017). *Style: Lessons in clarity and grace* (12th ed.). Pearson.

https://doi.org/10.17323/jle.2025.27858

Exploring the Affordances of ChatGPT in Developing Language Teachers' Awareness and Pedagogical Practices

Dara Tafazoli [1], **Lee McCallum** [2]

[1] The University of Newcastle, Callaghan, Australia

[2] The University of Edinburgh, Scotland, United Kingdom

ABSTRACT

Background: The field of language education is experiencing substantial advancements with the introduction of Generative Artificial Intelligence (GenAI). While there is a significant amount of research on teachers' perceptions in language education, there is a noticeable gap in understanding the integration of ChatGPT into this field.

Purpose: This study investigated the perceptions of EFL teachers in Iran regarding how ChatGPT can contribute to the enhancement of language teaching practices and pedagogical awareness.

Method: A qualitative case study examined teachers' perspectives on using ChatGPT in language education. A Zoom webinar on AI in teaching was conducted for Iranian EFL teachers, using tools like Focusky. From 73 registered teachers, 23 were selected based on ≥2 years' teaching experience and voluntary participation. Data were gathered through focus groups, semi-structured interviews (n=10, in Persian), and reflective essays (n=10) submitted two weeks post-webinar. Inductive thematic analysis was performed using NVivo 12, with rigor ensured via peer debriefing, member checking, cross-case comparison, and 83% inter-coder agreement.

Results: Data analysis revealed that ChatGPT can promote pedagogical awareness (i.e., enhanced professional development opportunities, enhancement of critical thinking skills, and autonomy and confidence) and pedagogical practices (i.e., understanding students' needs and personalizing teaching and content creation and resource development).

Conclusion: These findings suggest that ChatGPT can positively impact language teachers without necessarily sacrificing their jobs, which is the dominant concern in the current academic literature.

KEYWORDS

ChatGPT; Generative Artificial intelligence (GenAI); Computer-Assisted Language Learning (CALL); pedagogical awareness; pedagogical practices

Citation: Tafazoli, D., & McCallum, L., (2025). Exploring the affordances of chatgpt in developing language teachers' awareness and pedagogical practices. *Journal of Language and Education, 11*(3), 86-97. https://doi.org/10.17323/jle.2025.27858

Correspondence:
Dara Tafazoli,
dara.tafazoli@newcastle.edu.au

Received: May 06, 2025
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

Artificial Intelligence and the Rise of Generative AI

The field of Artificial Intelligence (hereafter AI) has the potential to revolutionize language education. The integration of emerging technologies like AI into language education has already shown promise through Intelligent Tutoring Systems (ITS) (Gamper & Knapp, 2010; Schulze, 2008; Stockwell, 2007), Automated Writing Evaluation (AWE) (Han & Sari, 2024), automatic speech recognition (Inceoglu et al., 2023), Robot-Assisted Language Learning (RALL) (Engwall et al., 2022), Chatbots (Jeon, 2023) and Extended Reality (XR) (Khazaei & Ebadi, 2023). However, there is still much-untapped potential for AI in education (Godwin-Jones, 2023), specifically the recent development of Large Language Models (LLMs) as a contributing factor to integrating AI-based tools into teachers' teaching practices (Bonner et al., 2023). AI-based systems known as LLMs are designed to understand and generate human-like text based on the input they receive (Liu et al., 2024). These systems use deep learning techniques, particularly transformer architectures, and are trained on vast amounts of textual data

to learn patterns, semantics, and syntax of human language (Kasneci et al., 2023; Yan et al., 2024).

Emergence of Generative AI and ChatGPT

One increasingly popular batch of LLM-AI-led technology is generative AI (GenAI), particularly and their subsequent tools, like ChatGPT, that use huge collections of past textual data to train the eventual model to produce human-like responses to prompts written by a user (see Kohnke et al., 2023). This breakthrough technology offers innovative pedagogical affordances for language learning, such as stimulating learners' interests (Kohnke, 2022), promoting higher-order thinking skills (Kasneci et al., 2023), providing personalized learning based on language learners' proficiency levels (Kuhail et al., 2023), and facilitating an authentic, interactive language-learning environment (Chiu et al., 2023), which calls for more research into this field (Tlili et al., 2023). Recent studies have primarily concentrated on evaluating the efficacy of this technology (Barrot, 2023, 2024; Kohnke et al., 2023) or emphasizing student-centered perspectives (Yan, 2023; Yuan et al., 2024). Against this backdrop, the present study seeks to explore the perceptions of English as a Foreign Language (EFL) teachers regarding the incorporation of ChatGPT into language teaching, aiming to offer insights that could guide the creation of effective policies within English education.

Affordances of ChatGPT for Language Learning

The field of language education is experiencing substantial advancements with the introduction of ChatGPT. This section of the literature review critically examines the contributions and limitations of various studies on the role of ChatGPT in language education. This technology has transformed traditional chatbots (see Gao et al., 2024) and offers remarkable affordances. Some of these affordances include providing students with opportunities to practice their written language via prompt-response 'dialogue' like exchanges (e.g., Alexander et al., 2023; McCallum, 2023). Students can also practice and enhance their grammatical and lexical knowledge by prompting these models to support them with planning written work, generating texts, and providing feedback on a text that is entered with a prompt (e.g., Barrot, 2023; Cotton et al., 2024; Huang et al., 2022). The affordances mentioned above, along with other functionalities like stimulating learners' interests in language learning (e.g., Kohnke, 2022), promoting higher-order thinking skills (Kasneci et al., 2023), being available 24/7, allowing students to practice their language skills anytime and anywhere (e.g., Haristiani, 2019), providing personalized learning based on language learners' proficiency levels (Kuhail et al., 2023), facilitating an authentic, interactive language-learning environment (Chiu et al., 2023), conducting formative assessments with immediate feedback (Cotton et al., 2024; Huang et al., 2022; Kuhail et al., 2023) mean that these tools have captured the attention of language teachers. AI-powered tools have become more appealing to teachers, specifically when recent research has also demonstrated their efficacy in assisting teachers with evaluating the teaching process as well as facilitating lesson planning (Celik et al., 2022; Zawacki-Richter et al., 2019).

Controversies and Ethical Concerns

However, despite the vast possibilities to aid language learning, ChatGPT's impact on the field has sparked substantial controversy Some experts acknowledge its potential as an educational support tool and a potential disruptor of the status quo (e.g., Kohnke et al., 2023), while others emphasize its disadvantages and potential risks, such as malpractice, trustworthiness, and cultural bias (Eke, 2023; Kostka & Toncelli, 2023). There is ongoing debate surrounding the ethical use of ChatGPT in language education. Some worry that it could facilitate malpractice (e.g., cheating) (see Eke, 2023; Saville, 2023) and have concerns about its impact on fair assessment (see Rudolph et al., 2023). For instance, Noam Chomsky referred to ChatGPT as "high-tech plagiarism"[1], while Weismann (2023) essentially decried the end of the teaching profession.

Another issue is the originality and reliability of ChatGPT's responses, particularly, for those whose first language is not English may be vulnerable to false positives from the AI-detection software[2] (Reed, 2023). Also, the tool does not provide sources or citations if not requested, leading some to question whether its responses are truly original or simply paraphrased from uncited sources. Even if requested, ChatGPT gives false citations in most of the cases, which in both cases leads to plagiarism and academic misconduct. Another debate focuses on the cultural bias inherent in the source database and algorithms (Rettberg, 2022) where most of the text in the database is derived from an English corpus and then translated into target languages. In addition, it includes many words that are written more often than spoken. These issues are particularly concerning in language education, as students are likely to come from diverse cultural backgrounds and may be unaware that ChatGPT and other AI-driven tools are not culturally neutral.

1 Stewart, J. (2023, February 17). Noam Chomsky says ChatGPT is a form of "high-tech plagiarism". My Modern Met https://mymodernmet.com/noam-chomsky-chatgpt/#:~:text=ChatGPT%20is%20basically%20high%2Dtech,to%20avoid%20doing%20- the%20work

2 Reed, B. (2023, July 10). Programs to detect AI discriminate against non-native English speakers, shows study. *The Guardian*. https://shorturl.at/hmnP2

Teachers' Role

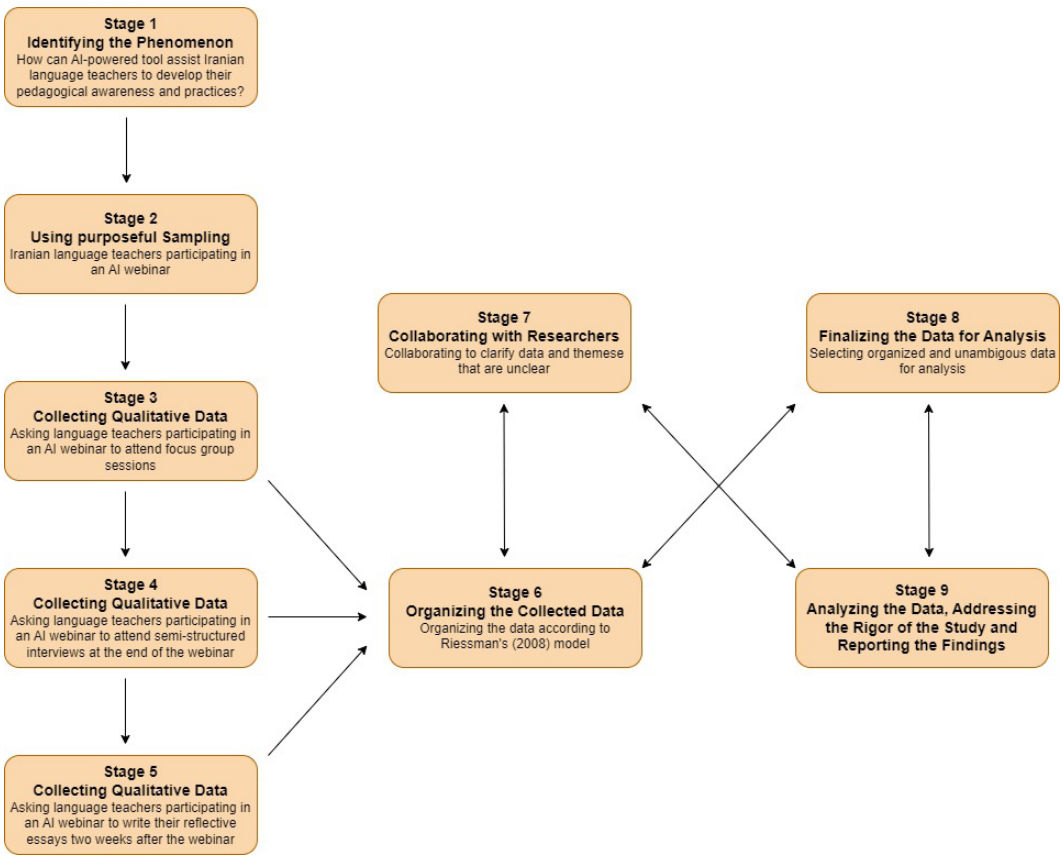
Considering the pros and cons of the integration of ChatGPT in language education, the impact of teachers should not be overstated as they have the potential to shape the use of ChatGPT in the classroom and curriculum with ethical and appropriate practices (Trust et al., 2023). However, incorporating AI-powered tools, like ChatGPT, into language teaching necessitates that language teachers to be well-versed in its pedagogical implications. Despite this necessity, few studies have explored teachers' perceptions of ChatGPT. As explained above, previous research has predominantly focused on learners.

It is important to note that while there is a significant amount of research on teachers' perceptions in language education, there is a noticeable gap in understanding the integration of ChatGPT into this field. Only one study has investigated teachers' perspectives on incorporating ChatGPT in English language teaching (Al-khresheh, 2024). However, this study had limitations in participant recruitment, such as selection bias, and involved participants from diverse contexts with varying technological perspectives. To address these limitations, Al-khresheh (2024) recommended conducting research with preliminary orientation sessions with ChatGPT to ensure feedback is based on firsthand experience. The study may have also underrepresented the cultural nuances intertwined with pedagogical methodologies across regions, indicating the need to explore culturally specific pedagogical contexts to shed light on region-specific complexities in AI tool adoption. The author also suggested that educational institutions should organize AI-centered pedagogical seminars to equip teachers with the latest skills and knowledge. Further research into the application of AI, including ChatGPT, is highly recommended to address unresolved areas identified in the study (Al-khresheh, 2024).

Informed by Al-khresheh (2024) and building on previous literature, this study is rooted in the significance of teachers' perceptions in shaping their teaching approaches and practices (Liu & Wang, 2024) and its impact on their professional identity and development (Heyder, 2019). Through this, the study aims to investigate the perceptions of EFL teachers in Iran regarding how ChatGPT can contribute to the enhancement of language teaching practices and pedagogical awareness. The central research question driving this qualitative study is: How do Iranian EFL teachers perceive the integration of ChatGPT in aiding the development of their pedagogical awareness and practices?

Figure 1

The Stages of the Study



METHOD

Research Design

In order to address the gap in understanding teachers' perspectives on incorporating ChatGPT into language education, a qualitative research design was chosen. This approach allows for a comprehensive examination of the intricate dynamics (Yin, 2018) involved in AI integration within real-world educational settings. The case study method is particularly suitable for this research as it facilitates a thorough exploration of the specific context in which ChatGPT is being utilized, offering valuable insights into its implications. Figure 1 shows the stages of the study.

Webinar, Sampling Technique and Participants

To address the necessity of teachers' professional development and to upskill their CALL competencies, the first author was invited by a private sector in language education to organize a series of webinars on CALL for Iranian language teachers. The webinar which is the focus of the current study titled 'Artificial intelligent advancement in language education.' The webinar was held using Zoom as the primary platform, and the instructor employed various tools, including Focusky and multimedia tools, to enhance the interaction with participants and the quality of instruction. The webinar consisted of different sections that explored topics such as the timeline of AI progress in language education, basic concepts in AI, critics of AI, the practical implementation of AI in language education, and the Q&A sessions.

According to Czarniawska (2004), purposeful sampling involves selecting a subset of participants who represent the population based on basic screening criteria. In this case study, purposeful sampling was used to choose the participants. The participants were chosen from this specific webinar participants according to the following criteria: (1) they needed to be a language teacher; (ii) they had to participate in the study voluntarily; (iii) they were required to have at least two years of teaching experience during the time when data was being collected.

The first author submitted an invitation email to around 200 language teachers who attended previous CALL workshops

Table 1

Research Participants' Demographic Information

		N = 23	%
Gender			
	Male	8	35%
	Female	15	65%
Age			
	20-24	1	4.3%
	25-29	4	17.5%
	30-34	4	17.5%
	35-39	5	21.7%
	40-44	8	34.7%
	45+	1	4.3%
Teaching experience			
	1-5 years	15	65.2%
	6-11 years	5	21.8%
	12-17 years	2	8.7%
	18-23 years	1	4.3%
Academic degree			
	MA student	5	21.7%
	MA	9	39.3%
	PhD student	8	34.7%
	PhD	1	4.3%

and/or webinars. Among them, 73 English as a Foreign Language (EFL) teachers completed the expression of interest (EOI) forms. The webinar link was sent out to 73 registered teachers, yet only 23 teachers attended and engaged in the webinar. Prior to conducting this study, all participants were given information about the study and agreed to participate. The researchers also obtained informed consent from each individual participant in the study to publish the results. All 23 EFL teachers agreed to participate voluntarily in the research project. This number of participants would be appropriate for a qualitative research study (Lichtman, 2013). These EFL teachers were selected to provide detailed and informative data on the central phenomenon of the study. Demographic information, such as age, gender, teaching experience, and highest qualification, was collected and categorized, and presented in Table 1.

Data Collection

This study employed three complimentary data collection methods: focus group sessions, semi-structured interviews, and reflective essays (see Figure 2). Using multiple methods helps validate results and increases the reliability of a study by capturing diverse perspectives and providing a richer data set (Salmons, 2014). The subsequent sections provide further details about these tools.

Focus Group Sessions

Focus group sessions were organized to gain initial insights into teachers' use of ChatGPT and to foster group discussions that could reveal collective attitudes and experiences. These sessions enabled participants to engage with one another and expand on each other's responses, creating a dynamic and interactive setting for data collection. Although no pressure was exerted on any participants to participate, all 23 teachers were involved and volunteered in the focus group after each session. After completing each section of the webinar, the focus group sessions were held through oral group interviews, which lasted for 20-30 minutes. The sessions were transcribed and coded, and around eight questions were asked during the focus group sessions, with teachers responding to the questions they chose, followed by feedback from other focus group members. All of the questions were created by the researchers and were rooted in the sessions' content.

Semi-Structured Interviews

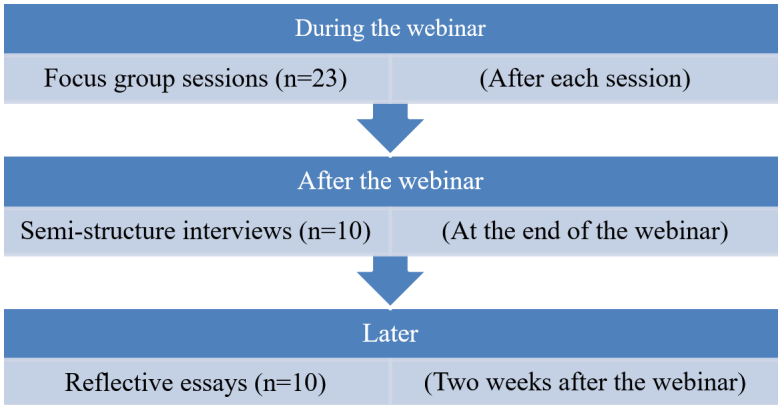
Semi-structured interviews were chosen to delve deeper into individual teachers' perceptions and experiences with ChatGPT. This method provides structure with predetermined core questions while allowing flexibility for follow-up questions and probing based on participants' responses. By conducting one-on-one interviews, the researchers ensured that participants could express their thoughts and feelings without external influence, such as peer pressure, which is essential for obtaining honest and detailed insights (Orvik et al., 2013). At the end of the webinar, ten participants took part in these 45-60 minute interviews individually. All interviews were recorded for transcription purposes. Since both the interviewer (first author) and the participants share Persian (or Farsi) as their mother tongue, all ten interviews were conducted in this language to ensure clarity of expression. To gather additional information, the interviewer used probes such as "Could you explain your response more?" or "Tell me more…" (Creswell, 2012, p. 221).

Reflective Essays

Reflective essays, written two weeks after the webinar, were employed to capture participants' more thoughtful and introspective feedback on their experiences with ChatGPT. The delayed writing of the reflective essays allowed participants the time to contemplate their interaction with ChatGPT and apply it in their actual teaching environments. This data collection method offered a longitudinal view and fostered deeper reflection, complementing the immediate feedback from focus groups and interviews. While the essays contained guiding questions, teachers were free to write as much or as little as they desired in response to the prompts. This section employed data from reflective essays

Figure 2

Stages of Data Collection



submitted by ten individuals. The first author opted to use a set of prompts for reflective essays, covering all areas of the course material rather than just the ChatGPT content. Teachers were asked to share their experiences with ChatGPT used during the webinar. Specific questions about AI were included:

(1) To what extent are you content with the ChatGPT for teaching and learning?
(2) What aspects of the AI did you appreciate the most? Can you give specific examples of activities that you enjoyed?
(3) Did the AI contribute to your understanding of teaching knowledge and skills?
(4) In what ways can AI enhance your experience with teaching practices?

Data Analysis

The researchers conducted the current study to investigate how language teachers in Iran could benefit from an AI-powered tool called ChatGPT, in terms of developing their pedagogical awareness and practices. The data collected from three sources were analyzed using thematic analysis (Braun & Clarke, 2006). To achieve this, the researchers utilized inductive thematic analysis. This method involves deducing meaning and recognizing patterns from raw data without any preconceived notions (Braun & Clarke, 2006). The advantage of using inductive thematic analysis is that it produces categories based on the information gathered and prioritizes the opinions and perspectives of the research participants over those of the researchers (Braun & Clarke, 2006). Essentially, inductive coding promotes the creation of coding categories that are derived from the initial data.

The researchers utilized NVivo 12 to follow Braun and Clarke's (2006) phases of thematic analysis. They approached the data thematization through multiple recursive phases. The researchers re-read the data multiple times during the initial phase after inputting it into NVivo 12. As per the study conducted by Braun and Clarke in 2006, the second phase of the research involved organizing the data obtained from the initial concepts and uncovering implicit extracts. These extracts were not explicitly stated but inferred from relevant literature. For instance, upon encountering a sentence like "ChatGPT offers various opportunities for my career growth," the researchers identified the first relevant idea - professional development. In the third phase of the research, the initial codes were grouped into potential themes. To do this, the researchers used software to create potential parent codes and themes. It was important to discuss these parent codes and potential themes with other experts to ensure that they accurately represented the participants' intentions. Two research assistants also read the data to identify common themes and created a list of examples for each theme. They evaluated the frequency of each theme both quantitatively and qualitatively by comparing their findings to those from the literature. The researchers shared their findings with the first researcher and formed a panel to reach a consensus on the final themes that all three agreed upon. After some modifications, deletions, and combinations based on the invited experts, the researchers created a comprehensive thematic map using NVivo 12 in the fourth phase.

After analyzing the data, the researchers checked the obtained themes for consistency across all data sets. For this purpose, the researchers renamed the themes based on interesting words or phrases in teachers' data. In the sixth phase, the researchers selected some vivid examples of extracts and provided the report in the Findings section.

The Rigor of the Study

To achieve reliable results in qualitative research, it is essential to ensure rigor. This study focused on three elements of rigor: credibility, transferability, and dependability, as suggested by Ary et al. (2014). To improve credibility (similar to internal validity in quantitative research), we used the peer debriefing technique, where two research assistants in applied linguistics reviewed parts of the data to check if the themes identified were consistent. We also used member checking to verify our themes with all of the participants to strengthen the thematization process. To address transferability (similar to external validity in quantitative research), we used cross-case comparison by including participants with different educational levels and teaching experiences from various parts of Iran to increase the generalizability of our findings. Finally, we ensured dependability (similar to reliability in quantitative research) by evaluating the in-

Table 2

Frequency of Themes and Sub-Themes

Theme	Sub-themes	Freq.
Pedagogical awareness	Enhanced professional development opportunities	140
	Autonomy and confidence	21
	Enhancement of critical thinking skills	15
Pedagogical practice	Understanding students' needs and personalizing teaching	44
	Content creation and resource development	36

ter-coder agreement. All data were independently coded by three coders: two research assistants and the primary researcher. The inter-coder agreement was assessed based on their coding of the entire dataset. We achieved an 83% inter-coder agreement, indicating strong consistency in the interpretation of the data (McHugh, 2012).

RESULTS

According to the research question of the study, data analysis revealed two main themes of teachers' pedagogical awareness and pedagogical practice that emerged from the focus groups, interviews, and reflective essays. As shown in Table 2, these main themes are categorized into various sub-themes: 1) Pedagogical awareness: a) enhanced professional development opportunities, b) enhancement of critical thinking skills, and c) autonomy and confidence, 2) Pedagogical practice: a) understanding students' needs and personalizing teaching, and b) content creation and resource development.

The participants also discussed several challenges they faced during their preparation, but this study did not focus on these challenges, and therefore they were not included in the findings.

Pedagogical Awareness

Based on the data analysis, it was evident that teachers view ChatGPT as a valuable tool for improving their pedagogical awareness. The feedback from the participants underscored various important areas where ChatGPT contributes to their professional growth and development, such as offering enhanced professional development opportunities, fostering autonomy and confidence, and promoting critical thinking skills.

Enhanced Professional Development Opportunities

In her reflective essay, Azadeh emphasizes how ChatGPT provides professional development opportunities tailored to teachers' specific goals by recommending suitable courses, conferences, workshops, and resources:

> ChatGPT provides me with diverse possibilities for professional development. When I sought its assistance, it proposed numerous internet-based courses that were consistent with my professional development objectives, suggested in-person and virtual conferences and workshops that were pertinent to my field of expertise and interests, and shared professional development resources, such as scholarly articles and instructional videos that provided me with knowledge on teaching methods and trends in education. (Azadeh, female – Reflective essay)

Teachers found ChatGPT to be a flexible tool that allowed them to learn at their own pace and according to their schedules, enabling continuous professional development and lifelong learning. In a focus group, Habib pointed out the advantages of this flexibility: "*I have the freedom to access ChatGPT at any time, day or night, which allows me to learn whenever I have the time and desire. This flexibility enables me to manage my professional development around my hectic schedules*" (Habib, male – Focus group). Vania echoed this sentiment, "*with ChatGPT, I can learn at a pace that suits me and take breaks whenever necessary. This flexibility ensures that I do not feel stressed or pressured, which can improve my overall learning experience and professional development*" (Vania, female – Focus group). Sama emphasized the importance of attaining an improved work-life balance: "*I believe that it is beneficial in achieving a better work-life balance. ChatGPT enables me to plan my professional development around my personal obligations, which can alleviate stress and improve the balance between my work and personal life*" (Sama, female – Interview).

Teachers highlighted that ChatGPT has the capability to personalize the learning process according to individual needs and preferences, facilitating personalized professional development. Hamid shared his firsthand experience: "*ChatGPT has the ability to be personalized in order to meet my unique learning needs and preferences. I have the option of selecting the topics I want to study and the depth of knowledge I desire, providing me with a tailored learning experience*." (Hamid, male – Interview).

Teachers have also observed that ChatGPT's platform undergoes consistent updates with the most recent trends, knowledge, and effective methodologies in the field of education. This resource enables them to improve their teaching strategies and remain abreast of the latest developments in language education. Kian highlighted this during his interview:

> The AI-based environment is continuously refreshed with the most recent education-related knowledge, trends, and effective techniques. This guarantees that I have access to the most pertinent and current information that can aid me in enhancing my teaching methods and staying up-to-date with the newest advancements in the education sector. (Kian, male – Interview)

Autonomy and Confidence

The data analysis demonstrates that ChatGPT can boost teachers' autonomy and confidence, enabling them to take charge of their own learning and professional development. This empowerment leads to greater job satisfaction and a more rewarding work experience, as expressed by Arvid in the focus group "*through the workshop, I discovered that I have the ability to manage and direct my own learning and professional development using AI technology. As a result, I feel a greater sense of satisfaction and investment in my role as an English language teacher*" (Arvid, male – Focus group). Soosan emphasized the significance of self-directed learning during her interview: "*my primary objective as a teach-*

er is to enhance my skills and knowledge through self-initiated learning. By doing so, I can gain greater confidence in my abilities and effectively implement my learning to improve my teaching practices and enhance my student's academic results" (Soosan, female – Interview).

Enhancement of Critical Thinking Skills

The participants recognized that ChatGPT has the potential to enhance their critical thinking abilities through the presentation of diverse viewpoints and opinions, thereby aiding in making well-informed decisions and effectively addressing problems. Tara highlighted this in her interview:

> The instructor explained that ChatGPT can provide teachers with a range of viewpoints and opinions on a topic, which can help them develop a more comprehensive understanding of the subject matter. By asking questions and seeking out new information, teachers can enhance their critical thinking skills and become better problem solvers and decision-makers. This, in turn, can lead to improved teaching outcomes for students and greater job satisfaction for teachers. (Tara, female – Interview)

Neda also described using an AI-powered decision-making framework to approach complex problems more critically:

> I used an AI-powered decision-making framework to help me approach complex problems in a more critical manner. The framework recommended a structured process, starting with problem identification, followed by problem analysis, generating possible solutions, evaluating options, selecting the best option, and finally implementing and monitoring the solution. (Neda, female – Reflective essay)

Pedagogical Practice

Understanding Students' Needs and Personalizing Teaching

Teachers have noted that AI-supported platforms enable them to gain valuable insights into students' needs and learning styles, which in turn allows for more personalized and effective teaching methods. Through the analysis of student performance data, educators are able to pinpoint strengths and areas for improvement, and provide tailored feedback. Aref described how he used ChatGPT for data analysis, "*following the workshop, I utilized ChatGPT to analyze my students' performance data, enabling me to pinpoint their strengths and weaknesses. This analysis provided me with personalized feedback that I could act on, resulting in improved teaching practices*" (Aref, male – Reflective essay). Nazanin, in her interview, shared that:

> According to the workshop instructor, ChatGPT can analyze student performance data to identify patterns and trends, enabling teachers to gain a deeper understanding of their students' needs and learning styles. This can assist in providing personalized feedback to students and helping them with their individual challenges, ultimately leading to more effective teaching practices. (Nazanin, female – Interview)

Content Creation and Resource Development

AI tools provide support to educators in multiple facets of content development, including idea generation, research, and editing. This assistance ensures the creation of compelling and exceptional educational content. Solmaz has implemented ChatGPT for lesson planning and said that "*it helped me receive feedback on my plans and suggested ways to make them more innovative and captivating. I believe that this can be a valuable resource for teachers to enhance their lesson-planning abilities and generate fresh ideas for upcoming classes*" (Solmaz, female – Reflective essay).

Amin and Eli offered specific use cases demonstrating how ChatGPT facilitates content creation.

> When I was preparing a language test for my students, I asked ChatGPT to create a daily conversation, and it did! Also, I asked it to proofread the whole test, and it did! Also, I asked to make three multiple-choice questions of that conversation, and it did it again! Isn't it amazing? I saved hours. (Amin, male – Interview)

> One of my main challenges in online courses is students' cheating! ChatGPT helped to create a test for each student. It was wonderful as it provided me with 43 reading comprehension tests at the same level of proficiency with the same topic, but different words and content. Humans cannot do it for sure. (Eli, female – Reflective essay)

DISCUSSION

Pedagogical Awareness

The findings indicate that ChatGPT offers substantial benefits for teachers' professional development. However, it is important to acknowledge that ChatGPT's suggestions may sometimes be superficial and not tailored to specific, nuanced teaching contexts. Therefore, in developing an appropriate contextualized program of training for teachers, actual human input would be essential to supplement and balance the training recommended by ChatGPT. When considering teacher training in challenging sociopolitical contexts (like Iran), this balance seems especially relevant to ensure appropriate and contextualized development for teachers (Hubbard, 2023; Nami, 2023).

More widely, our findings are evidence of a paradigm shift in thinking about how AI and teachers can collaborate to achieve particular educational goals. In this case, the added support from ChatGPT may further foster self-directed professional development among teachers and, therefore, put into practice the now long-standing message that teachers need to take responsibility for educating themselves with technology and also responsibility for their wider development (Stockwell, 2009). In this respect, one hurdle that ChatGPT may help overcome is recommending a wide enough range of teacher resources that are accessible and

low-cost/free. This may help maintain teachers' motivation and deter them from giving up pursuing self-directed learning because recommended resources are unavailable or difficult to locate, particularly in print form (Stockwell, 2009).

The flexibility afforded by ChatGPT allows teachers to engage with flexible and convenient professional development. This kind of development is markedly different from the often rigid, scheduled professional development that is provided by institutions. When conceptualized along a cline from face-to-face input to online provision (e.g., see Floris, 2021), we see here that teachers tend to favor online provision over the opposite end of the cline (i.e., traditional face-to-face input). This finding has important implications for how we envisage the provision and delivery of teacher professional development initiatives. It also reiterates and reminds us of the significance of previous observations from Compton (2009) who acknowledges the changing roles and responsibilities of professional development stakeholders in light of the increasing uptake of online development opportunities.

When it comes to personalized professional development, teachers mentioned that ChatGPT could offer tailored opportunities. Alongside the first theme of professional development and growth mentioned above, this capability of ChatGPT may be one way of practically maintaining and protecting teacher motivation along the continuous professional development cycle, with this latter point heavily advocated for in similar contexts to the Iranian one we are studying here (e.g., see Lamb & Wyatt, 2019). These findings are intriguing as the existing literature lacks information on tailored professional development opportunities for educators, with the emphasis solely on personalized learning environments for students (Kuhail et al., 2023).

Connected to personalized professional development is that teachers felt that ChatGPT could provide them with opportunities to develop autonomy and confidence because they can direct their own learning. This finding has important implications for current teacher training in language education. It opens up the possibility of teachers taking control of the design of personalized professional development plans in place of mandatory generic training programs which may be too broad for teachers who want to specialize in particular skills or interests (Borg, 2015). Mandatory generic training may negatively impact their professional development goals if it does not meet their current needs, and thus, suggestions from AI tools may offer greater autonomy and confidence as teachers take control of their own development and shift from focusing on institutional to individual professional development (Borg, 2015).

Teachers appreciated that ChatGPT could provide answers to prompts and include a range of different opinions within the texts. Teachers believed this could help support developing their critical thinking skills and also support them in developing a comprehensive understanding of particular topics. This finding, like the other themes, is further evidence that AI tools like ChatGPT have the ability to shift the focus from text creation as a teacher to focus on using text to develop content understanding and critique, thus promoting a focus on higher-order thinking skills for these teachers (Kasneci et al., 2023).

Pedagogical Practices

Teachers articulated that ChatGPT had the potential to understand student performance and their needs. ChatGPT can allow teachers to generate patterns in student performance data and then they can focus on interpreting that data for their particular course or learning context. In line with the other themes which emerged from the data, this finding highlights how ChatGPT may be able to reduce time spent on collation activities, allowing teachers to spend more time on cognitively demanding tasks like interpreting students' weaknesses and strengths and formulating appropriate strategies for managing those (see Alexander et al., 2023; Haristiani, 2019; Kasneci et al., 2023).

The vast expanse and power of ChatGPT for performing demanding tasks such as providing feedback were illuminated under the theme of creativity. Teachers reported that their lessons had become more creative and innovative methodologically, as a result of lesson plan feedback from ChatGPT. This finding again has implications for teacher trainer roles and the relationship between teacher, trainer, and professional development. If ChatGPT is able to provide feedback to make lessons more innovative and creative, we believe there is a possibility that the relationship and roles between teachers and teacher trainers will change to accommodate greater input and support from AI. The findings align with the integration of other CALL tools, such as Intelligent Tutoring Systems (see Gamper & Knapp, 2010; Schulze, 2008; Stockwell, 2007) and Automated Writing Evaluation (Han & Sari, 2024), and also support the recent research on AI-powered tools (see Cotton et al., 2024; Huang et al., 2022; Kuhail et al., 2023).

Teachers also expressed surprise that ChatGPT could create impressive content. In reflective essays and interviews, teachers mentioned a broad range of content creation prompts that ChatGPT had been able to produce output for. ChatGPT's capabilities to produce test content, proofread, and manipulate text to provide different versions of texts for assessments clearly have important implications for test creation and administration which are in line with other studies (Huang et al., 2022; Kuhail et al., 2023). As highlighted by one teacher, the amount of human time needed to prepare different text versions for individualized student assessment is immense and inefficient if ChatGPT can create the same output in a quicker time. This finding has important implications for test validity and test security because it can help teachers reduce peer-to-peer collusion and plagiarism, as

each student has an individualized text and task. A further point to note is that these thoughts of Eli also remind us that the rapid generation of texts can also help assessors focus on creating questions and tasks around the content of these texts, which can then, in turn, be used to promote assessing students' analytical, critical, communicative, and problem-solving abilities (Cotton et al., 2024).

CONCLUSION

This study aimed to investigate how ChatGPT can enhance language teachers' pedagogical awareness and professional practices. The findings suggest that teachers view ChatGPT as a valuable, self-directed development tool that supplements existing pedagogical frameworks, facilitating personalized, ongoing learning. ChatGPT's integration into teachers' professional routines has the potential to support classroom management and streamline traditionally time-consuming tasks, such as creating reading materials, generating test questions, and proofreading, thereby allowing teachers to dedicate more time to higher-order instructional activities like evaluating student performance and engaging deeply with curriculum design. Although some limitations were noted, including ChatGPT's limited ability to address cultural and affective factors, its challenges in facilitating speaking skills, and teachers' concerns regarding academic integrity, participants generally regarded the tool's disruptive nature as an opportunity for professional growth rather than a threat.

These findings indicate that ChatGPT can foster a shift toward a more autonomous, flexible, and personalized approach to professional development for language teachers, moving away from top-down models that may not always address individual needs effectively. The tool's adaptability and accessibility make it especially suitable for fostering lifelong learning habits that align with teachers' unique goals and contexts. Future research could build on this work by examining student perceptions of ChatGPT in language learning, as understanding students' views is crucial to comprehending the tool's broader ethical and pedagogical implications. Comparative studies across diverse educational settings would also be valuable, helping to assess how differences in resources and institutional contexts shape the affordances and limitations of ChatGPT. Such studies would provide a more comprehensive picture of ChatGPT's potential and challenges, ultimately contributing to a more effective and responsible integration of AI in language education.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interest.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTIONS

Dara Tafazoli: conceptualization; methodology; formal analysis; investigation; resources; writing - original draft; writing - review & editing; visualization; supervision; project administration.

Lee McCallum: conceptualization; methodology; writing - original draft; writing - review & editing; supervision.

REFERENCES

Al-khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. *Computers and Education: Artificial Intelligence, 6*, 100218. https://doi.org/10.1016/j.caeai.2024.100218

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *Teaching English with Technology, 23*(2), 25–43. https://doi.org/10.56297/BUKA4060/XHLD5365

Ary, D., Jacobs, L. C., Sorensen, C. K., & Walker, D. (2014). *Introduction to research in education*. Cengage Learning.

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English with Technology, 23*(1), 23-41. https://doi.org/10.56297/BKAM1691/WIEO1749

Borg, S. (2015). Overview - beyond the workshop: CPD for English language teachers. In S. Borg (Ed.), *Professional development for English language teachers: Perspectives from higher education in Turkey* (pp. 5–13). BC.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66*(4), 616-630. https://doi.org/10.1007/s11528-022-00715-y

Chiu, T. K. F., Moorhouse, L. M., Chai, C. S., & Ismailov, M. (2023). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environment*. https://doi.org/10.1080/10494820.2023.2172044

Compton, L. K. (2009). Preparing language teachers to teach language online: A look at skills, roles, and responsibilities. *Computer Assisted Language Learning, 22*(1), 73-99. https://doi.org/10.1080/09588220802613831

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Czarniawska, B. (2004). *Narratives in social science research*. SAGE.

Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology, 13*, 100060. https://doi.org/10.1016/j.jrt.2023.100060

Engwall, O., Lopes, J., Cumbal, R., Berndtson, G., Lindström, R., Ekman, P., Hartmanis, E., Jin. E., Johnston, E., Tahir, G., & Mekonnen, M. (2022). Learner and teacher perspectives on robot-led L2 conversation practice. *ReCALL, 34*(3), 344-359. https://doi.org/10.1017/S0958344022000027

Floris, F. D. (2021). Online continuing professional development. In H. Mohebbi., & C. Coombe (Eds.), *Research questions in language education and applied linguistics*. Springer. https://doi.org/10.1007/978-3-030-79143-8_115

Gamper, J., & Knapp, J. (2002). A review of intelligent CALL systems. *Computer Assisted Language Learning, 15*(4), 329-342. https://doi.org/10.1076/call.15.4.329.8270

Gao, Y., Wang, Q., & Wang, X. (2024). Exploring EFL university teachers' beliefs in integrating ChatGPT and other large language models in language education: a study in China. *Asia Pacific Journal of Education, 44*(1), 29–44. https://doi.org/10.1080/02188791.2024.2305173

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology, 27*(2), 6–27. https://hdl.handle.net/10125/73501

Han, T., & Sari, E. (2024). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning, 37*(4), 961-985. https://doi.org/10.1080/09588221.2022.2067179

Haristiani, N. (2019). Artificial intelligence (AI) chatbot as language learning medium: An inquiry. *Journal of Physics: Conference Series, 1387*(1), 012020. https://doi.org/10.1088/1742-6596/1387/1/012020

Heyder, A. (2019). Teachers' beliefs about the determinants of student achievement predict job satisfaction and stress. *Teaching and Teacher Education, 86*, 1–9. https://doi.org/10.1016/j.tate.2019.102926

Hubbard, P. (2023). Contextualizing and adapting teacher education and professional development. In D. Tafazoli & M. Picard (Eds.), *Handbook of CALL teacher education and professional development* (pp. 3-14). Springer. https://doi.org/10.1007/978-981-99-0514-0_1

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning – are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Inceoglu, S., Chen, W., & Lim, H. (2023). Assessment of L2 intelligibility: Comparing L1 listeners and automatic speech recognition. *ReCALL, 35*(1), 89-104. https://doi.org/10.1017/S0958344022000192

Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning, 36*(7), 1338-1364 https://doi.org/10.1080/09588221.2021.1987272

Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education, 55*(1), 48-63. https://doi.org/10.1080/15391523.2022.2142873

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Stadler, M. (2023). ChatGPT for good? On opportunities and challenges of Large Language Models for education. *Learning and Individual Differences, 103*(102274). https://doi.org/10.1016/j.lindif.2023.102274

Khazaie, S., & Ebadi, S. (2023). Exploring the feasibility of augmented reality game-supported flipped classrooms in reading comprehension of English for medical purposes. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2023.2173612

Kohnke, L. (2022) A pedagogical chatbot: A supplemental language learning tools. *RELC Journal*. https://doi.org/10.1177/00336882211067054

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 5*(4), 537-550. https://doi.org/10.1177/00336882231162868

Kostka, I., & Toncelli, R. (2023). Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations. *TESL-EJ, 27*(3). https://doi.org/10.55593/ej.27107int

Kuhail, M. A., Alturki, N., Alramlawi, S., Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies, 28*, 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Lamb, M., & Wyatt, M. (2019). Teacher motivation: The missing ingredient in teacher education. In S. Mann & S. Walsh (Eds.), *The Routledge handbook of English language teacher education* (pp. 522-535). Routledge.

Lichtman, M. (2013). *Qualitative research in education: A user's guide* (3rd ed.). SAGE.

Liu, G. L., & Wang, Y.(2024). Modeling EFL teachers' intention to integrate informal digital learning of English (IDLE) into the classroom using the theory of planned behavior. *System, 120*, 103193. https://doi.org/10.1016/j.system.2023.103193

Liu, G. L., Darvin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2024.2310288

McCallum, L. (2023). New takes on developing intercultural communicative competence: Using AI tools in telecollaboration task design and task completion. *Journal for Multicultural Education, 18*(1-2), 153-172. https://doi.org/10.1108/JME-06-2023-0043

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-282. https://pubmed.ncbi.nlm.nih.gov/23092060

Nami, F. (2023). CALL teacher education models, methods, and theoretical groundings: A systematic review of the studies in under-represented contexts. In: D. Tafazoli & M. Picard (Eds.), *Handbook of CALL teacher education and professional development* (pp. 483-503). Springer. https://doi.org/10.1007/978-981-99-0514-0_28

Orvik, A., Larun, L., Berland, A., & Ringsberg, K. C. (2013). Situational factors in focus group studies: A systematic review. *International Journal of Qualitative Methods, 12*(1), 338-358. https://doi.org/10.1177/160940691301200116

Rardin, J. (1977). The language teacher as facilitator. *TESOL Quarterly, 11*(4), 383–387. https://doi.org/10.2307/3585734

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Teaching & Learning, 6*(1), 342-362. https://doi.org/10.37074/jalt.2023.6.1.9

Salmons, J. (2014). *Qualitative online interviews: Strategies, design, and skills* (2nd ed.). SAGE.

Saville, N. (2023). The future of language learning in the era of ChatGPT. *Journal of e-Learning and Knowledge Society, 19*(3), 6-10. https://doi.org/10.20368/1971-8829/1135875

Schulze, M. (2008). Interfaces in Intelligent CALL. *Computer Assisted Language Learning, 21*(4), 301-303. https://doi.org/10.1080/09588220802393608

Stockwell, G. (2007). Vocabulary on the move: Investigating an intelligent mobile phone-based vocabulary tutor. *Computer Assisted Language Learning, 20*(4), 365-383. https://doi.org/10.1080/09588220701745817

Stockwell, G. (2009). Teacher education in CALL: Teaching teachers to educate themselves. *International Journal of Innovation in Language Learning and Teaching, 3*(1), 99-112. https://doi.org/10.1080/17501220802655524

Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education, 23*(1), 1-23. https://www.learntechlib.org/p/222408

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*, 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, Z., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112. https://doi.org/10.1111/bjet.13370

Yin, R. K. (2018). *Case study research: Design and methods*. SAGE.

Yuan, Y., Li, H. & Sawaengdist, A. (2024). The impact of ChatGPT on learners in English academic writing: opportunities and challenges in education. *Language Learning in Higher Education, 14*(1), 41-56. https://doi.org/10.1515/cercles-2023-0006

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

https://doi.org/10.17323/jle.2025.19075

Assessment of Intercultural Communicative Competence in High-Stakes Speaking Tests: The case of the Czech Republic

Tomáš Mach

Jan Evangelista Purkyně University, Czech Republic

ABSTRACT

Background: For a long time, concepts such as the intercultural communicative competence have challenged the view of culture as purely fact-based, objective content to be transmitted to students, from whom only retention is expected. Today, cultural knowledge plays virtually no role in international standardised English exams such as TOEFL or the Cambridge English Qualifications. It is unknown to which extent this is true for examinations at the national level, and the materials used to prepare students for these examinations.

Method: The study provides a quantitative account of exam topics in the speaking part of the final exam in Czech grammar schools. The sample consists of 206 grammar schools in the Czech Republic (58% of all such schools). The percentage of knowledge-oriented, culture-specific topics was calculated for each school. The study further analyses the content of four books used to prepare students for the exam. The sample comprises four frequently used preparatory books. The text in the chapters that deal with cultural topics is classified using Larzén's (2005) typology of culture.

Results: The study has demonstrated that approximately 50% of the final exam topics are focused on specific English-speaking countries, their geography, history and literature. A similar tendency, but significantly more pronounced, has been found in the preparatory materials. More than 95% of their content was heavily knowledge-oriented with no discernible difference between older and recently published materials. Virtually no evidence of a skill-based or intercultural approach to culture was found.

Conclusion: The study has confirmed what is often only an unfounded assumption by researchers and it has shown that the traditional factual view of culture is still immensely popular with no change in sight. Recommendations for material writers, teachers, and teacher educators are provided at the end of the study.

KEYWORDS

culture in language education; intercultural communicative competence (ICC); culture in high-stakes assessment; standardized English exams; content knowledge in language assessment; assessment of speaking

Citation: Mach, T. (2025). Assessment of intercultural communicative competence in high-stakes speaking tests: The case of the Czech Republic. *Journal of Language and Education, 11*(3), 98-109. https://doi.org/10.17323/jle.2025.19075

Correspondence:
Tomáš Mach,
tomas.mach@mail.muni.cz

Received: January 29, 2024
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

The role of culture within English language teaching (ELT) has been the subject of extensive discussion over the past two decades (Baker, 2009; Bayyurt, 2006; Byrd et al., 2011; Larzén, 2005; Murphy, 1988; Paige et al., 1999; Zotzmann, 2015). As the field has shifted from a focus on communicative competence to intercultural communicative competence (ICC), and as English has increasingly functioned as a lingua franca (ELF), new challenges have emerged in conceptualising, eliciting, and evaluating the cultural dimension of language proficiency (Liddicoat & Scarino, 2010; Scarino, 2017). While a wide array of theoretical models and pedagogical approaches to ICC have been proposed, its role in language assessment remains relatively underexplored, especially in large-scale or high-stakes testing contexts. Historically, cultural content has been embedded in

both curricula and assessment practices, although some international standardised tests have sought to reduce cultural bias (e.g., Weir, 2005). Whether similar developments have taken place in national testing systems, however, is less well understood.

This study addresses this gap by examining the representation of cultural content in high-stakes English language assessment, focusing on the Czech maturita exam as a representative example. Despite growing curricular emphasis on ICC, there is limited empirical research on how cultural knowledge is reflected in assessment tasks and preparatory materials. In particular, little is known about how far assessment practices align with curricular aims, or about the extent to which they prioritise knowledge-based, attitudinal, or behavioural dimensions of culture. To explore these questions, the study investigates (1) to what extent knowledge-oriented cultural topics are represented in the maturita speaking exam, and (2) how knowledge-oriented the preparatory materials for the exam are. The study applies Larzén's (2005) typology as an analytical framework to examine a corpus of exam topics and commercially available textbooks. In doing so, it offers a theoretically grounded, data-driven contribution to current debates about culture and assessment in language education.

LITERATURE REVIEW

Defining Culture in Language Education

To understand how culture affects assessment, it is necessary to define what culture represents. Traditional views often describe cultures as static artifacts tied to specific regions (Canale, 2016; Murphy, 1988; Paige et al., 1999), enabling generalisations that reinforce us–them dichotomies. Yet cultures are not synonymous with superficial knowledge of language communities. As Baker (2022, p. 289) notes, "cultures are heterogeneous and cultural characterisations are contestable." They change over time and are continuously interpreted and co-constructed. For English in particular, there is no single English-speaking culture; not even among native speakers, long viewed as custodians of the language (Llurda & Calvet-Terré, 2022; Seidlhofer, 2005). ELF further perspectives challenge this view, positioning native speakers as participants in global culture, not its curators.

Many newer definitions of culture seek to move beyond static models but often remain vague. Importantly, how culture is conceptualised, whether as fixed knowledge or as a dynamic communicative skill, has direct consequences for assessment. Static views lend themselves to factual testing, whereas processual approaches require more complex and authentic formats that are harder to implement, especially in high-stakes contexts.

Teachers' Conceptions and Classroom Practices

Byram's (1997) influential model of intercultural communicative competence (ICC) identifies knowledge, skills, and attitudes as core components. Larzén (2005) links these to classroom practice, showing how each shapes goals and teacher–student roles (see Table 1). In practice, however, it is factual knowledge, especially about target cultures, that dominates (Johnson & Rinvolucri, 2010, p. 15; Kramsch, 1995, p. 89; Scarino, 2017, p. 16; Sercu, 2002, p. 62). Although Byram (2014) and Gu & Zhao (2021) report shifts towards more complex assessment (e.g., in China, Bulgaria, Argentina), gaps exist elsewhere, likely due to insufficient teacher training or discrepancies between policy and practice (Safa & Tofighi, 2021).

Table 1

Teachers' Conceptions of ICC in TEFL

	Cognitive orientation	Action orientation	Affective orientation
	Quantitative view	**Qualitative view**	
WHAT? (Teachers' conceptions of «culture»)	Factual knowledge	Skills	Bi-directional perspective
WHY? (Teachers' beliefs about cultural objectives)	Providing general background information	Preparing for future intercultural encounters	Promoting tolerance and empathy
HOW? (Teachers' classroom practice)	Pedagogy of Information Teacher in centre	Pedagogy of Preparation Teacher and pupil in centre	Pedagogy of Encounter Pupil in centre

Note. Adapted from Larzén (2005).

Empirical studies across national contexts further illustrate that teachers' conceptualisations of culture correspond to this fact-based instructional focus. Decke-Cornill (2003), for example, found that teachers at German *Gymnasien* resisted ELF-oriented perspectives and instead focused on conveying information about specific English-speaking countries. One participant emphasized that "[w]hat they [students] expect from us, at school, is cultural knowledge. They want information about the specific countries" (p. 63). In Finland, similar attitudes led Larzén (2005) to conclude that teachers viewed culture as a static product to be transmitted. Roughly half of the participants specifically described culture explicitly in terms of "facts about the history, geography, religion and political conditions of the English-speaking countries" (p. 103). Zerzová (2012) reported nearly identical views among lower-secondary teachers in the Czech Republic. Bayyurt's (2006) study added further nuance, noting that while some teachers valued local or international cultural references, others excluded culture altogether unless it clearly supported linguistic objectives.

These knowledge-oriented beliefs are also reflected in classroom practice. In the UK, Driscoll, Earl, and Cable (2013) found that cultural instruction was typically treated as an *ad hoc* supplement, with most activities focused on factual knowledge. Larzén (2005) and Zerzová (2012) both observed that approximately 90% of instructional time devoted to culture was spent transmitting factual content. Safa & Tofighi (2021) compared teachers' beliefs about ICC with their implementation, and found that "the implementation of ICC in their pedagogical practices was evidently quite poor and far from their stated beliefs" (p. 170). These findings point to the widespread dominance of what has been termed the "Pedagogy of Information" – a model in which culture is presented as a body of knowledge rather than a set of communicative skills. This over-reliance on the factual approach to cultural instruction, despite the availability of alternative models, warrants an examination of the role of culture in testing and assessment, and learning materials. The following section turns to an analysis of teaching materials as another domain where cultural ideologies are embedded and transmitted.

Textbooks as Vehicles of Cultural Ideology

The dominance of the Pedagogy of Information approach in language instruction is further supported by analyses of learning materials, which primarily focus on expanding factual knowledge about specific English-speaking countries (Lee 2009; Rodríguez & Espinar 2015), often without deeper exploration or consideration of an ELF intercultural perspective. Byrd, Cummings, Watzke, and Montes (2011) attribute this deficiency to university teacher-training programs, which, while often covering cultural content in their curricula, rarely provide training specifically focused on how to teach that content effectively. This lack of hands-on teaching preparation further widens the existing gap between research and classroom practice and may ultimately lead teachers to shy away from addressing cultural topics altogether.

Turning to the materials themselves, Canale (2016) analysed foreign language textbooks across six languages and concluded that:

> a common pattern of homogenisation which – almost paradoxically – can be achieved by means of adopting politics of exclusion or inclusion of cultural diversity and heterogeneity. This homogenisation manifests itself both in verbal and non-verbal language, and also in the activities and exercises surrounding the text itself. (p. 239)

Kang-Young's (2009) analysis of EFL textbooks revealed that general culture, i.e. culture not specific to a particular country, as well as the subjective (small-c) culture, are consistently neglected. Instead, preference was given to what could be memorised (arts, history, and customs) with a strong bias towards facts about the USA. Shin, Eslami, & Chen (2011) examined seven EFL textbook series, distinguishing between knowledge-oriented and communication-oriented culture. The knowledge-oriented culture accounted for the average of 83.1%, with 58.7% and 100% as the minimum and maximum respectively. Since teachers generally display reluctance to go against the textbook (Forman, 2014), inferences about their classroom practice can, among other things, be made based on the materials they use.

These studies collectively reveal a pattern in which textbooks tend to present culture as a static body of factual knowledge, rather than as a dynamic set of communicative skills. This representation aligns closely with the Pedagogy of Information conception and serves as evidence for the limited intercultural focus in many ELT materials. Moreover, given teachers' reliance on textbooks, the predominance of static cultural content likely influences both classroom practices and the nature of culture-related assessment tasks.

Culture in Language Assessment and ICC Testing

Compared to teaching, relatively little is known about how culture should be assessed – or whether it should be assessed at all (Scarino, 2017; Sercu, 2010). As in teaching, two core questions arise: what should be assessed and how. However, unlike teaching, where cultural content is often expected, in assessment the very inclusion of culture remains contested. Zotzmann (2015), for example, argues that intercultural learning is an embodied, value-laden, and unpredictable process that "certainly [does] not lend [itself] to assessment" (p. 186). Borghetti (2017, p. 10) echoes this, likening interculturality to values such as honesty or generosity – qualities that should be promoted but not necessarily tested.

These concerns go beyond the familiar objection that ICC is too latent to be measured (Schauer, 2020). Here, ICC is

framed not just as a complex construct but as a value, and thus fundamentally unsuited to assessment. While such positions are in the minority, they highlight key tensions. Even advocates of ICC assessment like Deardorff (2009) stress of ICC assessment like Deardorff (2009) stress that "assessment is ultimately about learning" (p. 490). This reflects a broader shift from cognitive (psychometric) to sociocultural views of assessment (Liddicoat & Scarino, 2012), prompting renewed attention to both *what* and *how* to assess.

Content recommendations often mirror ICC dimensions, and while assessing all components is desirable (Sercu, 2010), no comprehensive measure exists. As Schauer (2020) notes, holistic assessment may not even be feasible in compulsory education. This makes it likely that knowledge will be prioritized as it is the easiest to assess and historically linked to cultural instruction (Byram, 1997; Scarino, 2010, 2017). But this raises dilemmas. As Larzén (2005) and Sercu (2010) point out, choosing which cultural facts to assess is complex and depends on teacher preferences and learner needs. Given the interdependence of teaching and assessment (McNamara, 2000), cultural knowledge tied to history, geography, and literature will likely dominate. Additional factors, such as resistance to ELF benchmarks (Ke, 2018) or treating Anglo-centric norms as "international" (Jenkins & Leung, 2017, p. 114) can further strengthen this bias.

As for *how* culture is assessed, alternative approaches aim to capture more dimensions of ICC. These include reflective diaries, self-assessment portfolios, role-plays, group discussions, and other tasks that foster learning and provide insight into development (Byram, 1997; Liu, 2021; Schauer, 2020). Yet testing cultural knowledge in isolation remains problematic. First, without support from other task types, such tests are one-sided, reducing assessment to mere factual recall. Second, they suffer from a complication shared by all approaches that seek to integrate language and content – construct irrelevant variance, where "the trait being assessed…is confounded with the irrelevant requirement of having knowledge of a particular topic" (McNamara, 2000, p. 53). Llosa (2017, p. 4) clarifies that construct irrelevant variance depends on whether the focal point of assessment is content or language:

> From the perspective of content assessment, language has typically been considered a source of construct-irrelevant variance – variance in scores that is not related to the construct being assessed. From the perspective of language assessment, content (also referred to as topical knowledge or background knowledge) has also been considered a potential source of construct-irrelevant variance. (p. 4)

If the primary goal of language education is to develop ICC or communicative competence, with both framed around language use, assessment should reflect this. Yet, as Fantini (2009) observes, this alignment between goals, instruction, and assessment is often absent (p. 460). Problems with validity and reliability compound the issue (Sercu, 2004). For instance, if a course targets communication skills, assessing cultural facts may undermine content validity. Likewise, if a test rewards memorized knowledge, its construct validity is compromised. Reliability suffers too when test objectives are unclear, e.g., one version emphasizes interaction, and another factual knowledge. These concerns are especially relevant in high-stakes exams, where washback can reinforce outdated, fact-based models of cultural instruction.

Research Gap

Despite growing theoretical interest in assessing ICC and the development of well-grounded recommendations, significant gaps remain regarding the implementation of these ideas in practice. On one hand, the relatively recent interest in assessing ICC, and, more importantly, offering theoretically grounded recommendations for doing so, is without a doubt a positive development. On the other hand, it remains unclear if and how these recommendations materialise in practice.

Research on the assessment of ICC in national speaking exams is scarce, especially regarding the status of cultural content in compulsory high-stakes speaking tests and how well exam content aligns with stated curricular goals relevant to ICC. This study addresses these gaps by empirically investigating the presence and nature of cultural content in Czech national high-stakes speaking tests and preparatory materials, and how this reflects curricular objectives. Similarly, teachers can hardly be expected to follow and act upon research findings (McIntyre, 2005; Vanderlinde & van Braak, 2010; Willingham & Daniel, 2021[1]). This study thus aims to make the first step, at least in the Czech context. Specifically, its purpose it to investigate the current state of compulsory high-stakes examinations because:

(1) These exams strongly influence the educational content leading up to them, as teachers and institutions often align their curriculum and instruction to prepare students for the test. Examining exam content thus provides critical insights into the cultural priorities that shape classroom teaching and learning.
(2) If the study reveals any inconsistencies or misalignments in how culture is addressed, implementing corrective changes in the exams could have a widespread impact. This would in turn benefit a large number of students who rely on these exams as a measure of their readiness and cultural competence in English.

[1] Willingham, D. T., & Daniel, D. B. (2021). *Making education research relevant: How researchers can give teachers more choices. Education Next, 21*(2), 28-33.

METHOD

Context of the Study

The study focuses on the secondary/high-school leaving exam in the Czech Republic called *maturitní zkouška* or *maturita*, for short. The scope has been narrowed down to grammar schools that are academically oriented (*gymnasium*) and aim to prepare their students to study at a university. This excludes a field-specific orientation of the final exam, which is often the case with other more vocation-oriented secondary schools. When taking the exam, students are 18-19 years old, and they have completed nine years of mandatory primary education and four years of grammar school. For the exam, four subjects are typically chosen by every student, with Czech language being compulsory for everyone. Presently (2024), students further have to opt for either Mathematics or a foreign language. They choose the two remaining subjects based on their interests.

Both the educational content and the educational goals of secondary education, and by extension the corresponding final exams, are regulated by the Framework Education Programme (MEYS, 2021). Much like the Common European Framework of Reference for Languages (CEFR), this curricular document specifies what a successful student should be able to do. Apart from this, it also provides information on the content to be covered during the four years of study. While cultural content is indeed mentioned as one of many areas to be addressed, it is not part of any of the expected learning outcomes listed. Instead, the following references to ICC can, among other things, be found as the goal of the area of *language and communication*:

(1) mastering the basic rules of interpersonal communication in a given cultural environment and respecting them;
(2) forming a general overview of the social and historical development of human society, which aids in learning respect and tolerance for the different cultural values of various language communities[2]

In line with Fantini's (2009) plea for the alignment of education goals, their implementation, and assessment, it seems sensible that the overarching goals (competence acquisition) as well as the expected outcomes should also be the object of assessment.

As for the final exam itself, it is divided into the so-called *state* and *profile* part. For English (or any other foreign language), this means that the state is responsible for creating a standardised test that every student regardless of school type is required to take. The test is set approximately at a B1 level and comprises tasks in the areas of listening, reading and use of English. This test is only graded pass or fail. The rest of the exam is currently administered by the school. The law stipulates that there has to be a writing task with an output of min. 200 words, and that the speaking part (which carries 60 % of the mark) should take max. 15 minutes. 20-30 topics are to be prepared by the school for the purposes of the speaking exam and a student chooses one of them at random. The speaking exam can be divided into more parts (monologic, dialogic, with/without visual aids etc.) as long as these are confined to the selected topic. For the *gymnasium* school type, the target level for these two parts is B2. The writing as well as the speaking sections are designed by teachers, which means that both the form and the content of the examination may differ across schools.

Research Questions

The prevalence of a cognitive, knowledge-based approach to culture is often mentioned in various studies (see above), but mostly no reference is made to any data supporting the view that this is indeed a common practice. Very little is also known about the national assessment policies with respect to ICC despite the number of students affected. Therefore, a quantitative account of what is assessed is needed. The following research questions have been formulated:

RQ 1: To what extent are knowledge-oriented cultural topics represented in the speaking part of the Czech *maturita* exam?

RQ 2: How knowledge-oriented are the materials that help students prepare for the Czech *maturita* exam?

Data Collection

In order to answer the first question, I examined the lists of topics for individual schools. In December 2022, I accessed the websites of 250 Czech grammar schools and succeeded in finding the lists in 206 cases. This was deemed representative enough given that there were 355 such schools in the country. Thus, approximately 58 per cent are covered with proportional representation across all regions. If the school had different streams (general, CLIL, bilingual), which also happened to differ with respect to the exam topics, the list for the general stream was taken. I tallied the topics for each school and determined how many of them were focused on specific cultural topics, which are commonly associated with the necessity to memorise facts. These topics included for example *the UK*, *the USA*, *Commonwealth countries*, *British lit-*

[2] MEYS - The Ministry of Education, Youth and Sports [Ministerstvo školství, mládeže a tělovýchovy České republiky]. (2021). Framework education programme for secondary general education (grammar schools) [Rámcový vzdělávací program pro gymnázia]. Retrieved from edu.cz: https://www.edu.cz/rvp-ramcove-vzdelavaci-programy/ramcove-vzdelavaci-programy-pro-gymnazia-rvp-g/

erature, *Czech Republic*, *Prague* etc. To control for the variety in the number of topics, the percentage of these topics was counted for each school. Some schools had made the corresponding materials publicly available, which I used to confirm that the content was indeed knowledge-oriented.

Although the first research question already provides insights into the exam focus, I also decided to analyse the materials marketed as preparatory for the Czech *maturita* exam. Not only are these commonly used by and recommended to students, but they are also often utilised as supplementary materials in schools. Unlike traditional textbooks, these materials are structured specifically around commonly tested topics, prioritising certain language aspects (mainly vocabulary) while neglecting others (such as grammar).

Several criteria were used to guide the selection of the maturita preparation books. First, only materials explicitly designed and marketed as dedicated exam-preparation tools were included. While there are other books with the word maturita in the title, many of these function primarily as comprehensive coursebooks with only a subset of exercises linked to the exam. Such titles were excluded, since their primary aim is general language instruction rather than systematic exam preparation. By contrast, the selected books (published in 2005, 2008, 2014, and 2020) were explicitly designed as exam-focused materials, and their distribution across different publication years makes it possible to observe developments in how maturita preparation has been approached over time. Second, the chosen titles represent the main publishers active on the Czech market, ensuring coverage of the principal options available to teachers and learners. Third, they are regularly recommended on the official websites of surveyed schools, appear in school libraries, and are cited in lists of required or supplementary reading, which demonstrates their adoption in practice. Finally, the limited number of specialized publications means that the list presented here covers the key materials available at the time of data collection and can thus be regarded as representative. A limitation of this selection is that newer materials published after the period of data collection may not yet have been widely adopted and therefore are not included. The following four books were included:

Book 1: Mrákota, J. (2005). *Maturitní otázky z angličtiny* [*Maturita questions for English*]. Jiří Mrákota – vydavatelství jazykové literatury.
Book 2: Smith-Dluhá, G. (2008). *Angličtina - otázky a odpovědi nejen k maturitě* [*English – question and answers not just for maturita*]. Infoa.
Book 3: El-Hmoudová, D. (2014). *Angličtina - edice Maturita* [*English – the maturita edition*]. Vyuka.cz.
Book 4: Baláková, L., Baron, U., Hamzová, A., Knápek, J., Pešková, D. (2020) *Maturita z anglického jazyka – Maturitní témata* [*Maturita in the English language – maturita topics*]. Didaktis.

Analytical Framework and Procedures

I searched the publications for cultural topics that appeared in at least three of them. Five such topics were identified: the United States of America, the United Kingdom, the Czech Republic, (other) English-speaking countries, and literature (of the English-speaking countries). These topics were chosen because they directly reference specific cultural contexts, making it easier to reliably identify cultural content – even if it is implicit. Broader categories, such as education or technology could indeed involve cultural aspects, but these vary greatly in emphasis and would require subjective interpretation to confirm their cultural relevance. By focusing on topics tied to distinct cultural settings, this approach ensures consistency in identifying culturally relevant material. Using Larzén's (2005) typology outlined above, I aimed to establish how much of the text on the corresponding pages is related to factual knowledge (Pedagogy of Information), skills (Pedagogy of Preparation), and the bi-directional perspective (Pedagogy of Encounter). After initial probing, looking for the last two categories seemed a moot point, as no such instances were present. This conclusion was confirmed by two colleagues specialising in applied linguistics and ELT methodology.

With Pedagogy of Information being the dominant conception in the materials, I instead turned my attention to the nature of factual knowledge; Larzén (2005) offers four subcategories:

(1) realia
(2) cultural products
(3) traditions and ways of life
(4) modes of thought

Realia refers to "facts about the history, geography, religion and political conditions of the English-speaking countries" (p. 103). The second category is represented by art, literature, film and music associated with the English-speaking world. Examples of the third category include, for example Christmas, Easter, and Halloween but also practices related to family life, work life, and leisure. Lastly, modes of thought "can be described in terms of the values, norms and beliefs underlying the way people live and act" (p. 105). The instances in these categories were counted for each book and topic. The counting proved challenging because there was rarely a one-to-one correspondence between a sentence and an idea. As I did not want to artificially inflate the numbers, I approached the counting relatively conservatively and decided to err on the side of caution; as shown in c) below, longer lists of items were counted as only two pieces of information. A couple of examples:

(a) Canterbury: The spiritual centre of the Church of England, the Cathedral was built between 1000s and the early 1500s. It was the destination of travellers in *The Canterbury tales* by Geoffrey Chaucer.

i. The spiritual centre of the Church of England was the Cathedral.
ii. The Cathedral was built between 1000s and the early 1500s
iii. It was the destination of travellers in *The Canterbury tales* by Geoffrey Chaucer.

(3 pieces of information)

(b) Legislature: Congress consists of two houses: the 100-member Senate, elected for a term of 6 years; the 435-member House of Representatives, elected for a term of 2 years.

i. Congress consists of two houses.
ii. The Senate has 100 members.
iii. Senators are elected for a term of 6 years.
iv. The House of Representatives has 435 members.
v. Th Members of the House of Representatives are elected for a term of 2 years.

(3 pieces of information)

(c) Important products: Agriculture – wool, beef, cattle horse breeding, potatoes, barley, sugar beet, wheat, oats.

i. [a list of words or short phrases]

(2 pieces of information)

Reliability and Validation

In order to enhance the reliability of the categorisation process, the coding scheme was first piloted on a small sample of texts and refined through discussion. A coding guide was created to ensure consistent application of Larzén's categories across the full dataset. To further validate the approach, approximately 20 % of the textbook content was independently coded by one of the two colleagues. Coding decisions were then compared, and differences in interpretation were discussed until consensus was reached. While formal intercoder reliability statistics were not calculated, this process of double coding and negotiation helped to reduce subjectivity and clarify ambiguous cases. Although some discrepancies emerged, these were mostly limited to the precise quantification of realia. For example, one coder counted 85 instances, while the other counted 93, for the USA topic in Book 1. In a few cases, the colleague also expressed scepticism about whether examples from the other three subcategories were present at all.

Even though the quantitative approach taken here allows for a systematic scrutiny of the cultural content in exam topics and preparatory materials, it cannot capture how teachers interpret or implement the materials in practice. The study as such does not endeavour to account for classroom dynamics or teacher perspectives, which could add more context and depth. These aspects are further discussed in Section 5.5.

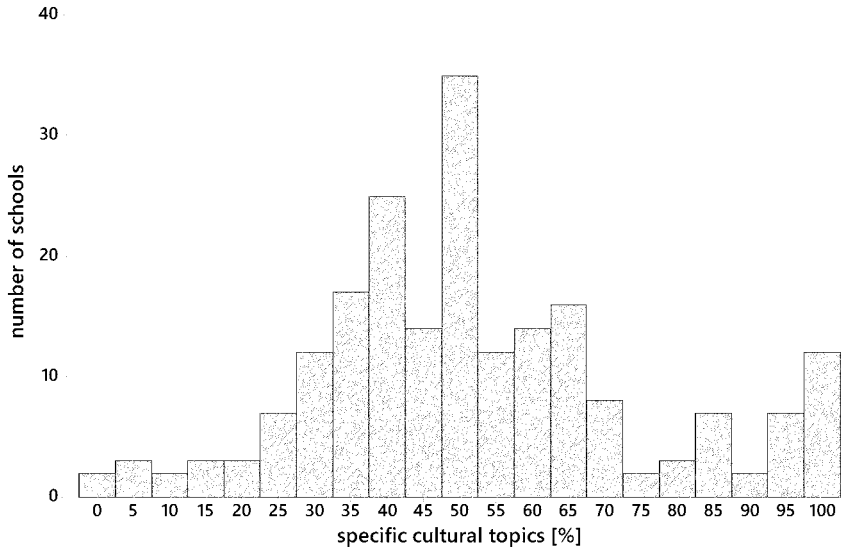
RESULTS

Cultural Topics Account for About Half of the Final Exam

Figure 1 below shows the relative proportion of cultural topics in the final exam across Czech grammar schools.

Figure 1

Proportion of Cultural Topics in the Final Exam in Czech Grammar Schools (n=206)



The mean proportion of cultural topics was 52.33% (SD = 22.45), with a median of 50%. This indicates that, on average, approximately half of the final speaking exam topics are cultural in nature. However, substantial variation was observed between schools. Some institutions included only culture-focused topics such as *The Politics of Australia, New Zealand, and Canada*, *American Literature after Romanticism*, *US Political System*, or *Early History of the USA*. In contrast, other schools included few or no explicitly cultural topics in their final exams, or integrated cultural content more implicitly within broader themes.

A Kruskal-Wallis test was conducted to examine whether the proportion of cultural content in final exams differed across the 14 Czech regions. The test indicated no significant differences (H = 13.06, df = 13, p = 0.443; adjusted for ties: H = 13.10, p = 0.440), suggesting that regional variation in exam content is negligible. Although the median proportions varied slightly between regions (from 35% to 59%), these differences were not statistically significant.

Cultural Content in Preparatory Materials Is Predominantly Factual and Realia-Based

Table 2 below presents the frequency of cultural content in four commonly used preparatory books, categorised by topic.

In this table, the "realia" category includes factual content related to history, politics, geography, and culture. The remaining categories of cultural products, traditions and ways of life, and modes of thought are grouped under "other" due to their low frequency. For example, cultural content related to literature was typically presented in a factual or historical style. The following excerpt from Book 3 is a case in point:

> Much American writing in the 1600s and 1700s focused on religious issues linked to the challenges of such life. During the 1700s some writers turned to political issues. Essays like Thomas Paine's pamphlet *Common Sense* (1776) encouraged Americans to seek independence from Britain. (Book 3)

This pattern of presentation was consistent across the materials surveyed.

DISCUSSION

The results paint a clear picture: on average, approximately half of the final exam topics in Czech grammar schools explicitly test cultural content. This cultural component, though nominally aligned with communicative goals, is overwhelmingly factual in nature. The preparatory materials reinforce this pattern, focusing heavily on realia such as political systems, historical events, and biographical details about authors. Literature, which might be expected to offer more affective cultural engagement, is presented almost entirely through factual summaries.

The findings also invite reflection on the theoretical framework guiding the analysis. Larzén's (2005) typology was originally developed to categorise teachers' conceptions of ICC. While the typology offers valuable conceptual clarity regarding different pedagogical orientations (cognitive, action, and affective), it was not designed with direct categorisation of assessment artefacts in mind. It follows from this that applying the model to exam tasks and textbooks involved inference. Furthermore, in practice, elements of different orientations may often appear in blended forms, even

Table 2

Amount of Information about Culture in Preparatory Materials for the Final Exam

		USA	UK	CR	English-speaking countries	Literature
Book 1						
	realia	93	109	134	91	116
	other	0	0	0	1	0
Book 2						
	realia	190	265	110	166	160
	other	0	0	0	0	0
Book 3						
	realia	163	104	192	130	80
	other	4	7	11	11	4
Book 4						
	realia	150	152	178	111	-
	other	3	8	6	0	-

within individual tasks, although this was rarely the case in the data. Acknowledging this hybridity offers a potential refinement of the typology as it underscores the need to move beyond ideal-type distinctions and at the same time consider how educators' conceptions may co-exist, interact, and shape real-world educational materials. Overall, Larzén's typology provides a useful lens for analysis. Its use here simultaneously points to the need for caution when applying belief-oriented models to real-life educational materials, especially when these may further be shaped by institutional pressures and constraints, which the individual-driven model can account for only to a limited extent.

The findings overall corroborate Scarino's (2017) observation that culture in language teaching is still often treated as a repository of specific and factual knowledge. They also extend Zerzová's (2012) work by showing that such tendencies persist well beyond lower secondary education and are now structurally embedded at multiple levels of the Czech school system. Decke-Cornill's (2003) account of similar patterns in other countries adds further weight to the claim that this is not merely a local anomaly but part of a broader issue in how language education conceives of and assesses culture.

The internal variation between schools is notable. Some institutions test almost exclusively knowledge-based content (e.g., "US political system," "American literature after Romanticism"), while others appear to downplay or even avoid such tasks altogether. This inconsistency suggests that teacher beliefs and institutional culture may play a larger role than formal curricula, a possibility that deserves further inquiry.

Implications

It is remarkable – and troubling at the same time – that such practices persist in high-stakes exit examinations, where one would expect adherence to research-based standards in assessment and ICC. Despite Scarino's (2017) argument that "teachers' assessment practices are heavily constrained by the requirements of the education systems in which they work" (p. 27), the Ministry's curricular documents do in fact allow for considerable teacher autonomy, offering guidelines rather than prescriptive rules. Yet this flexibility appears underutilised. Rather than capitalise on this space for pedagogical innovation, many teachers continue to rely on traditional, information-heavy approaches, perhaps out of inertia, or perhaps due to deeper institutional pressures.

The implications are serious. A narrow interpretation of culture undermines not just the spirit of ICC but – perhaps more importantly – the validity, reliability, and authenticity of the final exam itself. If (intercultural) communicative competence is to be meaningfully assessed, then culture must be approached not as inert content but as something to be interpreted, negotiated, and situated in real-world interaction.

Limitations

The limitations of the study stem mainly from the sampling. First, the study focuses exclusively on the situation in the Czech Republic and, although the discovered tendencies may be similar to those in other countries, the results cannot be generalised. Second, the study was narrowed down to gymnasium-type schools (secondary/further education), which are academically oriented and do not offer vocational specialisations. For other schools, the level and format of the final exam differ; specifically, the exam topics must reflect the particular focus of the programme (business, gastronomy, construction, etc.).

Equally important is the fact that this study relies solely on document analysis, specifically exam topics and preparatory materials. While this method is effective for identifying what is being assessed and how, it does not provide insight into why teachers and schools design their assessment in these ways. The absence of qualitative data, such as teacher interviews or classroom observations, limits the interpretative depth of the findings and makes it difficult to fully understand the beliefs, constraints, or institutional pressures that may influence assessment practices. To address this, a qualitative follow-up study would be beneficial. Investigating teachers' motivations and conceptions of culture and assessment through interviews, focus groups, or classroom-based research could enrich the present findings and provide a stronger foundation for possible interventions. By understanding the rationale behind current practices, the chances of meaningful change and alignment with ICC-informed assessment principles could increase.

CONCLUSION

With intercultural communicative competence (ICC) becoming increasingly central in English language teaching (ELT), there is a need to consider how its development is reflected in assessment practices. This study analysed the final exam topics of 206 Czech grammar schools, alongside the cultural content found in preparatory materials. The findings confirm what many earlier studies have suggested anecdotally: the pedagogical focus on factual knowledge remains dominant. Despite the rhetoric of communicative and intercultural aims in curricular documents, there is still approximately a 50:50 chance that the most consequential component of the final exam – its speaking part – will assess students' factual knowledge about an English-speaking country, rather than their ability to communicate meaningfully across cultures.

This tendency is mirrored in the preparatory materials. Four widely used books were analysed, and their content was found to rely almost exclusively on factual, country-specific knowledge. Literature, for instance, was largely presented through historical summaries and biographical detail rather than through interpretive engagement. These findings

suggest that the testing of culture rests on a narrow and static conception of knowledge, effectively leaving ICC out of assessment. This has two significant consequences for assessment quality. First, it undermines construct validity: if the intended construct is intercultural communicative competence, then tests that reward factual recall misrepresent what they purport to assess. Second, it creates problematic washback effects. When high-stakes exams emphasise memorisation, teachers and learners are likely to prioritise factual knowledge over communicative and intercultural skills in the classroom, regardless of curricular intentions.

The results carry the following implications for different stakeholders in the educational landscape: Teachers should critically reflect on the content and methods they use for assessment. In high-stakes tests, they should avoid rote memorisation of cultural knowledge and instead aim to assess students' communicative abilities in diverse contexts.

Textbook authors and material designers are in a position to re-introduce ICC into preparatory materials. Moving beyond encyclopaedic content, they can develop resources that foster intercultural awareness, empathy, and negotiation skills through open-ended questions, critical comparisons, and situated dialogues.

Policy-makers and national exam designers must ensure coherence between curricular documents and actual assessment.

Teacher educators play a crucial role in shaping future teachers' conceptualisations of culture and assessment. They should prioritise critical engagement with culture and equip their students with the skills to assess ICC meaningfully.

While this study focuses on the Czech Republic, the challenges identified extend to other European countries, as documented in prior research from Germany, Finland, Spain, for instance. What sets this study apart is its large-scale empirical evidence base, demonstrating the dominance of factual cultural knowledge in both high-stakes testing and preparatory materials. The Czech case thus serves as both a cautionary example and a point of reflection for other systems that strive to genuinely embed ICC.

Future research should focus on understanding the interplay of factors that influence the assessment of intercultural communicative competence. As these are likely to differ across countries, data from other educational contexts and comparative studies across different countries could highlight shared obstacles and/or provide examples effective practices. Such work is essential to guide the development of policy, teacher education, and instructional materials that support ICC in language assessment, while upholding construct validity and reducing harmful washback.

DECLARATION OF COMPETING INTEREST

None declared.

REFERENCES

Baker, W. (2009). The cultures of English as a lingua franca. *TESOL Quarterly*, *43*(4), 567-592. http://dx.doi.org/10.1002/j.1545-7249.2009.tb00187.x

Baker, W. (2022). From intercultural to transcultural communication. *Language and Intercultural Communication*, *22*(3), 280-293. http://dx.doi.org/10.1080/14708477.2021.2001477

Bayyurt, Y. (2006). Non-native English language teachers' perspective on culture in English as a Foreign Language classrooms. *Teacher Development, 10*(2), 233-247. http://dx.doi.org/10.1080/13664530600773366

Borghetti, C. (2017). Is there really a need for assessing intercultural competence? Some ethical issues. *Journal of Intercultural Communication*, *17*(2), 1-15. http://dx.doi.org/10.36923/jicc.v17i2.736

Byram, M. (1997). *Teaching and assessing intercultural communicative competence.* Multilingual matters.

Byram, M. (2014). Twenty-five years on – from cultural studies to intercultural citizenship. *Language, Culture and Curriculum*, *27*(3), 209-225. http://dx.doi.org/10.1080/07908318.2014.974329

Byrd, D., Cummings, A. H., Watzke, J., & Montes, M. F. (2011). An examination of culture knowledge: A study of L2 teachers' and teacher educators' beliefs and practices. *Foreign Language Annals*, *44*(1), 4-39.

Canale, G. (2016). (Re)Searching culture in foreign language textbooks, or the politics of hide and seek. *Culture and Curriculum*, *29*(2), 225-243. http://dx.doi.org/10.1080/07908318.2016.1144764

Deardorff, D. K. (2009). Implementing intercultural competence assessment. In D. K. Deardorff (Ed.), *The SAGE handbook of intercultural competence* (pp. 477-491). SAGE. http://dx.doi.org/10.4135/9781071872987

Decke-Cornill, H. (2003). 'We would have to invent the language we are supposed to teach': The issue of English as lingua franca in language education in Germany. In M. Byram & P. Grundy (Eds.), *Context and culture in language teaching and learning* (pp. 59-71). Multilingual matters.

Driscoll, P., Earl, J., & Cable, C. (2013). The role and nature of the cultural dimension in primary modern languages. *Language Culture and Curriculum*, *26*(2), 146-160. https://doi.org/10.1080/07908318.2013.799675

Fantini, A. E. (2009). Assessing intercultural competence: Issues and tools. In D. K. Deardorff, *The SAGE Handbook of Intercultural Competence* (pp. 456-476). SAGE.

Forman, R. (2014). How local teachers respond to the culture and language of a global English as a Foreign Language textbook. *Language, Culture and Curriculum*, *27*(1), 72-88. http://dx.doi.org/10.1080/07908318.2013.868473

Gu, X., & Zhao, Y. (2021). Integration of Teaching and Assessing Intercultural Communicative Competence. *Chinese Journal of Applied Linguistics, 44*(2), 241–258. https://doi.org/10.1515/CJAL-2021-0014

Jenkins, J., & Leung, C. (2017). Assessing English as a lingua franca. In E. Shohamy, I. G. Or & S. May (Eds.), *Language testing and assessment* (pp. 103-117). Springer.

Johnson, G., & Rinvolucri, M. (2010). *Culture in our classrooms: Teaching language through cultural content.* Delta publishing.

Ke, I.-C. (2018). Has English been increasingly tested as an international language? Evidence from 1956–2016. *Language, Culture and Curriculum*, *32*(2), 191-206. http://dx.doi.org/10.1080/07908318.2018.1512609

Kramsch, C. (1995). The cultural component of language teaching. *Language, Culture and Curriculum, 8*(2), 83-92. http://dx.doi.org/10.1080/07908319509525192

Larzén, E. (2005). *In pursuit of an intercultural dimension in EFL-teaching: Exploring cognitions among Finland-Swedish comprehensive school teachers.* Åbo akademi university press.

Lee, K.-Y. (2009). Treating culture: What 11 high school EFL conversation textbooks in South Korea do. *English Teaching: Practice and Critique*, *8*(1), 76-96.

Liddicoat, A., & Scarino, A. (2010). Eliciting the intercultural in foreign language education at school. In A. Paran & L. Sercu (Eds.), *Testing the untestable in language education* (pp. 52-73). Multilingual matters.

Liddicoat, A., & Scarino, A. (2012). Assessing intercultural language learning. In J. Jackson (Ed.), *The Routledge handbook of language and intercultural communication* (pp. 123-142). Routledge.

Liu, K.-L. (2021). Teaching and Assessing Intercultural Communicative Competence (ICC): Experience with Student-Authored Critical Incidents. *Taiwan Journal of TESOL, 18*(1), 1–27. https://doi.org/10.30397/TJTESOL.202104_18(1).0001

Llosa, L. (2017). Assessing students' content knowledge and language proficiency. In E. Shohamy, I. G. Or & S. May, *Language testing and assessment* (pp. 3-14). Springer.

Llurda, E., & Calvet-Terré, J. (2022). Native-speakerism and non-native second language teachers: A research agenda. *Language Teaching, 57*(2), 229-245. https://doi.org/10.1017/S0261444822000271

McIntyre, D. (2005). Bridging the gap between research and practice. *Cambridge Journal of Education*, *35*(3), 357-382. http://dx.doi.org/10.1080/03057640500319065

McNamara, T. (2000). *Language testing.* Oxford University Press.

Murphy, E. (1988). The cultural dimension in foreign language teaching: Four models. *Language, Culture and Curriculum*, *1*(2), 147-163. http://dx.doi.org/10.1080/07908318809525033

Paige, M. R., Jorstad, H., Siaya, L., Klein, F., & Colby, J. (1999). Culture learning in language education: A review of the literature. In R. M. Paige, D. L. Lange & Y. A. Yeshova, *Culture as the Core: Integrating Culture into the Language Curriculum* (pp. 47-113). University of Minnesota.

Rodríguez, A. R., & Espinar, Á. M. (2015). General and specific culture learning in EFL textbooks aimed at adult learners in Spain. *Studia Anglica Posnaniensia*, *50*(1), 5-25.

Safa, M. A., & Tofighi, S. (2021). Intercultural communicative competence beliefs and practices of Iranian pre-service and in-service EFL teachers. *Innovation in Language Learning and Teaching, 16*(2), 164–175. https://doi.org/10.1080/17501229.2021.1889562

Scarino, A. (2010). THE ISSUE: Assessing intercultural capability in learning languages: A renewed understanding of language, culture, learning, and the nature of assessment. *The Modern Language Journal*, *94*(1), 324-329. http://dx.doi.org/10.1111/j.1540-4781.2010.01023.x

Scarino, A. (2017). Culture and language assessment. In E. Shohamy, I. G. Or & S. May (Eds.), *Language testing and assessment* (pp. 15-32). Springer.

Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal, 59*(4), 339-341. https://doi.org/110.1093/elt/cci064

Sercu, L. (2002). Autonomous learning and the acquisition of intercultural communicative competence: Some implications for course development. *Language, Culture and Curriculum*, *15*(1), 61-74. http://dx.doi.org/10.1080/07908310208666633

Sercu, L. (2004). Assessing intercultural competence: A framework for systematic test development in foreign language education and beyond. *Intercultural Education*, *15*(1), 73-89. http://dx.doi.org/10.1080/1467598042000190004

Sercu, L. (2010). Assessing intercultural competence: More questions than answers. In A. Paran & L. Sercu (Eds.), *Testing the untestable in language education* (pp. 17-34). Multilingual Matters.

Shin, J., Eslami, Z. R., & Chen, W.-C. (2011). Presentation of local and international culture in current international English-language teaching textbooks. *Language, Culture and Curriculum*, *24*(3), 253-268. http://dx.doi.org/10.1080/07908318.2011.614694

Schauer, G. A. (2020). Measuring intercultural competence. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 359-370). Routledge.

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Education Research Journal*, *36*(2), 299-316. http://dx.doi.org/10.1080/01411920902919257

Weir, C. J. (2005). Testing as validity. In C. J. Weir, *Language testing and validation: An evidence-based approach* (pp. 1-40). Palgrave Macmillan.

Zerzová, J. (2012). *Interkulturní komunikativní kompetence a její rozvíjení v hodinách anglického jazyka na 2. stupni ZŠ* [Intercultural communicative competence and its development in Czech lower-secondary English classes]. Munipress.

Zotzmann, K. (2015). The impossibility of defining and measuring intercultural competencies. In D. J. Rivers (Ed.), *Resistance to the known: Counter-conduct in language education* (pp. 168-191). Palgrave Macmillan. http://dx.doi.org/10.1057/9781137345196_8

ANALYSED TEXTBOOKS

Baláková, L., Baron, U., Hamzová, A., Knápek, J., Pešková, D. (2020) Maturita z anglického jazyka – Maturitní témata [Maturita in the English language – maturita topics]. Didaktis.

El-Hmoudová, D. (2014). Angličtina - edice Maturita [English – the maturita edition]. Vyuka.cz.

Mrákota, J. (2005). Maturitní otázky z angličtiny [Maturita questions for English]. Jiří Mrákota – vydavatelství jazykové literatury.

Smith-Dluhá, G. (2008). Angličtina - otázky a odpovědi nejen k maturitě [English – question and answers not just for maturita]. Infoa.

https://doi.org/10.17323/jle.2025.22570

From "Spicing Up My Writing" to "Convincing My Supervisors": EFL Learners' Motivations for Using Promotional Language ('Hypes') in Academic Texts

Cita Nuary Ishak [1], Yazid Basthomi [1], Nurenzia Yannuar [1], Amelia Abdullah [2]

[1] Universitas Negeri Malang, Malang, Indonesia
[2] Universiti Sains Malaysia, Pulau Pinang, Malaysia

ABSTRACT

Background: The trend of promotional language (hypes) in academic discourse, such as *critical*, *robust*, *new*, *discover*, and *undoubtedly*, has raised concerns about the changing nature of scholarly communication. While previous studies have documented this trend in published texts, the motivations driving hype usage among developing academic writers, especially in EFL context, remain underexplored.

Purpose: This paper investigates Indonesian EFL learners' use of hypes in theses and dissertations, examines their perceptions and the factors that motivate hype usage in unpublished academic texts.

Method: Through purposive sampling, we conducted in-depth semi-structured interviews with 12 Indonesian EFL learners whose theses and dissertations contained hypes, focusing on the intentions and reasons for using hypes. Hypes were analyzed using Millar et al.'s (2020) functional framework, and reflexive thematic analysis was conducted to identify motivational patterns underlying their usage.

Results: Through thematic analysis of in-depth interviews, this study reveals that participants generally viewed hypes positively and strategically used them across all functional categories, with a novel category targeting the research gaps. While viewing hypes positively as persuasive tools, they expressed concerns about appearing overconfident. Five external motivational factors were identified: supervisory expectations shaped by hierarchical power dynamics, audience awareness, AI tool influences, classroom instructions, and published writing conventions.

Conclusion: This study demonstrates that Indonesian EFL learners use hypes mainly to meet supervisory expectations rather than publication pressures. The findings offer three key theoretical contributions: first, that hype usage represents identity construction where EFL learners negotiate academic and cultural expectations; second, that power asymmetries in hierarchical context manifest linguistically through rhetorical compliance; and third, that AI tools now function as rhetorical agents alongside traditional human influences in academic discourse socialization.

KEYWORDS

academic identity; EFL learners; hypes; promotional discourse; reflexive thematic analysis; rhetorical stance

Citation: Ishak, C.N., Basthomi, Y., Yannuar, N., & Abdullah. A. (2025). From "Spicing up my writing" to "Convincing my supervisors": EFL learners' motivations for using promotional language ('hypes') in academic texts. *Journal of Language and Education, 11*(3), 110-126. https://doi.org/10.17323/jle.2025.22570

Correspondence:
Cita Nuary Ishak,
cita@polinema.ac.id

Received: September 24, 2024
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

The spread of 'publish or perish' principle, which many scholars view as a negative and harmful pressure (Lambovska & Todorova, 2021), has contributed to making academic writing more promotional (Millar et al., 2019; 2020; 2022; Hyland & Jiang, 2021a; 2021b). In addition, emerging technologies have enabled unprecedented schemes in research production and promotion (Raitskaya & Tikhonova, 2023), which further foster self-promotion in academic texts.

In academic texts, promotion can be identified through language that highlights and embellishes positive aspects of the text. In the literature, this phenomenon is commonly referred to as hype (Rinaldi, 2012; Millar, et al. 2019). The plural term, hypes, comprise any subjective language (Millar et al., 2019), such as descriptive words e.g., *excellent*, *sustainable*, *outstanding* (Scott & Jones, 2017), boosters and intensifiers e.g., *clearly*, *obviously*, *completely* (Hyland, 2005), emphatics e.g., *for sure*, *no way* (Hinkel, 2005), positive words e.g., *innovative*, *unique* (Vinkers et al., 2015), and theatrical words e.g., *reveal*, *discover* (Wheatley, 2014).

Most previous studies have focused on hypes in texts that belong to the high-stakes genre, such as published articles, high impact and grant submissions (e.g., Millar et al., 2019; 2020; 2022; Hyland & Jiang, 2021a; 2021b), with limited attention to the lower-stakes genre like unpublished theses and dissertations. These texts are written to convince supervisors and examiners while meeting institutional publication expectations, particularly in Indonesia, where competitive research funding supports student publication goals (Miasari et al., 2018; Arsyad, 2019). Such pressures may encourage writers to use hypes to accentuate research value and significance (Hyland & Jiang, 2021b) and to overstate the predicted outcomes of their work (Wäscher et al., 2020).

Our pilot analysis of hypes in Indonesian EFL learners' theses did confirm a growing trend of hypes over the last ten years (Ishak et al., 2022). Hence, we speculated that the institutional requirement to publish was one of the factors contributing to the rise of hypes in this unpublished genre. However, this speculation remain untested as existing research on hypes has largely focused on frequency analysis rather than writer's intentionality and reasons underlying hype usage. While studies like Millar et al. (2020) have begun to explore authors' perspectives on hypes, studies into the motivational factors driving EFL writers' use of hypes remain limited.

This gap is significant because writers' rhetorical choices in academic discourse reflects the processes of identity construction and disciplinary socialization (Hyland, 2002; Ivanič, 1998). Therefore, understanding why EFL student-writers choose to use hypes can give some insights into how they position themselves as legitimate members of academic communities while navigating institutional expectations and cultural constraints.

To address the gap, the present study explores how student-writers perceive hypes in theses and dissertations, and why they choose to hype. Studies on sub-academic communities, such as the Indonesian EFL learners, where emergent academic writers are to gain more recognition through their academic works, are worthy of investigation. This is because such studies can give information about how the learners interact with and how they manage academic interactions in a wider academic community (Avena & Yumarnamto, 2022). In addition, they can yield better understanding about learners' preferred rhetorical practices in academic writing in order to meet the expectations of disciplinary communities in an international publication context (Martín & Pérez, 2014).

Furthermore, as most Indonesian EFL learners aim to enhance their international standing (Yannuar et al., 2014; Miasari et al., 2018), it is crucial to understand how and why hypes are used in their academic texts. This issue becomes more critical given the shift in international writing conventions from objective and impersonal to more interactive and interpersonal styles (Malmir et al., 2019). This study will investigate whether hype usage is encouraged by the desire for international recognition, as is common among other L2 writers (Martín & Pérez, 2014), or if the production of hypes is driven by other internal and external factors. These findings may help EFL learners make informed choices when using promotional language in theses and dissertations, especially when they aim to boost international recognition.

Given this background, this study explores hypes in theses and dissertations and the writers' intentions when using hypes in these texts. Following Millar et al. (2020), this study investigates contributing factors to the use of hypes, including external and internal factors. To further investigate these issues, this paper aims to answer the following research questions.

RQ1: How do Indonesian EFL learners perceive hypes?
RQ2: What parts of the text are the most and least targeted by hypes?
RQ3: What are the factors that influence the writers to use hypes?

LITERATURE REVIEW

Definitions of Hypes

Hypes are commonly defined as an overestimation of the significance of research findings, and focus more on the benefits than on the risks (Master & Resnik, 2011). They are often used in a hypercompetitive environment where writers are to increase their chances of having their work published and are usually shown by descriptive words (Scott & Jones, 2017). More broadly, hypes involve an exaggeration about the certainty of research findings, the promise of future application of the findings, and various beneficial aspects of science (Intemann, 2022). Millar et al.'s (2019) conceptualize hypes as "hyperbolic and/or subjective language that authors use to glamorize, promote, embellish and/or exaggerate aspects of their research" (p. 139). In this study, hypes encompass linguistic features that embellish research significance, design of the study, novelty, emphasize the credentials of the researchers, or simply reflect a positive attitude towards the results.

To explore hypes in the participants' academic texts, we compiled a set of lexical items that have been previously identified as hypes (e.g., Hyland, 2005; Hyland & Jiang, 2021b, McCarthy, 2015; Millar et al., 2019; 2020), such as *important*, *interestingly*, *critical*, *robust*, *new*, *discover*, *unveil*, and *undoubtedly*. While most of these studies mainly focused on the textual analysis of reports of randomized controlled trials (Millar et al., 2019), successful grant applications (Millar et al., 2022), highly cited articles (Hyland & Jiang, 2021a), and journal articles (Vinkers et al., 2015; Wang & Yang, 2015), the present study goes further by focusing on writers' intentions and motivations for using hypes. As Millar et al. (2020) argued, what should be further explored in the study of hypes is authors' intentions.

From a discourse analytic perspective, lexical items identified as hypes cannot be fully understood apart from the intentions that guide their use. To better understand how and why EFL learners use hypes in their theses and dissertations, this study draws on three complementary theoretical perspectives (see Table 1).

Martin and White's (2005) Appraisal Theory, which focuses on how writers express attitude, judgement, and graduation, provide a lens for analyzing hypes as evaluative language that amplifies meaning through graduation. Swales' (1990) genre analysis approach helps explain strategic use of hypes at specific textual locations where writers must present the value and contribution of their work. Finally, Lillis and Curry's (2010) academic literacies framework conceptualizes academic writing as social practice that is shaped by power relations, cultural expectations, and institutional contexts. Taken together, these perspectives position hypes not merely as lexical choices, but as rhetorical resources that writers use to meet disciplinary conventions, position their research, and negotiate their academic identities within complex sociocultural contexts.

Functions of Hypes

Previous studies have shown that hypes serve multiple promotional functions in academic texts. At the most basic level, they stress shared information (Hyland, 1999) and express authors' certainty in their claims (Hyland, 2005). More strategically, hypes amplify the novelty, scale, significance, and rigor of research projects, while conveying authors' evaluative stance (Millar et al., 2022). Additionally, hypes engage readers and draw more attention to the salient research features, which contributes to the readability and persuasiveness of claims by reinforcing interactivity (Hyland & Jiang, 2021a; Rinaldi, 2012).

These empirical findings align with the theoretical frameworks outlined above. Martin and White's (2005) graduation concept explains how hypes function as linguistic resources that allow writers to "turn the volume up or down" on their meanings, which makes them powerful tools for academic persuasion. From a genre perspective, Swales' (1990) CARS (Create a Research Space) model further shows why hypes may concentrate in specific textual locations. When writing academic texts, writers must establish territory, identify niches, and occupy those niches, all of which naturally invite the use of hypes. The strategic use of hypes thus reflects writers' genre awareness and their attempts to fulfill communicative purposes within established rhetorical structures.

Table 1

Theoretical Frameworks for Analyzing Hypes and Motivations to Hype in Academic Writing

Theoretical fcramework	Key concepts	Relevance to the concept of hypes	Contribution to analysis	Application to EFL learners' writing
Appraisal theory (Martin & White, 2005)	Attitude. Judgement. Graduation. Evaluative language.	Views hypes as evaluative resources that amplify meaning through graduation	Provides a basis for analyzing how writers express their stance and intensify their claims	Explains how EFL learners use graduation to strengthen evaluative positions in theses and dissertations
Genre analysis (Swales, 1990)	CARS model. Rhetorical moves. Communicative purposes. Discourse community expectations.	Explains strategic placement of hypes in textual locations	Shows why hypes are likely to concentrate in particular sections and serve genre-specific functions	Reveals how EFL learners navigate disciplinary conventions and rhetorical structures through hype usage
Academic literacies framework (Lillis & Curry, 2010)	Social practice. Power relations. Cultural. Expectations. Institutional contexts.	Conceptualizes hypes usage as shaped by sociocultural factors, including supervisory relationships, assessment criteria, and institutional pressures	Contextualizes linguistic manifestations of hypes within social practices and power dynamics	Explains how Indonesian institutional contexts, publication pressures, and cultural expectations influence EFL students' hype usage

However, as emphasized by Lillis and Curry (2010), linguistic choices of writers cannot be separated from the social practices and power relations that shape academic writing. The institutional contexts in which theses and dissertations are produced, including supervisory relationships, assessment criteria, institutional and cultural expectations may influence how and why EFL writers use hypes. From these frameworks, it is apparent that hype usage reflects complex negotiations between individual rhetorical goals and sociocultural constraints.

Ethical Debates of Hype Use

Despite their functions, the promotional nature of hypes has generated few concerns in academic communities. This is because the central feature of hypes is exaggeration (Intemann, 2022) and exaggerating is rarely considered a good thing (Rinaldi, 2012). Scott and Jones (2017) argued that scholars need to acknowledge that misinterpretation of findings through hype words such as *excellent*, *remarkable*, and *extraordinary* is a serious matter for scientific integrity and a disturbing trend in scientific writing.

In addition, Millar et al. (2022) argued that some rhetorical devices that function as hypes fall under the concept of spin (see Boutron et al., 2010), which potentially misreport and misrepresent research findings. Even though hyperbolic use of language is basically a natural tendency in human speech, it may lead to "a distortion of the truth" when used in texts (Leech, 1983, p. 148) and can damage the credibility of the texts (Hinkel, 2005). It may also undermine the real importance and objective of a study, subjugate authors' attitude (Millar et al., 2020; Hyland & Jiang, 2021a), and bias readers' interpretation of knowledge (Millar et al., 2019).

Another concern about hypes is they may create some air of excitement (Wheatley, 2014). For example, the word *reveal* in research articles, although grammatically correct, can give more dramatical effect compared to less emotive words such as *show*, *indicate*, or *tell*. Wheatly argued that claims such as "*We have revealed for the first time…*" is irritating because despite its truth, the authors declare themselves "as true pioneers by making prior claims, when the whole purpose of a primary research paper is to communicate new findings" (p. 14). Additionally, when hypes are used too frequently, they can become cliché and jargon-like to the point that they lose their force.

Existing Studies on Hype Use in EFL Contexts

In the last two decades, there have been some studies that investigate the linguistic manifestations of hypes in various texts within academic genre (e.g., Hinkel, 2005; Master & Resnik, 2013; Rinaldi, 2012; Martín & Pérez, 2014; Wheatley, 2014; Scott & Jones, 2017). These studies suggest that the linguistic manifestations of hypes differ from one text type to another, and that certain parts of the text have a different range of hypes. Recent studies on hypes such as Millar et. al. (2019, 2020, 2022), Moreno (2021), and Hyland and Jiang (2021a, 2021b) indicated a growing number of hypes used in academic texts over the years. What can, and should, be explored further, according to Millar et al. (2020), is writers' reasons to employ hypes.

The answers to why writers hype can be elicited in two ways. The first approach is by examining what prior research indicates about the possible factors that foster hypes in academic texts. Literature on hypes have revealed that in academic texts, hypes are generally associated with the intentions to portray research in a positive light and gain public support for research (Master & Resnik, 2011). These tendencies can also be better understood by referring to the perspectives that view academic writing as socially constructed, where writers must navigate individual, local, institutional, and global expectations that shape how they present their work (Swales, 2004; Lillis & Curry, 2010). The second, more crucial, approach involves gaining direct insights to writers' perspectives through in-depth interview. Unfortunately, this approach remains limited in hype research, with Millar et al. (2020) being among the few studies to use it.

Using in-depth interviews, Millar et al. (2020) interviewed writers of medical research articles who published research reports containing hypes. The study found that the reasons for hype are multifaceted and may be influenced by internal and external factors such as writers' struggle for objectivity, interventions from editors, linguistic ability and replication of academic writing conventions. Furthermore, less sensitivity to the evaluative nuances of certain words among some non-native speakers of English, local context, and local academic cultures may also play a part. To confirm these findings, they suggest further studies to investigate "the language choices of authors from different backgrounds working in different contexts (p. 62)", which the present study attempts to address.

METHOD

Design

This study employed a qualitative approach with a purposive sampling method to explore hypes in Indonesian EFL learners' theses and dissertations and the motivations and factors that influence their uses. For this purpose, our participant selection included both current students and alumni from the English Language Education (ELE) program at Universitas Negeri Malang, Indonesia. This is to ensure maximum representational diversity in academic writing practices. This approach allowed us to capture a broader spectrum of academic writing experiences and contexts, rather than limiting our analysis to a specific timeframe or cohort.

Semi-structured interviews and text analysis were used as the primary data collection instrument. The interviews were done to gain deeper insights into the students' perceptions, reasons, and decision-making process when using hypes that textual analysis alone could not reveal. The interview guide (see Appendix) was adapted from Millar et al.'s (2020) study on why authors of medical papers choose to hype. Additionally, the use of interviews aligns with our exploratory research objective of understanding the contextual factors that influence the use of hypes rather than generating broadly generalizable findings. Through means of a purposive sampling method, 12 students were selected and interviewed. This sample size and approach align with common practices in qualitative research, particularly those involving interviews (Lincoln & Guba, 1985). This sample size also follows Sandelowski (1995) who recommends that qualitative sample sizes be large enough to yield rich, new understandings of the studied phenomenon, yet small enough to allow for deep, case-oriented analysis.

The number of participants was also determined based on the principle of data saturation (Saunders et al., 2018). We found that after the ninth interview, there were no new themes emerged, which confirmed that informational redundancy (Lincoln & Guba, 1985) has been achieved. We continued the interviews with another three participants to confirm this data saturation. This approach is consistent with qualitative research principles, where the focus is on depth and quality of data rather than quantity or generalizability (Patton, 2015). The data collected from the interviews are read and analyzed by two of the authors to categorize the factors that influence students' use of hypes. This practice also confirms saturation as the interview transcriptions are reordered and reanalysed several times (Constantinou et al., 2017).

Lastly, it is important to note that the limitation of this study is that it does not aim to generalize findings to all EFL learners or academic writing contexts. Instead, the intention is to provide a detailed, context-specific understanding of how hypes are perceived and why they are used in this particular setting. Thus, this study attempts to give some insights into a previously less explored area of academic writing among Indonesian EFL learners, which can serve as a foundation for future, more extensive studies in this field.

Classification of Hypes in Theses and Dissertations

As previously stated, this study is guided by Millar et al.'s (2019) concept of hypes. To identify hypes in the participant papers, we compiled a list of lexical items previously identified as hypes, which is largely based on the sources of hype in Hyland and Jiang (2021). Of these hypes were boosters and attitude markers (Hyland, 2005), positive words (Vinkers et al., 2015), and superlatives (McCarthy, 2015). We extended the list by including theatrical words (Wheatley, 2014), conversational intensifiers and overstatements (Hinkel, 2005), hyperbolic terms and hype items (Millar et al., 2019, 2020, 2022).

To ensure these items functioned as hypes, we established exclusion criteria: words were excluded if 1) they were supported by immediate justification (e.g., statistical evidence, expert validation, or pilot study findings), 2) they appeared in quotations from previous studies, or 3) they were extracted from interview excerpts. To code hypes in the participants' writings, we used Millar et al.'s (2020) five functional categories (Table 2). The sixth category, Research Gap (RG), was added based on our pilot analysis of the sample writings to capture instances where authors emphasized gaps in existing research to justify their work.

The process of assigning hypes to these categories was conducted collaboratively by two of the authors. Both coders independently applied Millar et al.'s (2020) framework to the

Table 2

Hypes' Functional Categories

Hype target	Definition	Example
Research Topic (RT)	comprises hypes referencing the overall domain of study and/or the specific research topic	*importantly, critical*
Research Method (RM)	comprises hypes targeting materials and methods used in the study, including the research design and the qualifications of the researchers themselves.	*experienced, certified, correct*
Research Outcome (RO)	efers to hypes that emphasise a positive aspect of the results	*importantly, highly*
Research Primacy (RP)	comprises hypes that refer to the research as superior or assign it priority, often in terms of newness	*notable, unique*
Research Conclusion (RC)	comprises hypes that magnify the implications of the results,	*essential, appropriate*
Research Gap (RG)	emphasized gaps in existing research to justify their work	*very, abundant, few*

Note. Adapted from Millar et al. (2020)

corpus of 12 participants' papers using a manual coding. Although each instance of hype was carefully examined and categorized based on its context within the text, discrepancies still occurred in several cases, which primarily involved borderline instances where context determined hype functions. These discrepancies were resolved through several rounds of discussions. This process involved rereading the original text and revising the initial coding results. This approach ensured that the final categorization of hypes was as accurate and consistent as possible. The example of hype identification and classification can be found in Appendix 2.

Participant Selection

To investigate the motivations and factors influential to the use of hypes in theses and dissertations, 12 Indonesian EFL learners were selected through purposive sampling (Palys, 2008). Open-ended interview questions, which are useful for generating richer data (Ogden & Cornwell, 2010), were used, making the sample size sufficient to answer the research questions. As noted by Morse (2000), fewer participants are needed when more detailed data is collected from each individual.

Participants were selected based on three criteria: 1) they had written or were currently writing a thesis or dissertation, 2) they consented to be interviewed, confirmed by signed consent forms, and 3) they used hypes in their academic writing, as proven by a preliminary analysis of their work. Twelve student writers meeting these criteria participated in semi-structured interviews. They were fully responsible for writing their theses or dissertations and are anonymized in this paper as S1 to S12. All specialized in English Language Education, were native Bahasa Indonesia speakers, and used English as a foreign language. The group included two undergraduates, five master's graduates, one current master's student, and four second-year PhD students. Except for one, all participants worked as English teachers in high schools or universities in Indonesia.

Interviews

In-depth semi-structured interviews were conducted between January and March 2024, both in person and online via Zoom, with follow-up clarifications through WhatsApp to ensure the accuracy of the transcriptions. Each interview lasted between 60 to 90 minutes. Prior to the interviews, participants signed informed consent forms, agreeing to the interviews and the analysis of their writing. We identified hypes in their theses and dissertations, coded and categorized them based on Millar et al.'s (2020) five functional categories, as explained earlier. These highlighted hype instances informed the development of the interview protocols.

During the interviews, participants were shown the highlighted hypes in their writing and asked the questions in the interview protocol (Appendix 1), which explored two main questions: 1) what they intended to communicate, and 2) why they chose those specific words. Follow-up questions explored their perspectives, expectations, and decision-making process regarding the use of hypes. All interviews were recorded and transcribed, with transcriptions sent back to participants for review and clarification. Following interview transcription and verification with participants, we conducted thematic analysis following Braun and Clarke (2006) to identify key themes related to the reasons and factors that influence hype usage.

Taking an inductive approach, we conducted line-by-line coding without predetermined themes (see Appendix 3). During this process, we assigned descriptive codes such as 'supervisor influence', 'competitive pressure' and 'classroom instruction'. This coding process involved iterative refinement over several weeks, with regular consultation to ensure consistency. To improve trustworthiness, we employed prolonged engagement with the data over several months and maintained regular discussions throughout the analytical process to ensure the emerging themes accurately reflected the complexity of participant experiences.

It is important to note, however, that interpretive member checking was not implemented given the sensitive nature of discussing self-promotional behaviors, where participants might feel reluctant to validate findings that could appear professionally compromising. Instead, we employed alternative validation strategies: triangulation with textual analysis of participants' actual academic writings, documentation of analytical decisions, and ongoing team consultation during interpretation phases.

RESULTS

The results are presented in three sections which correspond to the research questions. First, we examine how Indonesian EFL learners perceive hypes in academic writing. Second, we analyze which parts of academic texts are most and least targeted by hypes. Finally, we explore the factors that influence writers' decisions to use hypes. This structure provides a systematic investigation of participants' attitudes, practices, and the contextual forces shaping their use of promotional language in academic discourse.

Participants' Perceptions on Hypes

Positive

Generally, eight participants had a positive view on hypes. These participants were aware of the concerns surrounding hypes, but believed that their benefits outweigh the concerns. S7 believed that hypes could make a text attractive. S12 also stated that hypes have a marketing force.

[1] "If I didn't use hypes, I'm afraid my writing would sound monotonous. People nowadays like these kinds of words. Using hypes helps us as authors get as many readers as possible." (S7)

[2] "Words and phrases that embellish parts of our research work basically add colors to our writing. This type of diction has a marketing force, so it is important to use it as it can attract as many readers as possible. I think authors should be able to promote the positive aspects of their work to have large readership and citation." (S12)

Similarly, S3 noted that hypes are linguistic features necessary to make a research report interesting to read.

[3] "Academic texts can be so boring and heavy to read, so we need to add some spice and it can be done through hypes. They spice up my writing." (S3)

S4 and S8 described the way hypes helped them during thesis and dissertation supervisory. Hypes emphasizing the research strengths could persuade their supervisors and board of examiners, and serve as a 'signpost' that easily direct supervisors' attention to certain parts of the text.

[4] "I think they [hypes] are necessary in academic texts. These words help me get my thesis approved. As a writer, I think it is important to show what's so good about our work so that people will be convinced that it is worth reading." (S4)

[5] "I don't think supervisors have the time to read everything in my thesis. Using hypes can help them locate the important points quickly and focus on those." (S8)

Ambivalences

Three participants reported to have some ambivalences when using hypes. S6 mentioned that the ambivalences in using hypes came from concerns about projecting overconfident authorial voice.

[6] "When hypes are overused, then the author is certainly exaggerating. Science must be communicated as it is. But if hypes are used in moderation and in the right place, then it's okay. They can be useful for convincing readers and attract them to our work, but overusing these words can make the authors sound overconfident." (S6)

In a similar line, S9 and S11 pointed out the need for moderation when using hypes to avoid appearing overconfident, and for argumentation and supporting piece of evidence to justify their uses.

[7] "Overhyping is bad, but not hyping at all is also bad. As a writer we must be mindful about how much we can use them. Hypes will be meaningless if we fail to provide argumentation and supporting evidence. Overhyping our claims will make us appear overconfident. Not using them at all can make the arguments sound weak and hence, unattractive to readers." (S9)

[8] "When we use hypes, it is important to think about the supporting evidence. Using too many hypes can look like we fabricate things, so it is important to base the hyped claims on facts." (S11)

Negative

Only one participant (S5) held a negative view on hypes. This view was influenced by his personal experiences and reflections as a reader of academic works.

[9] "I often read papers that contained hypes and when I read the whole text to check if hyped claims were proven, I found nothing. It turned out that the research was not as 'in-depth' as it claimed, which was disappointing. It felt like false marketing when the real product is far from the hyped claims made in the paper." (S5)

Parts of the Text Targeted by Hypes

The frequency distribution shows a clear pattern of hypes in participants' texts (Table 3). The high frequency of hypes targeting the research topic reflects participants' primary concern with establishing research relevance. As S12 explained: *"I want to show that this topic is really important nowadays, because it is used in everyday life. This is my way of making the readers interested to read my dissertation"* [10]. Similarly, S9 deliberately escalated from "important" to "crucial" to "*show readers that this was more than just important*". [11].

Hypes targeting the research method constituted the second-largest category. Using these hypes, participants emphasize procedural rigor to show their credibility. S8 noted: *"I explicitly mentioned the advantage of using this method because I want to show the readers that my research was on the right track"* [13], while S9 emphasized showing *"how careful I was as a researcher"* [14].

Furthermore, we identified three distinct rhetorical functions from the analysis. Justification strategies (Topic, Gap) legitimized research urgency; validation strategies (Method, Outcome) established trustworthiness and promoted applicability; differentiation strategies (Primacy, Conclusion) reinforced innovation and research value.

Participants demonstrated high audience awareness. Hypes targeting the research gap specifically anticipated thesis supervisors and examiners' questions related to justification for conducting the study. In contrast, outcome-focused hypes targeted audiences who are practitioners. S7 explained that emphasizing the technique proposed in her thesis was necessary because she *"wanted the technique to be used by teachers and pre-service teachers"*.

Hypes targeting the research conclusion and primacy had the lowest frequency but served emphasis functions. S6 argued that readers "*needed to be ensured that the technique is really beneficial*" even when benefits had been previously stressed. The primacy-focused hypes appeared exclusively in studies using research and development design, where participants like S2 used the term "potential revolution" to "*sell the product or idea*" [17].

Table 3

Parts of Theses and Dissertations Targeted by Hypes

Target	Frequency	Range	Sample hypes	Hyped claims	Primary function	Key motivation
Research Topic	39	11/12	*crucial*; *fundamental*; *important*; *paramount*, *urgent*	This study is *paramount* for the continuity of the subsequent research… [S1]	establish significance, show relevance	convince readers of importance
Research Method	22	10/12	*careful*; *comprehensive*; *deeper*	I will make a *careful* analysis both for discussing verbs and rhetorical moves… [S9]	show rigor	demonstrate research(er) credibility
Research Outcome	16	8/12	*many*; *successfully*; *significantly*	In the field of TEFL, the use of ICT has been demonstrated to have *many* benefits for both learners and teachers. [S10]	promote results	encourage application
Research Gap	7	5/12	*abundant*; *very*	There has been so far *very* limited research showing this activity… [S3]	emphasize urgency	justify research rationale to supervisors and examiners
Research Conclusion	3	2/12	*successfully*; *easy*	The implementation of making connection technique *successfully* achieved the criteria of success… [S6]	reaffirm benefits	reinforce product value
Research Primacy	3	3/12	*new*	This research will give a *new* insight on how GOs might be tailored to specific learning styles… [S12]	highlight innovation	establish novelty, encourage adoption

Factors Influencing the Production of Hypes

Our analysis revealed five external factors that foster EFL learners' use of hypes: supervisory expectations, audience awareness, digital writing tools, classroom instruction, and published writing conventions.

Supervisory Influence and Gatekeeping

Supervisory expectations emerged as the primary reason for hype usage. These expectations were either in the form of explicit directives or implicit modeling. Participants like S1 and S12 described using hypes in direct response to supervisory feedback about research significance and argumentative strength.

> [18] "My supervisor used to question my research's significance during supervision. By hyping the research significance, I could persuade her to approve my arguments and allow me proceed to thesis defense. If I did not emphasize the research significance, I was afraid I might have to revise the entire manuscript." (S1)

> [19] "Since then, my choice of promotional words aligns with their expectations, simply because I want my work to be approved by my supervisors. If I do not highlight the important parts from my work, my supervisors won't approve it." (S12)

Sometimes supervisors' expectation about how much promotion should be used was not explicitly stated. When this happened, participants would observe their supervisors' published works and senior students' theses to decode acceptable promotional styles. By doing so, they were able to follow their writing styles, especially their uses of hypes. As S2 described:

> [20] I used to read my supervisors' published works just to observe the promotional language used in their work, so I can understand what and how positive aspects were highlighted. (S2)

The influence of supervisors on the use of hypes in student theses and dissertations goes beyond a matter of their expectations about which hype words to use and which part to hype. Their presence during face-to-face supervisory has created pressures. S6 and S3 described that the heightened pressure came from established relationships as well as supervisory expectations.

> [21] "During face-to-face supervisory, I would get comments about my work right away. This gave me more pressure. My advisors were also my lecturers, with whom I have had a close relationship throughout the years. They knew the quality of my undergraduate thesis, so if I failed to show the strengths of my master's thesis, it would be embarrassing." (S6)

> [22] "I would always picture who would read my thesis and that would be my supervisors. Sometimes I pictured how they would respond to my work during supervisory and that made me overthinking. I was always aware that I would be judged from my work." (S3)

Furthermore, supervisors' gatekeeping roles directly determine how much hypes are allowed, as stated by S7, or if they

are allowed at all, as mentioned by S5, which demonstrates supervisors' power over students' rhetorical choices.

[23] "During consultation with my supervisors, they were all aware that the technique that I proposed in my thesis was new. They asked me to emphasize the benefits of this technique more to persuade the readers." (S7)

[24] "I didn't really have the intention to hype certain parts of my thesis and I had supervisors who thought the same way. If they were the type that encourage the supervisees to do so, then I might have used hypes." (S5)

Audience Awareness

Participants' use of hypes reflect their awareness of multiple readerships and identity concerns. The anticipation of evaluation from supervisors, colleagues, and peers, as articulated by S2 and S4 below, has generated pressure to demonstrate competence through strategic emphasis.

[25] "I would feel embarrassed if my work lacked good quality, as it would be read by my supervisors and colleagues." (S2)

[26] "Sometimes the pictures of people who would read my work, like supervisors, my fellow teachers, and future university students, or other researchers make me quite pressured. I need to be really convincing so I can catch up to their level of expectations, which inevitably influenced my decision to hype" (S4)

Audience awareness also manifested in some kind of defensive strategies, with participants like S9 anticipating perceived methodological weaknesses and S11 ensuring clarity for readers who were not familiar with her research topics.

Digital tool Influence

This theme captures the role of (digital) writing tools and resources in shaping participants' hype usage. Participants explained that their use of hypes is sometimes influenced by AI tools, which often suggest options for word choices. To vary their diction and avoid repetition, they used dictionaries, thesauruses, and AI tools like Grammarly and Quillbot. While the final decision to follow these suggestions rests with the writers, the presence of these recommendations demonstrates the tools' impact on their use of hypes. S10 narrated as follows.

[27] "When writing some claims in my dissertations, I used Grammarly and Quillbot to help me paraphrase my own sentence. But I was not dependent ...and very selective with the words that they suggest." (S10)

This influence highlights how AI tools subtly shapes academic discourse toward more promotional language. By offering more persuasive or emphatic word choices, these tools may have led writers towards a style that emphasizes the strengths and significance of their research.

Classroom Instruction

This theme addresses how the decision to hype or not comes from academic instruction. The analytic logic focuses on participants' accounts of how promotional strategy is explicitly addressed in classroom settings rather than solely during individual supervisory meetings. Participants mentioned that their decisions to use or avoid hypes in their theses and dissertations were also influenced by research courses they took. For instance, S9 and S10 had a similar experience of being advised by their professors in proposal development course to avoid making overly hyped claims and using strong words like *must* and *no* as in *we must* or *there is no study*.

[28] "As far as I remember, my own supervisors did not really comment on my use of hypes. It was the professor in the Dissertation Proposal Seminar course who reminded us to avoid making claims that are too extreme." (S9)

[29] "In terms of my choice of words, including hypes, I always try to not make claims that are too strong, like *we must*. I got this advice from my professor in class." (S10)

Public Writing Conventions

This theme examines how participants deliberately model promotional strategies of published authors. S6 noted that she would often imitate the way published writers structure their arguments and the type of language used their works to make her thesis appealing for readers. Similarly, S10 stated that the use of hypes in her dissertation was mostly influenced by the research papers she had read. Whenever her peers successfully published their work in journals, she would read and observe the language used in their work and try to imitate them. This replication represents deliberate engagement into disciplinary discourse conventions.

DISCUSSION

This study explores Indonesian EFL student-writers' perspectives on hypes, their strategic use across textual locations, and the factors that shape their use in theses and dissertations. Overall, Indonesian EFL student-writers generally viewed hypes positively, though some voiced concerns about the risks of overusing them. Despite the ambivalences, hypes were found to concentrate on research topic. This particular finding aligns with Arianto et al. (2023), who found that appeals to salience was the most common promotional strategy used by Indonesian authors writing in Indonesian or English.

Drawing on Martin and White's (2005) Appraisal Theory, Swales' (2004) genre analysis, and Lillis and Curry's (2010) academic literacies framework, the following sections exam-

ine three emerging issues in the motivations underlying the use of hypes.

Supervision as Gatekeeping: Power Dynamics and Rhetorical Compliance

The most significant finding of this study is perhaps the central role of supervisory expectations in shaping students' use of hypes, which reveals supervision as a gatekeeping practice that has a direct influence on students' promotional rhetoric. This finding extends beyond simple compliance to reflect how hierarchical power relations, as theorized by Lillis and Curry (2010), shape students' use of hypes in their academic writing. As supervisors are likely to occupy a higher rank and greater knowledge, students often find themselves in a subordinate position (Deirich, 2023; Zhang & Hyland, 2021). Besides, the hierarchical nature of supervision cannot be fully eliminated, which makes supervisory relationships inherently imbalanced (Nangimah & Walldén, 2023). This asymmetrical power explains why participants in this study reported a higher pressure in promoting their theses and dissertations to the supervisors, compared to writing articles for publication.

The cultural context is particularly significant in this case. It is worth noting that based on Hofstede's cultural dimension, Indonesia scores 78 on Power Distance, which reflects significant power dynamics and causes hierarchical pattern (Liyanti, 2019). This hierarchical structure extends to Indonesian educational system which prioritizes respecting authority figures, including teachers and supervisors. In academic writing, this manifests through students' heightened awareness of supervisory approval, which leads to what can be understood as rhetorical compliance, that is, a strategic use of hypes to align with supervisory and institutional expectations. Thus, unlike the competitive publishing environment described by Scott and Jones (2017) and Hyland and Jiang (2021a; 2021b), where hypes often emerge from pressures to publish, the thesis and dissertation context reveals supervision as a site of discourse socialization where students learn disciplinary conventions through asymmetrical power negotiations.

As discussed in the earlier part of this paper, our diachronic study of hypes in Indonesian EFL learners' theses revealed an increase in promotional language, particularly in the year when publishing parts of theses in reputable journals became mandatory Ishak et al. (2024). Nevertheless, findings from the present study suggest that this increase is not directly related to the competition to publish, which negates our initial speculation. Instead, participants indicated that their use of hypes was primarily motivated by supervisory expectations, rather than to succeed in journal publication or secure research funding.

Additionally, face-to-face meetings with supervisors add to the pressure, as students must be prepared to answer questions directly. This heightened pressure may also be attributed to their personal relationships with the supervisors with whom they strive to maintain harmony, which is valuable in a collectivistic society like Indonesia (Hofstede, 1991). The participants' accounts of using hypes to catch up to their level of expectations and avoid embarrassment reveal the identity work embedded in promotional language choices. This aligns with Swales' (1990) understanding of genre as social action, where writers must fulfill communicative purposes as well as position themselves appropriately within disciplinary communities.

This tension between individual rhetorical choices, supervisory and institutional demands, though explored here in the Indonesian context, is likely to surface in other EFL contexts characterized by hierarchical academic cultures. In Iran, for example, the teacher-student relationship remains highly asymmetrical, with students occupying a largely passive role (Yusofi et al., 2018). Similarly, Chinese students often report authoritative supervisory interactions which limits dialogic negotiation of ideas (Huang et al., 2023). These supervisory dynamics may exert similar rhetorical pressures. It is plausible that in such settings, students also strategically use hypes not only to meet academic expectations but also to preserve interpersonal harmony and secure institutional success.

Identity Negotiation through Hypes

The participants' articulation of audience awareness (i.e., anticipating supervisors, peers, and future practitioners) demonstrates what Martin and White (2005) describe as heteroglossic engagement, where writers put themselves in relation to multiple voices and perspectives. Students' use of hypes to target different audiences reflect their understanding of academic writing as a social practice that requires careful navigation of competing expectations and identity positions.

Particularly important is the finding that students who employed research and development design tend to concentrate hypes on research outcomes and primacy, with the intention to sell the product or idea to encourage adoption. This reveals genre-specific identity positioning, where student-writers adopt entrepreneurial identities alongside academic ones. Such positioning challenges traditional boundaries between objective academic discourse and promotional communication, which suggests that contemporary thesis writing involves multiple, sometimes conflicting, identity negotiations.

Furthermore, the ambivalent attitudes expressed by some participants, who recognize both benefits and risks of hypes, indicate developing rhetorical sophistication. Their concerns about appearing overconfident and damaging credibility demonstrate emerging awareness of the importance of having balanced academic self-presentation. This

ambivalence reflects what Lillis and Curry (2010) identify as the tension between individual rhetorical goals and institutional constraints, where writers, in the context of this study, must negotiate competing demands for self-promotion and scholarly objectivity.

Digital Tools as Rhetorical Agents

An unexpected finding concerns the influence of digital writing tools, especially AI-powered platforms such as Grammarly and Quillbot, on Indonesian EFL students' use of hypes. This finding shows a unique trend in academic discourse socialization that moves from traditional human agents like supervisors, peers, and published authors, to include algorithmic influences on rhetorical choices. The participants' accounts of their selective adaptation of AI suggestions for word choices and paraphrasing reveal that technology plays an active part in shaping hyping practices.

The implication of this finding is in understanding contemporary academic literacy practices. While the participants maintained agency in their final word choices, the consistent availability of more promotional or persuasive alternatives through AI tools has the potential to shape academic discourse into more accepting, if not vulnerable, of hypes. This technological influence may also operate alongside, and to some degree amplify, the other socializing forces identified in this study, such as supervisory expectations, audience awareness, and published writing conventions.

The integration of AI tools into academic writing practices represents what Lillis and Curry (2010) would recognize as an emerging literacy practice that reshapes traditional power relations in academic writing. AI tools, unlike human supervisors or published authors, can provide quick, algorithmic feedback that may lack contextual awareness of disciplinary norms or ethical considerations around promotional language. This raises important questions about the role of technology in academic discourse socialization and the need for critical digital literacy in thesis writing instruction.

Pedagogical Implications for Thesis Writing Instruction and Supervisory Practices

As a whole, the findings of this study have several implications for thesis writing pedagogy and supervisory processes. The central role of supervisory expectations in shaping students' use of hypes suggests the need for explicit discussion of hypes, especially during supervisory practices. Supervisors should be more aware of their powerful influence on students' rhetorical choices and the potential consequences of implicit expectations for hype use. More importantly, thesis writing instruction should help students raise their critical approach to hypes, and include discussions about both the strategic functions and potential risks of using hypes as identified by Intemann (2022) and Scott and Jones (2017). Students' limited awareness of broader implications on ethical concerns indicates that there is a need for more comprehensive coverage about hypes and ethics in academic writing curricula.

Finally, the influence of AI tools on promotional rhetoric calls for the need to include digital literacy components in thesis writing instruction. Teachers should equip students with critical frameworks for evaluating AI suggestions and understanding the impact of algorithmic influences on their use of hypes in academic writing. Moreover, institutional interventions should establish clear guidelines for AI tool usage in thesis writing and provide training programs that help students understand how AI algorithms are trained on existing academic texts, many of which already contain hypes. This could potentially create a feedback loop that amplifies hype usage over time. In addition, supervisory practices need to address AI-mediated writing, for instance, by asking students to identify which suggestions they accepted and why, and if AI suggestions genuinely improve accuracy, or merely add promotional embellishment. This approach can help students become thoughtful users of AI tools rather than passive recipients of their algorithmic recommendations.

CONCLUSION

This study demonstrates that Indonesian EFL students strategically use hypes, primarily targeting research topic, to navigate complex power dynamics within supervisory relationships rather than responding to publication pressures as found in previous literature. The dominant influence of supervisory gatekeeping, amplified by Indonesia's hierarchical culture, shows that hypes in student academic writing stem from complex power dynamics rather than simple stylistic choices.

These findings contribute to our understanding of promotional discourse in EFL academic settings in three important ways. First, regarding student identity formation, this study shows that EFL students engage in identity negotiation through hypes, as they adopt multiple identities: academic, entrepreneurial, and culturally compliant. Second, regarding power asymmetries, the findings reveal that supervisory gatekeeping operates through rhetorical compliance, where hierarchical cultural contexts create conditions for promotional language as students navigate asymmetrical power relations to maintain harmony and secure approval. Third, regarding promotional discourse in EFL settings, this study shows that hypes are heavily influenced by both humans (supervisors, published conventions) and algorithmic agents (AI tools).

The influence of AI-powered tools also requires immediate pedagogical attention. Writing instruction should include digital literacy that helps students critically evaluate AI sug-

gestions, particularly those that are related to hyping practice. As academic discourse is becoming increasingly promotional, the objective should not be eliminating hypes at all but raising critical awareness among students, supervisors, and institutions about their use in scholarly communication.

These findings challenge current approaches to academic writing instruction and supervisory processes. Rather than considering hypes as problematic, educators should help students critically understand when and how to use them appropriately. Furthermore, supervisors should recognize their powerful influence on student rhetoric and discuss hype usage explicitly. It is also important for them to clearly communicate their expectations regarding hypes and their impact on the credibility of academic work. Policymakers who evaluate student academic work should recognize that hypes are embedded in thesis and dissertation writing practices, which requires a critical evaluation of research claims rather than accepting them at face value, particularly when making academic and funding decisions.

The findings from this study have significant implications for understanding how EFL students use hypes to construct their academic identities and navigate power relationships in academic settings. This study shows that when EFL students use hypes, they are not merely following writing conventions but actively positioning themselves within hierarchical academic structures while managing cultural expectations and supervisory relationships. While this study focuses on Indonesian contexts, the interplay between supervisory gatekeeping, cultural hierarchies, and promotional discourse likely extends to other EFL academic settings where power asymmetries and cultural values shape student-supervisor relationships. Future research should examine these motivational patterns in other EFL contexts, particularly those with different cultural values and supervisory practices, to better understand how promotional language functions in student academic writing across diverse educational, and cultural settings.

USE OF AI STATEMENT

No generative AI tools were used in the design, analysis, or interpretation of this study.

ACKNOWLEDGMENTS

We would like to sincerely thank Professor Elena V. Tikhonova and the anonymous reviewers for their valuable feedback and immense guidance that helped us improve the quality of this manuscript. We would also thank all the participants of this study for the information and time they provided.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Cita Nuary Ishak: conceptualization; data curation; formal analysis; investigation; methodology; writing – original draft; writing – review & editing.

Yazid Basthomi: formal analysis; project administration; resources; supervision; validation; writing – review & editing.

Nurenzia Yannuar: data curation; investigation; resources; supervision; validation; writing – original draft.

Amelia Abdullah: methodology; visualization; resources; supervision.

REFERENCES

Arianto, M. A., Haq, M., & Jufrizal, J. (2023). Research article introductions in applied linguistics: A comparative study on the use of appeals. *Journal of Language and Education*, *9*(2), 12-24. https://doi.org/10.17323/jle.2023.16163

Arsyad, S. (2019). Struggling for international publication: The potential rhetorical problems for Indonesian scholars in social sciences and humanities when writing in English. *Advances in Social Science, Education and Humanities Research*, (301), 469-478.

Avena, B. & Yumarnamto, M. (2022). Indonesian EFL students' thesis conclusions: organizational preferences and pedagogical implications. *LLT Journal: A Journal on Language and Language Learning*, *25*(2), 709-728. https://doi.org/10.24071/llt.v25i2.3715

Bhatia, V. K. (2005). Generic patterns in promotional discourse. In H. Halmari & T. Virtanen (Eds.), *Persuasion across genres: A linguistic approach* (pp. 213-225). John Benjamins Publishing.

Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*, *303*(20), 2058 – 2064. https://doi.org/10.1001/jama.2010.651

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Constantinou, C.S., Georgiou, M., & Perdikogianni, M. (2017). A comparative method for themes saturation (CoMeTS) in qualitative interviews. *Qualitative Research, 17*, 571 - 588. https://doi.org/10.1177/1468794116686650

Deyrich, M.-C. (2023). Addressing power asymmetries in doctoral supervision. In C. J. Craig, J. Mena, & R. G. Kane (Eds.), *Studying teaching and teacher education* (Advances in research on teaching, vol. 44, pp. 243-254). Emerald Publishing. https://doi.org/10.1108/S1479-368720230000044024

Fairclough, N. (1993). Critical discourse analysis and the marketization of public discourse: The universities. *Discourse & Society*, *4*(2), 133-168. https://doi.org/10.1177/0957926593004002002

Hinkel, E. (2005). Hedging, inflating, and persuading in L2 academic writing. *Applied Language Learning*, *15*(1 & 2), 29-53.

Hofstede, G. (1991). *Cultures and organizations: Software of the mind*. McGraw Hill.

Huang, J., Li., N., Shumin, W., & He, Z. (2023). The roles of cultural capital in teacher-student interactions in China: A qualitative study of students in higher vocational colleges. *Behavioral Sciences*, *13*(8), 1-17. https://doi.org/10.3390/bs13080690

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, *34*(8), 1091-1112. https://doi.org/10.1016/S0378-2166(02)00035-8

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7, 173-192. https://doi.org/10.1177/1461445605050365

Hyland, K. & Jiang, K. (2023). Hyping the REF: Promotional elements in impact submissions. *Higher Education*. https://doi.org/10.1007/s10734-023-01030-y

Hyland, K. & Jiang, K. (2021a). The Covid infodemic: Competition and the hyping of virus research. *International Journal of Corpus Linguistics*, *26*(4), 444-468. https://doi.org/10.1075/ijcl.20160.hyl

Hyland, K. & Jiang, K. (2021b). 'Our striking results demonstrate …'" Persuasion and the growth of academic hype. *Journal of Pragmatics*, *182*(2021), 189-202. https://doi.org/10.1016/j.pragma.2021.06.018

Intemann, K. (2022). Understanding the problem of "hype": Exaggeration, values, and trust in science. *Canadian Journal of Philosophy, 52*(3), 279-294. https://doi.org/10.1017/can.2020.45

Ishak, C. N., Basthomi, Y., & Yannuar, N. (2024). Hypes in undergraduate thesis abstracts by Indonesian students across years. *Discourse and Interaction*, *17*(1), 51-72. https://doi.org/10.5817/DI2024-1-51

Ivanič, R. (1998). *Writing and identity: The discoursal construction of identity in academic writing (Studies in written language and literacy, vol. 5)*. John Benjamins Publishing Company.

Lambovska, M., & Todorova, D. (2021). 'Publish and Flourish' instead of 'Publish or Perish': A motivation model for top-quality publications. *Journal of Language and Education*, *7*(1), 141-155. https://doi.org/10.17323/jle.2021.11522

Leech, G. (1983). *Principles of pragmatics*. Longman.

Lillis, T., & Curry, M. J. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. Routledge.

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic Inquiry*. SAGE.

Liyanti, L. (2019). Information seeking behavior of santriwati of Madrasah Aliyah Pondok Pesantren Jakarta: A study of Hofstede's cultural dimensions perspective in information seeking behavior. *Jurnal Ilmu Informasi, Perpustakaan dan Kearsipan, 21*(1), 11-18. https://doi.org/10.7454/JIPK.v21i1.002

Lo, Y. Y., Othman, J., & Lim, J. W. (2021). Mapping the use of boosters in academic writing by Malaysian first-year doctoral students. *Pertanika Journal Social Sciences & Humanities*, *29*(3), 1917-1937. https://doi.org/10.47836/pjssh.29.3.23

Malmir, B., Khany, R., & Aliakbari, M. (2019). Journal article highlights in applied linguistics: An exploration into the rhetorical moves and their lexico-grammatical features. *Iranian Journal of English for Academic Purposes*, *8*(4), 49-63.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan. https://doi.org/10.1057/9780230511910

Martín, P. & Pérez, I. K. L. (2014). Convincing peers of the value of one's research: A genre analysis of rhetorical promotion in academic texts. *English for Specific Purposes*, *34*, 1-13. https://doi.org/10.1016/j.esp.2013.09.002

Miasari, S., Arsyad, S., & Arono. (2018). Indonesian authors' stances in citing English research article introductions literature in sciences. *EduLite:Journal of English Education, Literature, and Culture*, *3*(2), 173-187. https://doi.org/10.30659/e.3.2.173-187

Millar, N., Batalo B., & Budgell, B. (2022). Trends in the use of promotional language (hype) in abstracts of successful National Institutes of Health Grant Applications, 1985-2020. *JAMA Network Open*, *5*(8), 1-11. doi:10.1001/jamanetworkopen.2022.28676

Millar, N., Budgell, B., & Salager-Meyer, F. (2020). Hype in reports of clinical research: The authors' perspectives. *English for Specific Purposes*, 60, 53-64. https://doi.org/10.1016/j.esp.2020.07.001

Millar, N., Salager-Meyer, F., & Budgell, B. (2019). "It is important to reinforce the importance of.": 'Hype' in reports of randomized controlled trials. *English for Specific Purposes*, *51*, 139- 151. https://doi.org/10.1016/j.esp.2019.02.004

Mohamad, R. & Saad, N. M. (2016). Power distance culture and the construction of the followership identity. *Asia Pacific Journal of Advanced Business and Social Studies*, *2*(1), 149-160.

Morse, J. M. (2000). Determining Sample Size. *Qualitative Health Research*, *10*(1):3-5. https://doi.org/10.1177/104973200129118183

Nangimah, M., & Walldén, R. (2023). Diverse views on supervision: Insights from interviews with EAL supervisors in Sweden and Indonesia. *Journal of Praxis in Higher Education*, *5*(1), 122-153. https://doi.org/10.47989/kpdc281

Ogden, J., & Cornwell, D. (2010). The role of topic, interviewee and question in predicting rich interview data in the field of health research. *Sociology of Health & Illness*, *32*(7), 1059–1071. https://doi.org/10.1111/j.1467-9566.2010.01272.x

Palys, T. (2008). Purposive sampling. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (vol. 2, pp. 697-698). SAGE.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.

Raitskaya, L., & Tikhonova, E. (2023). Academic integrity: Author-related and journal-related issues. *Journal of Language and Education*, *9*(4), 5-10. https://doi.org/10.17323/jle.2023.18489

Rinaldi, A. (2012). To hype, or not to(o) hype. *EMBO Reports*, *13*, 303-307. https://doi.org/10.1038/embor.2012.39

Sandelowski M. (1995). Sample size in qualitative research. *Research in Nursing & Health*, *18*(2), 179-183. https://doi.org/10.1002/nur.4770180211

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, *52*(4), 1893-1907. https://doi.org/10.1007/s11135-017-0574-8

Scott, S. L. & Jones, C. W. (2017). Superlative scientific writing. *ACS Catalysis*, *7*, 2218-2219. https://doi.org/10.1021/acscatal.7b00566

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge University Press.

Tan, H. & Eng, W. B. (2014). Metadiscourse use in the persuasive writing of Malaysian undergraduate students. *English Language Teaching*, *7*(7), 26-39. http://dx.doi.org/10.5539/elt.v7n7p26

Vinkers, C.H., Tijdink, J.K., & Otte, W.M. (2015). Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis. *British Medical Journal*, 351, h6467. https://doi.org/10.1136/bmj.h6467

Wheatley, D. (2014). Drama in research papers. *European Science Editing*, *40*(1), 14-16.

Wäscher, S., Biller-Andorno, N., & Deplazes-Zemp, A. (2020). "I Don't Want to Do Anything Bad." Perspectives on scientific responsibility: Results from a qualitative interview study with senior scientists. *Nanoethics*, *14*, 135-153. https://doi.org/10.1007/s11569-020-00365-5

Yannuar, N., Shitadevi, I. A., Basthomi, Y., & Widiati, U. (2014). Active and passive voice constructions by Indonesian student writers. *Theory and Practice in Language Studies*, *4*(7), 1400-1408. https://doi.org/10.4304/tpls.4.7.1400-1408

Yusofi, M., Zarghami-Hamrah, S., Ghaedy, Y., & Mahmudnia, A. A quasi-transcendental approach for removing hierarchical teacher-student relation. *Policy Futures in Education*, *16*(3), 346-359. https://doi.org/10.1177/1478210317736205

Zhang, O & Hyland, K. (2021). Advice-giving, power and roles in theses supervisions. *Journal of Pragmatics*, *172*, 35-45. https://doi.org/10.1016/j.pragma.2020.11.002

APPENDIX 1

The Interview Protocol with EFL Learners

Section 1. Demographic and other contextual information

1. What is your name?
2. How long have you been studying English?
3. What is your occupation?
4. How often do you use English?

Section 2. Writing experience

1. During your study, what sorts of academic texts have you written so far?
2. How often do you write academic texts now?
3. Who do you usually picture in your mind when you are writing academic the texts?
4. If you have the picture of your target readers, are you aware of their expectations?
5. How do this picture and your awareness of their expectation influence the language that you use in the texts?

Section 3. Reasons for hype

1. Why did you use this particular word of phrase [i.e., the item identified as hype]? Are you aware that these words are promotional?
2. What were you trying to communicate or emphasize in this sentence?
3. Was the use of this word in this section intentional? If so, what were your intentions?
4. What do you expect from using this particular word or phrase?
5. Why do you choose to emphasize this part [i.e., the section in the text that is hyped], and not the other parts?
6. During the supervising/ review process, has any of the promotor/ editor/ reviewer commented on these words?
7. Did you use any tools such as Grammarly and Quillbot? Did you use thesaurus? To what extent these tools influence your choice of words in the text?

Section 4. Follow up questions

1. Is it important for authors to show readers the positive parts of their work?
2. As a writer of academic texts, what is your opinion about hypes?
3. What are the benefits of using these words for authors?
4. Are you aware of their implications? What are they?

APPENDIX 2

Example of Hypes Coding

From a teacher's point of view obtained from the needs analysis of this study, it was stated that they had ***very little***[1] experience with online learning. Finding an app that is suitable for the needs of the curriculum at school and tailored to students' needs are other difficulties that make them seldom use digital device as well as mobile applications for teaching. This is associated with the key issues found in a research on digital materials in the Australian context (Oakley, Pegrum, Faulkner, & Striepe, 2012). There are ***very limited***[2] apps that are specifically developed for a particular group of students with appropriate content, pedagogically focused, and linked to the curriculum and learning framework. Especially for EFL learners in general, a small number of English learning apps for a particular skill and specific users in a certain context is available in app stores. Nevertheless, the ***bombastic***[3] number of language learning apps users indicated a new learning interest through mobile applications. As digital technology will continually evolve, expanding the variety of language learning application is ***urgent***[4] to bridge the limitations of learning in the classroom and to suit learners' needs. [S6]

No.	Item	Hype	Category	Justification
1.	*very little*	x	-	The claim is a quotation from a participant from the pilot study.
2.	*very limited*	v	Research gap	The claim is not immediately supported by evidence.
3.	*bombastic*	v	Research topic	The claim promotes the relevance of the research topic, but is not immediately justified, carries exaggerated tone, and is replaceable with less promotional words such as *large* or *high*.
4.	*urgent*	v	Research topic	The claim promotes the urgency of the research topic.

APPENDIX 3

Example of Coding Framework for S3's Responses to Questions in Section 2 and 3

Raw data (participant quotes)	Key words	Initial codes	Final themes
"Obviously I would always picture who would be likely to read my thesis... I know that there are experts out there, including my supervisors, who will read my work"	Experts, supervisors, lecturers	Reader awareness	Audience awareness
"I'm afraid that I will be judged through my works... this is what makes me overthinking"	Afraid, judged, overthinking	Reader awareness, fear of judgement	Audience awareness
"When I wrote my thesis, the pictures of my supervisors came to mind... [compared to article writing,] the pressure was bigger when I wrote my thesis, because it would be very embarrassing to have it read by my supervisors"	Pressure, embarrassing, supervisors, gatekeeping	Authority, gatekeeping, fear of judgement	Supervisory influence
"So I was more careful when writing my thesis, compared to the article"	careful	Strategic writing	Supervisory influence
"For the word *'reveal',* I use it because based on my scholarly reading activities, it is largely used in academic texts... As I told you earlier that I often read and imitated the words that are often used in papers from highly reputable journals"	Scholarly reading, papers, reputable journals	Published text modeling, imitation, adhering to writing convention	Published writing conventions
"Moreover, when writing my thesis, I would read a lot of papers from Q1 journals, especially those about narrative inquiries. From these papers, I observed that many authors would highlight this value to the readers... I would imitate the practice and I believe it is a good idea because they are papers published in scopus-indexed journals"	Q1 journals, observed, imitate the practice, Scopus-indexed	Published text modeling, imitation strategy	Published writing conventions

https://doi.org/10.17323/jle.2025.18317

Navigating Motivation and Autonomy in Language Learning: Unveiling the Impact of Gradeless Assessment

Soroor Rostami [1], Mansooreh Hosseinnia [2]

[1] Agder University, Kristiansand, Norway

[2] Azad University of Bojnourd, North Khorasan, Iran

ABSTRACT

Background: Traditional grading systems often undermine intrinsic motivation and learner autonomy in language education, prompting interest in alternatives like gradeless assessment. While descriptive feedback and self-correction show promise in enhancing learner development, their impact on specific motivational constructs, such as ideal and ought-to L2 selves, remains underexplored, particularly in non-Western EFL contexts. This study addresses this gap by examining how gradeless assessment influences motivation and autonomy, contributing to global debates on formative assessment practices.

Purpose: This study examines the effects of a gradeless assessment approach, utilizing descriptive feedback and self-correction, compared to conventional grading, on the motivation (ideal and ought-to L2 selves) and autonomy of EFL learners. It aims to address the pedagogical challenge of designing assessments that foster sustained motivation and learner agency, particularly in diverse educational settings like Iran, to inform global language education practices.

Method: Forty-one upper-intermediate EFL students, selected via the Quick Placement Test, were randomly assigned to an experimental (gradeless, n=21) or control (graded, n=20) group for a reading and writing course. Both groups received identical instruction and materials from the same instructor, differing only in assessment: the experimental group received written descriptive feedback with self-correction opportunities, while the control group received numerical grades. Pre- and post-test questionnaires assessed motivation and autonomy, with data analyzed using independent-sample t-tests.

Results: The gradeless approach significantly enhanced ideal L2 self-motivation ($t = -2.70$, $p < .05$, $\eta^2 = .15$, 9.8% mean increase) and autonomy ($t = -2.06$, $p < .05$, $\eta^2 = .15$.1% mean increase), but reduced ought-to L2 self-motivation ($t = 3.05$, $p < .01$, $\eta^2 = 6.2$ % mean decrease), suggesting a shift toward ideal L2 self-motivation.

Conclusion: These findings highlight the potential of gradeless, feedback-oriented assessment to foster intrinsic motivation and autonomy in Iranian EFL contexts, suggesting a need to reconsider traditional grading in favor of formative strategies. However, due to the study's cultural and contextual specificity, generalizability should be approached cautiously, warranting further research across diverse settings.

KEYWORDS

gradeless assessment; descriptive feedback; self-correction; L2 motivation, autonomy

Citation: Rostami, S., & Hosseinnia, M. (2025). Navigating motivation and autonomy in language learning: unveiling the impact of gradeless assessment. *Journal of Language and Education, 11*(3), 127-137. https://doi.org/10.17323/jle.2025.18317

Correspondence:
Mansooreh Hosseinnia,
mansoorehhossinnia@yahoo.com

Received: November 10, 2023
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

Traditional grading systems, which emphasize numerical scores, have been widely critiqued for their psychological and pedagogical limitations. The grading system remains a focal point of controversy and robust debate (Cain, Medina, Romanelli, & Persky, 2022; Jaschik, 2009; Lim, 2024; McMorran & Ragupathi, 2020). Research indicates that grades have detrimental psychological effects (e.g., Goos, Pipa, & Peixoto, 2021; Linder, Gerdtham, & Heckley, 2023; Marquez & Garcia, 2023; Supiano, 2019), raising concerns about the credibility of the conventional "A" to "F" grading system, highlighting well-documented trends in

grade inflation patterns (Rojstaczer & Healy, 2012). Jaschik (2009) encapsulated this debate by arguing that grades often create a superficial semblance of coherence. These findings have prompted growing interest in alternative assessment approaches, such as gradeless assessment, which prioritizes descriptive feedback and self-correction to foster learner-centered environments. In English as a Foreign Language (EFL) contexts, where motivation and autonomy are critical for sustained language acquisition, such alternatives hold particular promise for addressing the limitations of conventional grading (Dörnyei, 2005; Lamb, 2010).

The detrimental effects of numerical grading have been well-documented. Butler (1988) found that grades, even when accompanied by comments, can negatively impact students' engagement and performance by fostering a performance-oriented mindset. Similarly, Pulfrey, Buchs, and Butera (2011) argue that grades often operate through external pressures, such as fear of failure or peer competition, which can undermine students' willingness to engage deeply with learning material. In contrast, descriptive feedback, which provides specific guidance on how to improve without assigning scores, has been shown to enhance student success by focusing on learning processes (Brookhart, 2008; Belton, 2022). Studies by Guskey and Lee (2013) and Sackstein (2015) further challenge the necessity of grades, suggesting that formative feedback can support learning without numerical evaluations. Despite these insights, the application of gradeless assessment in diverse EFL contexts remains underexplored, particularly in terms of its impact on specific psychological constructs like motivation and autonomy.

Motivation and autonomy are pivotal in language learning, shaping students' academic success and long-term engagement (Lerner et al., 2022; Pratomo, & Kuswati, 2022; Quesada et al., 2025; Rostami et al., 2015; Smith et al., 2023). It seems that these two psychological variables are influenced by the grading system (Mendoza, Yan, & King, 2023; Muho & Taraj, 2022; Wu et al., 2014; Zedan, 2021). Dörnyei's (2005) L2 Motivational Self System (L2MSS) provides a robust framework for understanding motivation, conceptualizing it through the ideal L2 self (learners' internalized aspirations as language users) and the ought-to L2 self (external obligations driving language learning). Similarly, Kluger and DeNisi's (1996) Feedback Intervention Theory (FIT) posits that feedback focused on task improvement, rather than performance evaluation, enhances learners' engagement and self-regulation. These frameworks suggest that assessment practices can significantly influence learners' motivational orientations and autonomous behaviors. For instance, descriptive feedback aligned with learners' goals may strengthen their ideal L2 self, while reducing reliance on external pressures associated with the ought-to L2 self. Likewise, autonomy-supportive assessments, which encourage self-monitoring and reflection, can foster learners' sense of control and responsibility (Lamb, 2010).

Despite the growing body of research on formative assessment, a critical gap remains in understanding its impact on L2 motivation and autonomy in non-Western EFL contexts. Most studies have focused on Western educational settings, where cultural and institutional norms differ significantly from collectivist societies like Iran, where grading systems dominate instructional practices (Tavallali & Marzban, 2015). In Iranian EFL classrooms, traditional grading often limits opportunities for learner agency, as students are conditioned to prioritize scores over learning processes (Rostami et al., 2015). Moreover, while L2MSS and FIT have been widely applied to study motivation and feedback, their integration in examining gradeless assessment in non-Western contexts remains limited. This gap is particularly significant given the unique cultural and educational challenges in Iran, where collectivist values and exam-driven curricula may influence how learners respond to alternative assessment practices.

This study addresses this gap by investigating the effects of a gradeless assessment approach, characterized by descriptive feedback and self-correction opportunities, compared to conventional grading, on the motivation (ideal and ought-to L2 selves) and autonomy of Iranian EFL learners. Drawing on L2MSS, it predicts that gradeless assessment will enhance ideal L2 self-motivation by aligning feedback with learners' internalized goals, potentially reducing ought-to L2 self-motivation by minimizing external pressures. FIT further suggests that descriptive feedback will promote autonomy by directing learners' attention to task improvement and self-regulation rather than performance evaluation. By exploring these constructs in an underexplored non-Western context, this study contributes to global discussions on assessment reform, L2 motivation, and autonomy-supportive pedagogy. It aims to provide empirical evidence to inform the design of assessment practices that foster sustained motivation and learner agency, offering implications for language education policies and teacher training worldwide.

LITERATURE REVIEW

Psychological Impacts of Traditional and Gradeless Assessment

Traditional grading systems in education have faced significant criticism for their psychological and pedagogical limitations. Scholars argue that numerical grades increase student anxiety, discourage risk-taking, and shift focus from learning to performance outcomes (Butler, 1988; Pulfrey et al., 2011). For instance, Butler (1988) found that grades, even when accompanied by comments, foster a perfor-

mance-oriented mindset, reducing engagement. Conversely, gradeless assessment, emphasizing descriptive feedback and self-correction, has shown promise in mitigating these issues. Sackstein (2015) reported that gradeless classrooms empowered lower-achieving students by reducing stigma, though high-achievers struggled to adapt. Similarly, Benz[1] (2019) found that gradeless assessment reduced anxiety and reliance on external rewards, fostering a learning-oriented environment. However, these studies, often qualitative and Western-centric, lack experimental rigor to establish causal links between gradeless assessment and psychological outcomes like motivation and autonomy, highlighting the need for more robust investigations.

Feedback and L2 Motivational Self System

Dörnyei's (2005) L2MSS provides a framework for understanding how assessment practices influence language learning motivation through the ideal L2 self (learners' internalized aspirations) and ought-to L2 self (external obligations). Research suggests that descriptive feedback aligns with the ideal L2 self by supporting learners' intrinsic goals, while grades reinforce the ought-to L2 self through external pressures (Butler & Nisan, 1986; Wu et al., 2014). For example, Butler and Nisan (1986) found that task-related feedback enhanced engagement and outperformed grade-based systems. Wu et al. (2014) further noted that descriptive feedback fosters a task-involving environment, unlike grades, which promote competition and anxiety. However, these studies rarely apply L2MSS to EFL contexts or examine how feedback types differentially impact ideal and ought-to L2 selves, leaving a gap in understanding their motivational dynamics in language learning.

Assessment and Autonomy in EFL Contexts

Learner autonomy, defined as the capacity to take responsibility for one's learning (Dickinson, 1987), is closely tied to assessment practices. Lamb (2010) distinguishes between "assessment of autonomy" (measuring autonomy) and "assessment for autonomy" (promoting it through self-monitoring and reflection). Descriptive feedback and self-correction, as autonomy-supportive strategies, encourage learners to engage actively in their learning process (Pishghadam et al., 2011). In the Iranian EFL context, where grading systems dominate, studies like Tavallali and Marzban (2015) highlight limited autonomy due to exam-driven curricula. Razavi (2016) found that descriptive evaluation in Iranian primary schools faced challenges, such as inadequate teacher training and weak student self-evaluation skills. These studies, while insightful, are often small-scale and qualitative, limiting their ability to generalize or address autonomy in relation to specific motivational frameworks like L2MSS in non-Western settings.

Methodological Gaps and the Need for Experimental Evidence

Despite the growing interest in gradeless assessment, significant methodological and contextual gaps persist. Most studies, such as Sackstein (2015, 2018[2]) and Spencer (2017)[3], rely on qualitative or observational designs, which, while rich in detail, cannot establish causal relationships between assessment practices and psychological outcomes. Moreover, their focus on Western contexts overlooks cultural nuances in non-Western EFL settings like Iran, where collectivist values and rigid grading systems shape learner experiences (Mortazavizadeh et al., 2017). No experimental study to date has examined the impact of gradeless assessment on learners' ideal and ought-to L2 selves and autonomy in non-Western EFL contexts using a theory-driven design. This study addresses these gaps by employing an experimental approach, grounded in L2MSS and FIT (Kluger & DeNisi, 1996), to investigate the effects of gradeless assessment (descriptive feedback and self-correction) versus traditional grading on Iranian EFL learners' motivation and autonomy. It poses the following research questions:

RQ1: Does gradeless assessment (using descriptive feedback) have a statistically significant impact on Iranian EFL learners' ideal and ought-to L2 selves?

RQ2: Does gradeless assessment (using descriptive feedback) have a statistically significant impact on Iranian EFL learners' autonomy?

METHOD

Participants

To achieve of the objectives of this study, 47 female students, aged between 21 and 33 (M=25.47, SD=3.19), were purposively selected from upper-intermediate level classes at a private language institute in Mashhad, Iran. The non-random selection aimed to ensure practical considerations. Six participants were excluded after the placement test, resulting in 41 homogenized individuals who were randomly divided into an experimental group (21 participants) and a control group (20 participants). The selection of up-

[1] Benz, G. (2019, October 31). *Going gradeless: A liberation from anxiety*. Retrieved from https://www.teachersgoinggradeless.com/blog/anxiety-benz

[2] Sackstein, S. (2018). Earning good grades versus learning. https://blogs.edweek.org/teachers/work_in_progress/2018/12/earning_good_grades_versus_learning.html

[3] Spencer, K. (2017). A new kind of classroom: No grades, no failing, no hurry. *The New York Times*. https://www.nytimes.com/2017/08/11/nyregion/mastery-based-learning-no-grades.html

per-intermediate students was informed by both practical and pedagogical considerations. Upper-intermediate learners possess sufficient linguistic competence to comprehend and respond to complex questionnaire items, ensuring reliable data collection, while their ongoing language development makes them particularly sensitive to motivational and autonomy-related interventions (Dörnyei, 2005). This level is theoretically suitable for examining shifts in motivation and autonomy, as learners are transitioning from guided to more self-directed learning, a critical phase for developing the ideal and ought-to L2 selves (Dörnyei, & Ryan, 2015). Informing participants about the aim of the research and the nature of the groups was handled with care to minimize potential biases. Participants were informed that they were part of a study exploring different teaching methods, emphasizing the importance of maintaining unbiased responses. The specifics of being in either a control or experimental group were intentionally withheld to prevent any awareness that might influence their answers.

Instruments

Two questionnaires served as both pre-tests and post-tests, chosen for their established validity and reliability within the Iranian context, as reported in prior research.

Motivational Self System Questionnaire

This questionnaire, translated and validated by Papi (2010), consists of 12 items encompassing two subscales: six items measure learners' ideal L2 self, and six items measure learners' ought-to L2 self. Participants rated their responses on a 6-point Likert scale ranging from «strongly disagree» to «strongly agree». Reliability and validity of this questionnaire have been analyzed and reported in previous studies in Iran (e.g., Ghanizadeh, Eishabadi, & Rostami,2015; Ghanizadeh & Rostami, 2015; Papi, 2010; Rostami, Ghanizadeh, & Ghonsooly, 2015). In the present study, Cronbach's reliability for ideal L2 self and ought-to L2 self were .82 and .91, respectively, indicating high reliability.

Learner Autonomy Questionnaire

The Learner Autonomy Questionnaire, designed and validated by Zhang and Li (2004), was used as both pre-test and post-test to assess participants' autonomy in learning English as a foreign language. Comprising 21 items rated on a five-point scale from 'Never' to 'Always', this questionnaire has demonstrated high content validity and reliability in the Iranian context (Nematipour, 2012). In this study, Cronbach's alpha reliability coefficient for this questionnaire was .75, indicating acceptable reliability. The questionnaire measures multiple dimensions of learner autonomy, including cognitive strategies (e.g., setting learning goals and selecting appropriate resources), behavioral engagement (e.g., initiating and sustaining learning tasks independently), and metacognitive planning (e.g., self-monitoring and evaluating progress).

Procedure

The researchers ensured the ethical considerations of the study, including informed consent and voluntary participation. The research was conducted over several weeks in a private language institute between September and October 2019, with two 1.5-hour sessions per week. It is worth mentioning that, to address ethical concerns, the study adhered to rigorous ethical standards by ensuring equal instructional content and time for both groups. Importantly, the researchers provided this course to students for free, covering the financial costs associated with study, emphasizing their dedication to a fair and unbiased investigation. The following steps were undertaken to conduct this research.

Homogenization of the Participants

The participant selection process began with the administration of the Quick Placement Test (QPT) as a placement test, targeting upper-intermediate students. Developed by Oxford University Press and Cambridge English for speakers of other languages (ESOL), the QPT is a valid and efficient English language proficiency test widely used for placement purposes. Forty-seven female students participated initially, but six were excluded as their score did not fall within the upper-intermediate range, leaving 41 homogenized participants. According to designers of the QPT test, participants who scored between 40 and 47 on the test are upper-intermediate. These homogenized participants were randomly assigned into either the experimental or control group with the same teacher assigned to both groups to minimize the teacher's potential impact on the results.

Control Group

In the control group, motivation and autonomy levels were assessed before the treatment via pre-test questionnaires. Both groups received identical instruction in reading and writing skills, with teacher-made tests administered for each session. The key distinction between the control and experimental groups was the assessment process. In the control group, students received grades for their test performance, accompanied by teacher-made tests with ten different topics, each requiring approximately 100 words of written response. However, these tests were evaluated using grades alone, without descriptive comments. Post-test questionnaires were administered after ten sessions.

Experimental Group

In the experimental group, pre-test questionnaires were used to assess motivation and autonomy levels. For this group, all 21 students received written descriptive comments, along with an opportunity for self-correction, on

teacher-made tests for each session. FIT was applied through the SE2R model, involving "summarize, explain, redirect and resubmit" (Barnes, 2015). This approach aimed to provide comprehensive feedback in line with the mastery learning approach. Implementing these four stages, the teacher first began by summarizing student work. This assist situates both learners and instructor when they refer to that piece of feedback; it describes the assignment that was attempted. Next comes the explanation. Here, the teacher provided evidence to explain the level of mastery that the students had achieved (Barnes, 2015). The level of mastery that students achieved depends on their demonstration of the specified learning outcomes. The instructor in this phase developed these results in collaboration with the students by allowing them time to reflect and discuss these outcomes with their peers and then re-wording them in more student-friendly language. When both the student and instructor agreed about the learning results, the instructor provided evidence directly from some parts of the work to mention which outcomes were met, which were not, and which need improvement. When the learners mastered all the learning objectives, then they did not proceed to the redirection step. When the instructor believed that their work needed improvement, they offered guidance or recommendations or prompts which encouraged the learners to work on special parts of tasks which do not yet show mastery. The instructor provided this redirection feedback to the learners, and they had a chance to revise their work with particular attention to the highlighted details. In the last phase, students resubmitted their work. To be clear, the final descriptive feedback using Mark Barnes' SE2R model included a sentence at the end which asked the learner to resubmit their work, once it has been revised (Clark, 2017). As stated before, in the experimental group, students' examinations were assessed using descriptive feedback by Mark Barnes' SE2R model which included a sentence instead of presenting grades. For example, the teacher wrote: "Please resend me when you've made the changes I asked you, so I can return to your text and re-evaluate it for you." Finally, after ten sessions, all 21 students of this group received the same motivation and autonomy questionnaires as post-tests to compare with the pre-test.

Data Analysis

Statistical Package for Social Sciences (SPSS 24) was used for analyzing data and computing normality, descriptive statistics, reliability analysis of the instruments, and independent-sample t-test. An independent sample t-test was used to make sure that the learners were at the same level of motivation and autonomy before the treatment. The homogeneity of variances for each combination of the groups was checked using Levene's test. To address the research questions concerning motivation and autonomy scores after treatment, another independent-sample t-test was employed.

Design

This study is categorized as quasi-experimental due to non-random selection of participants based on the QPT and random assignment into experimental and control groups. Both groups participated in pretest and post-test assessments. Dependent variables encompass students' ideal and ought-to-l2-selves' motivation and autonomy, while the independent variable involves teachers' adoption of the gradeless approach through descriptive feedback.

RESULTS

The Shapiro-Wilk test confirmed normal data distribution, with p-values above .05 for both groups in motivation and autonomy questionnaires. This indicates the appropriateness of employing parametric tests. The assumptions of the t-test were checked for both pre-test and post-test scores. Table 1 presents the descriptive statistics for both groups in the pre-test assessments of motivation and autonomy.

Table 1 shows the baseline pre-test data for 41 Iranian EFL learners across the experimental (n=21) and control (n=20) groups, which presents mean scores and standard deviations for the variables of ideal L2 self, ought-to L2 self, and autonomy. Specifically, the experimental group recorded a pre-test mean of 27.66 (SD = 3.91) for ideal L2 self and 27.38 (SD = 3.39) for ought-to L2 self, while the control group reported means of 26.90 (SD = 4.47) and 25.30 (SD = 4.29) respectively; for autonomy, the experimental group's pre-test mean was 52.76 (SD = 9.75) compared to 56.40 (SD = 9.65) for the control group.

Pre-Test Comparisons: Baseline Equivalence

To determine the significance of these differences, independent sample t-tests were conducted with results presented in Table 2.

As outlined in Table 2, no significant differences were observed between the groups concerning ideal L2 self (t= -.58, p= .56), ought-to L2 self (t= -1.72, p= .09) and autonomy (t= 1.20, p= .23) during the pre-test phase. This suggests that, with a 95% confidence interval, there were no significant differences between the mean scores of the control and experimental groups in terms of ideal L2 self, ought-to L2 self, and autonomy prior to the treatment.

Post-Test Differences: Treatment Effects

In order to answer the research questions, another independent sample t-test was used. Table 3 presents the descriptive statistics for both groups in the post-test assessments of motivation and autonomy.

The table presents the post-test data for the experimental and control groups, consisting of 41 Iranian EFL learners

Table 1

Descriptive Statistics of Two Groups in Motivation and Autonomy in the Pre-Test Assessments

Variable	Group	N	Mean	Std. Deviation	Std. Error Mean
Ideal L2 Self	Control	20	26.90	4.47	.99
	Experimental	21	27.66	3.91	.85
Ought-to L2 Self	Control	20	25.30	4.29	.95
	Experimental	21	27.38	3.39	.74
Autonomy	Control	20	56.40	9.65	2.15
	Experimental	21	52.76	9.75	2.12

Table 2

Results of the Independent Sample t-test of Motivation and Autonomy in the Pre-Test Assessments

Variable	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Ideal L2 Self	-.58	39	.56	-.76	1.31
Ought-to L2 Self	-1.72	39	.09	-2.08	1.20
Autonomy	1.20	39	.23	3.63	3.03

Table 3

Descriptive Statistics of Two Groups in Motivation and Autonomy in the Post-Test Assessments

Variable	Group	N	Mean	Std. Deviation	Std. Error Mean
Ideal L2 Self	Control	20	27.15	4.54	1.01
	Experimental	21	30.38	2.97	.64
Ought-to L2 Self	Control	20	28.90	4.02	.90
	Experimental	21	25.66	2.65	.57
Autonomy	Control	20	55.10	8.96	2.00
	Experimental	21	60.76	8.61	1.88

(experimental: n=21; control: n=20), across the variables of ideal L2 self, ought-to L2 self, and autonomy. For ideal L2 self, the experimental group recorded a post-test mean of 30.38 with a standard deviation of 2.97, while the control group reported a post-test mean of 27.15 with a standard deviation of 4.54. For ought-to L2 self, the experimental group's post-test mean was 25.66 with a standard deviation of 2.65, compared to 28.90 with a standard deviation of 4.02 for the control group. Regarding autonomy, the experimental group exhibited a post-test mean of 60.76 with a standard deviation of 8.61, whereas the control group's post-test mean was 55.10 with a standard deviation of 8.96.

Effect Size and Interpretation

To determine the significance of these differences, independent sample t-tests were conducted with results presented in Table 4. Eta squared was used to check the effect size.

As presented in Table 4, post-test comparisons revealed significant differences between the experimental (gradeless) and control (graded) groups across all measured outcomes. For ideal L2 self-motivation, the experimental group outperformed the control group ($t = -2.70$, $p = .01$, $\eta^2 = .15$), indicating a large effect size (Cohen, 1988). This

Table 4

Results of the Independent Sample T-Test of Motivation and Autonomy in the Post-Test Assessments

		t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference
Post-test	Ideal L2 Self	-2.70	39	.01	-3.23	1.19
	Ought-to L2 Self	3.05	39	.00	3.23	1.05
	Autonomy	-2.06	39	.04	-5.66	2.74

suggests a substantial pedagogical impact, as the gradeless approach markedly enhanced students' intrinsic aspirations to achieve language proficiency. For ought-to L2 self-motivation, the control group showed higher scores ($t = 3.05$, $p = .00$, $\eta^2 = .19$), also reflecting a large effect size, implying that traditional grading reinforced extrinsic obligations more effectively, though potentially at the cost of intrinsic motivation. For autonomy, the experimental group demonstrated greater improvement ($t = -2.06$, $p = .04$, $\eta^2 = .09$), with a medium effect size, indicating a meaningful increase in students' ability to self-direct their learning processes. These findings underscore the differential impact of gradeless assessment, involving descriptive feedback and self-correction, in fostering intrinsic motivation and autonomy while reducing reliance on external pressures in the Iranian EFL context

DISCUSSION

Main Findings and Theoretical Interpretations

In this study, the researchers conducted experimental research to examine the impact of implementing the gradeless method on Iranian EFL students' ideal and ought-to L2-selves motivation and autonomy. Grounded in theoretical considerations, the hypothesis suggested that providing descriptive feedback without grades would maintain or enhance students' motivation and autonomy. The findings revealed a significant increase in ideal L2 self-motivation and autonomy, alongside an unexpected decrease in ought-to L2 self-motivation, offering insights into the psychological effects of assessment practices in the Iranian EFL context. This study also found that providing numerical grades enhances ought-to L2 self-motivation at the expense of ideal L2 self-motivation. These results directly address the research questions by highlighting how gradeless assessment reshapes motivational constructs. Drawing on the L2 Motivational Self System (L2MSS), the rise in ideal L2 self reflects learners' strengthened internalized aspirations, as theorized by Dörnyei (2005), particularly when external grading is removed. Self-Determination Theory (SDT) further supports this, indicating that descriptive feedback enhances competence and relatedness, boosting intrinsic motivation. This extends Wu et al. (2014), who linked task-involving environments to motivation, by demonstrating these effects in a collectivist Iranian context. In addition, the statistically significant increase in autonomy aligns with Lamb's (2010) concept of assessment for autonomy. He defined assessment for autonomy as any assessment for which the priority in its design and practice is to serve the purpose of promoting pupils' autonomy. Jácome (2012) also reported similar findings emphasizing the role of teacher-student partnership assessment using feedback and self-evaluating skills in promoting learner autonomy. Furthermore, according to Lamb (2010), providing feedback not only enhances students' language learning but also their autonomy and motivation. He also continued that to enable students to become more autonomous learners, teachers should encourage them to involve not only what they have learned, but also how they learned it through a meaningful assessment such as providing feedback.

Unexpected Findings and Explanations

Regarding the impact of the type of assessment on students' motivation, the findings align with FIT and support previous research across various contexts (e.g., Butler & Nisan, 1986; Lipnevich & Smith, 2008; Stefanou & Parkes, 2003; Taylor & Nolen, 2008; Wu et al., 2014). The findings suggest that when students receive constructive feedback focused on improvement rather than grades, they are more likely to be driven by a genuine desire to attain their ideal language proficiency. Wu et al. (2014) emphasized the close relationship between classroom assessment and motivation, highlighting the positive impact of task-involving environments, diverse assessment tasks, descriptive feedback, and clear learning objectives on intrinsic motivation. However, providing a competitive classroom environment, instructing based on the test, and providing students grades are associated with extrinsic motivation. In line with their findings, our study supports the idea that creating a task-involving classroom environment, and providing descriptive feedback enhances students' ideal L2-self motivation. Therefore, the results of our study contribute to the existing body of knowledge by providing additional evidence of the positive impact of gradeless assessment using descriptive feedback on students' motivation, aligning with the principles highlighted by Wu et al. (2014). Additionally, Stefanou and Parkes (2003) highlighted the potential impact of the grading system on learner behavior and motivation, and

our findings substantiate this claim. Specifically, our results emphasize the significant effect of assessment types on motivation, reinforcing the importance of adopting assessment practices that prioritize descriptive feedback over grades. In connecting our findings to the broader literature, we corroborate the view that an overemphasis on grades, as indicated in prior studies (Coffield, 2012; Matthews & Noyes, 2014; McMorran & Ragupathi, 2020), indeed leads to detrimental impacts on the learning process. The present study further validated that learners who engage in self-correction assessments exhibit higher levels of motivation. Stiggins (2005) mentioned that "If students play even a small role in setting the (learning achievement) target...we can gain considerable motivational and therefore achievement benefits" (p.244). Therefore, teachers can increase students' motivation by engaging them in the process of assessment, which can be done by the self-correction assessment method. Stiggins (2005) also emphasized the importance of self-correction which increases students' responsibility for improving their sense of control over their success and accordingly enhances their motivation. The decrease in the mean score of ought-to L2 self-motivation in the experimental group from 27.38 in the pre-test to 25.66 in the post-test warrants deeper analysis. This unexpected finding might be attributed to several factors that could have influenced participants' perceptions and expectations during the study. Potential reasons could include variations in how students interpreted or internalized descriptive feedback, the nature of the self-correction process, or even individual differences in response to the gradeless assessment method. The concept of ought-to L2 self, as derived from Dörnyei's L2 Motivational Self System model, represents the traits learners believe they should possess due to various duties, obligations, or expectations. While the decrease in ought-to L2 self-motivation may initially seem counterintuitive, it aligns with the theoretical perspectives suggesting that traditional grading systems, emphasizing extrinsic rewards and external obligations, may negatively impact intrinsic motivation (Lee, Lallie, & Michaelides, 2023). This finding demonstrates a potential trade-off in gradeless systems, where increased autonomy and intrinsic drive (ideal L2 self) may reduce compliance-based motivation (ought-to L2 self). Using L2MSS, the de-emphasis on grades likely weakens the ought-to self, which thrives on external pressures. SDT complements this, suggesting that removing extrinsic regulation may destabilize ought-to motivation unless supported by autonomy-enhancing practices. In Iran's collectivist culture, this decline may reflect reduced societal or familial expectations, though our data lacks direct evidence to confirm this, highlighting a need for further investigation.

Implications for EFL Assessment Practices

This section discusses these findings in light of existing literature, addresses their cultural implications, and considers potential limitations and adverse effects of the gradeless approach. Our findings imply that gradeless assessment, with its focus on descriptive feedback, can enhance EFL learning by boosting ideal L2 self and autonomy, challenging traditional grading paradigms. This extends Butler and Nisan (1986), who highlighted feedback's role in motivation, by showing its efficacy in a non-Western context. Practically, educators might integrate self-correction to foster agency, though balancing autonomy with accountability remains critical. This study suggests self-correction as a valid proxy for learner agency, offering a novel approach to assessment design. At the institutional level, language programs could explore integrating gradeless methods to create environments that support sustainable motivation, though careful implementation is needed to address diverse learner needs.

Cultural Considerations

The findings put forward several implications for curriculum design, teacher training, and assessment policy, particularly in contexts seeking to foster sustainable motivation and independent learning. Language instructors are encouraged to adopt descriptive feedback practices to nurture students' intrinsic motivation and self-directed learning. Teacher training programs should emphasize formative assessment techniques, equipping educators to implement feedback-driven approaches effectively. In the Iranian EFL context, where traditional grading often limits learner agency, the gradeless method's focus on self-correction likely fostered metacognitive strategies, aligning with Oxford's (2011) framework for autonomous learning. However, constructs like autonomy and the ought-to L2 self are culturally embedded, and their interpretation may differ in other educational systems. For instance, in individualistic cultures like Scandinavian countries, where autonomy is already emphasized, the impact of gradeless assessment may be less pronounced (Hofstede, 2011). Conversely, in assessment-driven East Asian systems, students may rely more heavily on grades for motivation, potentially limiting the efficacy of descriptive feedback (Cheng & Curtis, 2010). The Iranian context, characterized by collectivist values, likely shapes how gradeless assessment impacts motivation, with the decline in ought-to L2 self reflecting a reduced emphasis on external approval, a cultural norm in collectivist settings.

Limitations and Directions for Future Research

While the gradeless approach yielded positive outcomes for ideal L2 self-motivation and autonomy, potential negative effects must be considered. The decrease in ought-to L2 self-motivation suggests that some students, particularly those reliant on external validation, may experience uncertainty or reduced motivation without grades. High-performing students accustomed to competitive environments may find the lack of grades demotivating, as grades often serve as a benchmark for achievement (Pulfrey et al., 2011). Additionally, less confident learners may feel anxious without the clear structure of numerical feedback, as noted by Sackstein (2015), who observed challenges for high-achiev-

ing students transitioning to gradeless systems. In the Iranian context, where grades are deeply tied to academic and social status, this shift could exacerbate uncertainty for some learners. Research indicates that gradeless learning can lead to uncertainty and identity loss among high-performing students accustomed to competitive environments, potentially destabilizing extrinsic motivation (Kjærgaard, et al., 2023). Furthermore, learners in collectivist cultures may struggle to adapt to feedback-driven systems without clear performance metrics, which can exacerbate anxiety and reduce motivation (McNall & Gravelin, 2024). These findings suggest that while gradeless systems have potential, their success depends on careful implementation and cultural adaptation. Several limitations affect this study's generalizability. The cultural specificity to Iranian EFL learners, a sample size of 41, and the short-term intervention without follow-up data constrain the findings. Reliance on self-report instruments may also introduce bias. Future research should employ longitudinal designs to assess long-term effects, use mixed-methods for deeper insights, and compare outcomes across diverse educational cultures to address these gaps and refine gradeless assessment practices.

CONCLUSION

This study explored the impact of gradeless assessment, utilizing descriptive feedback and self-correction, on the motivation and autonomy of Iranian EFL learners. The findings suggest that descriptive feedback and self-correction practices can significantly enhance learners' ideal L2 selves and foster greater autonomy. These preliminary results underscore the potential of formative assessment to reorient language education toward meaningful learning rather than grade-driven outcomes. Looking forward, these findings pave the way for further exploration of alternative assessment models in language education. Future research should investigate the long-term effects of gradeless assessment across diverse cultural and proficiency contexts to enhance its applicability. By building on these insights, educators and researchers can advance learner-centered pedagogies that foster motivation and autonomy, contributing to the evolution of EFL assessment practices globally. This study offers new insights into gradeless assessment by demonstrating its potential to enhance L2MSS dimensions in under-researched collectivist contexts like Iran, while revealing the challenge of balancing autonomy with accountability in assessment systems.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Soroor Rostami: conceptualization; formal analysis; investigation; methodology; resources; writing – original draft; writing - review and editing.

Mansooreh Hosseinnia: conceptualization; investigation; methodology; resources; writing – original draft; writing - review and editing.

REFERENCES

Barnes, M. (2015). *Assessment 3.0: Throw out your gradebook and inspire learning.* Corwin.

Belton, T. R. (2022). *The impact of grading systems on middle school students' summative mathematics achievement test scores* [Unpublished doctoral dissertation]. Regent University.

Brookhart, S. M. (2008). *How to give effective feedback to your students*. Association for Supervision and Curriculum Development.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*(1), 1-14. https://doi.org/10.1111/j.2044-8279.1988.tb00874.x

Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology, 78*(3), 210-216. https://doi.org/10.1037/0022-0663.78.3.210

Cain, J., Medina, M., Romanelli, F., & Persky, A. (2022). Deficiencies of traditional grading systems and recommendations for the future. *American Journal of Pharmaceutical Education, 86*(7). https://doi.org/10.5688/ajpe8850

Cheng, L., & Curtis, A. (2010). *English language assessment and the Chinese learner.* Routledge.

Clark, M., R. (2017). *Reducing the use of and emphasis on grades in high school physics classes* [Unpublished MA thesis]. University of Toronto.

Coffield, F. (2012). To grade or not to grade. *Adult Learning, 23*(4), 38–39.

Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge University Press.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum.

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited.* Routledge.

Ghanizadeh, A., Eishabadi, N., & Rostami, S. (2015). Motivational dimension of willingness to communicate in L2: The impacts of criterion measure, ideal L2 self, family influence, and attitudes to L2 culture. *International Journal of Research Studies in Education, 4*(5), 13-24. https://doi.org/10.5861/ijrse.2015.1261

Ghanizadeh, A., & Rostami, S. (2015). A Dörnyei-inspired study on second language motivation: A Cross-comparison analysis in public and private contexts. *Psychological Studies, 60*(3), 292-301. https://doi.org/10.1007/s12646-015-0328-4

Goos, M., Pipa, J., & Peixoto, F. (2021). Effectiveness of grade retention: A systematic review and meta-analysis. *Educational Research Review, 34*, 100401. https://doi.org/10.1016/j.edurev.2021.100401

Guskey, T. R., & Lee, A. J. (2013). *Answers to essential questions about standards, assessments, grading, and reporting*. Corwin.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture, 2*(1), 8. https://doi.org/10.9707/2307-0919.1014

Jácome, E. P. (2012). Promoting learner autonomy through teacher-student partnership assessment in an American high school: A cycle of action research. *Profile Issues in TeachersProfessional Development, 14*(2), 145-162.

Jaschik, S. (2009). *Imagining college without grades.* Inside Higher Education.

Kjærgaard, A., Mikkelsen, E. N., & Buhl-Wiggers, J. (2023). The gradeless paradox: Emancipatory promises but ambivalent effects of gradeless learning in business and management education. *Management Learning, 54*(4), 556-575. https://journals.sagepub.com/doi/pdf/10.1177/13505076221101146

Kluger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 2*, 254-284. https://doi.org/10.1037/0033-2909.119.2.254

Lamb, T. (2010). Assessment of autonomy or assessment for autonomy? Evaluating learner autonomy for formative purposes. In A. Paran & L. Sercu (Eds.), *Testing the untestable in language education* (pp. 98-119). Multilingual Matters.

Lee, D., Lallie, H. S., & Michaelides, N. (2023). The impact of an employee's psychological contract breach on compliance with information security policies: Intrinsic and extrinsic motivation. *Cognition, Technology & Work, 25*(2), 273-289. https://doi.org/10.1007/s10111-023-00727-5

Lerner, R. E., Grolnick, W. S., Caruso, A. J., & Levitt, M. R. (2022). Parental involvement and children's academics: The roles of autonomy support and parents' motivation for involvement. *Contemporary Educational Psychology, 68*, 102039. https://doi.org/10.1016/j.cedpsych.2021.102039

Lim, K. (2024). Assessing beyond grades: Unravelling the implications on student learning and engagement in higher education. *Assessment & Evaluation in Higher Education, 49*(5), 665-679. https://doi.org/10.1080/02602938.2024.2314703

Linder, A., Gerdtham, U. G., & Heckley, G. (2023). *Adolescent mental health: Impact of introducing earlier compulsory school grades.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4499202

Lipnevich, A. A., & Smith, J. K. (2008). *Response to assessment feedback: The effects of grades, praise, and source of information.* Educational Testing Service.

Marquez, E., & Garcia, S. (2023). Keeping calm and staying balanced: exploring the academic pressures faced by engineering students to attain high grades and their impact on mental health. *Paper presented at 2023 ASEE Annual Conference & Exposition, Baltimore, Maryland.* https://peer.asee.org/43908

Matthews, R., & A. Noyes. (2014). To grade or not to grade: Balancing formative and summative assessment in Post-16 teacher trainee observations. *Journal of Further and Higher Education, 40*(2), 247–261. https://doi.org/10.1080/0309877x.2014.953456

McMorran, C., & Ragupathi, K. (2020). The promise and pitfalls of gradeless learning: Responses to an alternative approach to grading. *Journal of Further and Higher Education, 44*(7), 925-938. https://doi.org/10.1080/0309877x.2019.1619073

Mendoza, N. B., Yan, Z., & King, R. B. (2023). Supporting students' intrinsic motivation for online learning tasks: The effect of need-supportive task instructions on motivation, self-assessment, and task performance. *Computers & Education, 193*, 104663. https://doi.org/10.1016/j.compedu.2022.104663

Mortazavizadeh, H., Nili, M., Nasr Isfahani, A., & Hassani, M. (2017). Teachers' lived experiences about teaching-learning process in multi-grade classes. *Journal of Education and Learning, 6*(4), 354-363. http://doi.org/10.5539/jel.v6n4p354

Muho, A., & Taraj, G. (2022). Impact of formative assessment practices on student motivation for learning the English language. *International Journal of Education and Practice, 10*(1), 25-41. https://doi.org/10.18488/61.v10i1.2842

McNall, L. A., & Gravelin, C. R. (2024). An exploratory study of student perceptions of ungrading in upper-level psychology courses. *Teaching of Psychology,* 00986283241293413. https://doi.org/10.1177/00986283241293413

Nematipour, M. (2012). A study of Iranian EFL learners' autonomy level and its relationship with learning style. *English Linguistics Research, 1*(1), 126-136. https://doi.org/10.5430/elr.v1n1p126

Oxford, R. L. (2011). Strategies for learning a second or foreign language. *Language Teaching, 44*(2), 167-180. https://doi.org/10.1017/S0261444810000492

Papi, M., (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System, 38,* 467-479. https://doi.org/10.1016/j.system.2010.06.011

Pishghadam, R., Hashemi, M. R., & Kermanshahi, P. N. (2011). Self-correction among Iranian EFL learners: An investigation into their preferences for corrective feedback. *Journal of Language Teaching & Research, 2*(5), 957-962. https://doi.org/10.4304/jltr.2.5.957-962

Pratomo, H. W., & Kuswati, Y. (2022). The effect of teacher motivation on student achievement in Islamic senior high school. *International Journal of Educational Qualitative Quantitative Research, 1*(2), 16-22. https://doi.org/10.58418/ijeqqr.v1i2.17

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: the mediating role of autonomous motivation. *Journal of Educational Psychology, 103,* 683-700. https://doi.org/10.1037/e549962013-069

Quesada, L. A., Oviedo, F. C., Nunez, S. V., Charpentier, K. A., & Atencio, M. F. Z. (2025). Extrinsic and intrinsic motivation and its impact on students' performance in the EFL classroom. *Ciencia Latina Revista Científica Multidisciplinar, 9*(1), 5670-5682. https://doi.org/10.37811/cl_rcm.v9i1.16244

Razavi, A. (2016). Three components of descriptive evaluation: Feedback, self-assessment, and peer-assessment. *Journal of Instruction and Evaluation, 9*(36), 49-74.

Rojstaczer, S, & Healy, C. (2012). Where A is ordinary: the evolution of American college and university grading, 1940–2009. *Teachers College Record, 114*(7), 1-23. https://doi.org/10.1177/016146811211400707

Rostami, S., Ghanizadeh, A., & Ghonsooly, B. (2015). External factors affecting second language motivation: The role of teacher burnout and family influence. *Iranian Journal of Applied Linguistics, 18*(2), 165-187. https://doi.org/10.18869/acadpub.ijal.18.2.165

Sackstein, S. (2015). Hack learning: Hacking assessment. Cleveland Times 10.

Smith, Z. R., Flax, M., Becker, S. P., & Langberg, J. (2023). Academic motivation decreases across adolescence for youth with and without attention-deficit/hyperactivity disorder: Effects of motivation on academic success. *Journal of Child Psychology and Psychiatry, 64*(9), 1303-1313. https://doi.org/10.13056/acamh.24604

Stefanou, C., & Parkes, J. (2003). Effects of Classroom assessment on student motivation in fifth-grade science. *The Journal of Educational Research, 96*(3), 152-162. https://doi.org/10.1080/00220670309598803

Stiggins, R. J. (2005). *Student-involved assessment for learning* (5th ed.). Merrill/Prentice Hall.

Supiano, B. (2019). *Grades can hinder learning. What should professors use instead*. Chronicle of Higher Education.

Tavallali, E., & Marzban, A. (2015). Becoming autonomous learners through self-regulated learning. *Journal of Applied Linguistics and Language Research, 2*(3), 72-83

Taylor, C. S. & Nolen, S. B. (2008). *Classroom assessment: Supporting teaching and learning in real classrooms* (2nd ed.). Pearson Education, Inc.

Wu, Y., Cheng, L., & Bettney, E. (2014). Exploring the relationships between assessment and motivation: Perspectives from teacher candidates. *Paper presented at the 58th International Council on Education for Teaching World Assembly, University of Ontario Institute of Technology, Oshawa, Ontario, Canada.*

Zedan, R. (2021). Student feedback as a predictor of learning motivation, academic achievement, and classroom climate. *Education and Self-Development, 16*(2), 27-46. https://doi.org/10.26907/esd.16.2.03

Zhang, L.X., & Li, X. X. (2004). A comparative study on learner autonomy between Chinese students and west European students. *Foreign Language World, 4,* 15-23.

https://doi.org/10.17323/jle.2025.22211

Detecting LLM-Generated Text with Trigram–Cosine Stylometric Delta: An Unsupervised and Interpretable Approach

Egor Salnikov [1], **Anastasiya Bonch-Osmolovskaya** [2]

[1] HSE University, Moscow, Russia

[2] Vinogradov Institute of Russian Language, Moscow, Russia

ABSTRACT

Background: Contemporary methods for detecting synthetic text, including model-specific detectors and transformer-based classifiers, often rely on intensive training or on features tied to particular language models, which restricts their generalizability to unfamiliar LLMs and diverse domains.

Purpose: To advance text attribution research by introducing a stylometry-based approach that utilizes trigram-based cosine delta as a lightweight and interpretable metric for distinguishing LLM-generated texts from human-written texts, irrespective of the underlying generation strategy.

Method: A corpus of Russian diary entries was compiled, encompassing both authentic human-written texts and synthetic counterparts generated through few-shot prompting and finetuned LoRA models. To evaluate the effectiveness of the proposed approach, multiple stylometric-delta variations were examined, integrating uni-, bi-, and trigram features with Manhattan and cosine distance metrics.

Results: The evaluation demonstrated that the trigram–cosine delta consistently achieved the highest performance across experimental conditions, reaching an Adjusted Rand Index of approximately 0.70. This markedly surpassed both the finetuned RuModernBERT baseline (ARI ≈ 0.28) and the classic unigram-based delta (ARI ≈ 0.53). Importantly, the method proved effective not only within the Russian diary corpus but also when applied to the RuATD benchmark, where it successfully separated human-authored and machine-generated texts and produced coherent clustering of related model families.

Conclusion: The findings confirm that trigram–cosine stylometric delta offers a robust, interpretable, and computationally efficient strategy for detecting LLM-generated texts across diverse generation strategies, including few-shot prompting and finetuning. By capturing discourse-level stylistic cohesion, the method advances beyond surface fluency and provides a scalable, unsupervised alternative to classifier-based detectors. While current validation is limited to Russian diaries and selected generation models, the approach demonstrates clear potential for broader application across domains, languages, and emerging state-of-the-art LLMs.

KEYWORDS

LLM; synthetic text detection; stylometry; Burrows delta; cosine delta

Citation: Salnikov, E., & Bonch-Osmolovskaya, A. (2025). Detecting LLM-generated text with Trigram–Cosine Stylometric Delta: An unsupervised and interpretable approach. *Journal of Language and Education, 11*(3), 138-151. https://doi.org/10.17323/jle.2025.22211

Correspondence:
Egor Salnikova,
esalnikov@hse.ru

Received: August 13, 2024
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

The rapid development of large language models (LLMs) has significantly advanced natural language generation, but it has also intensified concerns regarding the authenticity and reliability of textual content across multiple domains such as education, journalism, and digital archiving (Aich et al., 2022; Bender et al., 2021; Huang et al., 2023; Sahoo et al., 2024; Gurioli et al., 2025). Despite ongoing progress (Fraser et al., 2025), the ability to reliably differentiate between human-authored and machine-generated texts remains limited (Wang, S. et al.,

2024; Wang, Y. et al., 2023). This challenge has direct implications for academic integrity, information credibility, and the security of communication systems (Gressel et al., 2024; Roy et al., 2024).

A wide range of detection methods has been proposed, yet none of them offers a universal and sufficiently robust solution (Wu et al., 2025; Tang et al., 2024, Sadasivan et al., 2023). Model-dependent strategies, such as watermarking, introduce identifiable patterns during text generation (Kirchenbauer et al., 2023; Zhao et al., 2022), but these techniques cannot be applied retroactively to external or proprietary models. Distribution-matching approaches achieve precision when logit information is accessible (Gehrmann et al., 2019; Mitchell et al., 2023), but they fail when applied to black-box or previously unseen models. Supervised classifiers, including BERT-based detectors and systems like DetectGPT, demonstrate strong performance on in-domain data (Antoun et al., 2023; Bethany et al., 2024), although they require large volumes of labeled data, act as black boxes with limited interpretability, and generalize poorly to out-of-domain or finetuned outputs (Bakhtin et al., 2019; Li et al., 2023; Shamardina et al., 2022). Feature-based unsupervised methods avoid reliance on labeled datasets and provide higher interpretability, but their effectiveness declines when the text domain or stylistic register shifts significantly (Ma et al., 2023; Muñoz-Ortiz et al., 2023; Guo et al., 2023).

Stylometric techniques offer an alternative pathway grounded in authorship attribution studies. Burrows' delta and its variations rely on distance-based comparisons of frequent words and other stylistic features (Burrows, 2002; Hoover, 2004; Craig & Kinney, 2009). These methods have been successfully applied across languages and registers (Rybicki & Eder, 2011; Eder et al., 2016) and remain valued for their simplicity, interpretability, and low computational requirements. Recent research has suggested that even basic stylometric measures can differentiate human texts from LLM outputs in certain contexts (Rebora, 2023; Salnikov & Bonch-Osmolovskaya, 2023; Wang, S. et al., 2024). However, such studies often focus on zero-shot prompting and fail to address more advanced generation strategies such as finetuning, which can more effectively mimic authorial style and therefore complicate detection (Schuster et al., 2020; Zhang et al., 2024; Przystalski et al., 2025).

This gap highlights the need for an unsupervised detection approach that preserves interpretability and scalability while demonstrating robustness against different generation strategies. The present study therefore sets out to systematically assess the potential of stylometric delta methods for distinguishing between human-written and LLM-generated texts. Specifically, the study aims to determine whether trigram-based cosine delta, in comparison with alternative n-gram and distance metric configurations, provides a reliable, interpretable, and computationally efficient solution for text attribution across both few-shot prompting and fine-tuning scenarios.

LITERATURE REVIEW

Model-Dependent Approaches

Model-dependent approaches include methods such as watermarking, which introduce traceable patterns into generated outputs by manipulating token selection or probability distributions (Kirchenbauer et al., 2023; Zhao et al., 2023). These techniques demonstrate high effectiveness when applied to text produced by models under the researcher's control. Nevertheless, their reliance on pre-embedded signals makes them unsuitable for detecting content generated by external or proprietary systems, since signals cannot be retroactively incorporated into outputs created by third-party models. A related group of strategies is distribution matching, exemplified by the methods proposed by Gehrmann et al. (2019) and Mitchell et al. (2023). These approaches compare statistical regularities between known model outputs and test samples. Although they can achieve high precision when full access to the model is available, their performance decreases substantially when applied to black-box systems or previously unseen models, which limits their applicability in more general detection settings.

Supervised Classifier-Based Detection

A popular strategy involves training classifiers like BERT, RoBERTa, or DetectGPT using labeled samples (Shamardina et al., 2022; Li et al., 2023; Antoun et al., 2023; Bethany et al., 2024; Emi & Spero, 2024). For example, the Pangram Text classifier outperforms DetectGPT and many commercial tools in accuracy and generalization tasks. Though effective in-domain, these models carry significant drawbacks. Firstly, they require large quantities of labeled synthetic and human-written text, which is expensive and time-consuming to generate. Secondly, they often act as black boxes, offering little interpretability for their decisions. And lastly, they struggle to generalize to new domains or adaptation techniques like finetuning or fewshot prompting.

Unsupervised, Feature-Based Methods

These methods rely on linguistically interpretable features such as function-word frequencies, syntactic complexity, lexical richness, and other stylistic indicators (Ma et al., 2023; Muñoz-Ortiz et al., 2023; Zaitsu & Jin, 2023; Guo et al., 2023; Fröhling & Zubiaga, 2021, Chhatwal & Zhao, 2025). They do not require training data, and their outputs can be directly traced back to specific linguistic features, which ensures transparency of interpretation (Kumarage & Liu, 2023; Opara, 2024; Weerasinghe et al., 2025). However, when applied across domains that differ in genre, topic, or authorial style,

the relevance and effectiveness of such features may decline substantially, which reduces the reliability of these methods in cross-genre applications and in contexts involving evolving LLM outputs.

Stylometry-Based Techniques

Stylometry-based techniques have their origins in authorship attribution, where stylometric delta was introduced as a method for comparing ranked-frequency profiles, most often based on function words, across different corpora (Burrows, 2002). Subsequent studies have demonstrated its applicability to synthetic text detection. For example, Rebora (2023) employed Burrows' delta to distinguish ChatGPT-generated texts from Dickensian prose in a zero-shot setting. Although the study provided valuable insights, its scope was narrow, since it focused on a single stylistic register and did not consider texts produced through finetuning or domain-adapted LLMs. The principal strength of this approach lies in its simplicity and interpretability, which makes it appealing as an unsupervised method. Nevertheless, its broader applicability has not been thoroughly examined, particularly in contexts involving more sophisticated generation strategies such as finetuning or few-shot prompting. Moreover, Schuster et al. (2020) demonstrated that stylometry can be vulnerable when LLMs produce stylistically homogeneous content, such as uniform misinformation, in which case the method fails to discriminate between sources. This limitation underscores the necessity of further evaluating and refining stylometric techniques for contemporary detection tasks.

Comparative Analysis and Research Gap

In order to contextualize the present study, it is necessary to compare the main categories of existing detection methods along key dimensions that define their practical relevance. These dimensions include the degree of access required to underlying language models, the dependence on labeled training data, the interpretability of detection outcomes, the adaptability across domains and languages, and the robustness under advanced generation strategies such as finetuning or few-shot prompting. Table 1 provides a structured overview of these categories, enabling a systematic assessment of their relative strengths and limitations.

The comparison reveals critical limitations across existing approaches. Model-dependent methods such as watermarking and distribution matching presuppose either direct control over text generation or access to internal model parameters. While effective in controlled conditions, these techniques are inapplicable when dealing with outputs from external or proprietary systems (Kirchenbauer et al., 2023; Mitchell et al., 2023). Supervised classifiers demonstrate strong performance in in-domain settings (Antoun et al., 2023; Bethany et al., 2024), yet they are constrained by their reliance on large volumes of labeled data, lack of transparency, and limited generalizability to unseen domains or finetuned models (Bakhtin et al., 2019; Li et al., 2023). Feature-based unsupervised methods rely on linguistically interpretable indicators such as lexical richness or syntactic complexity, but they tend to degrade under shifts in genre or authorial style (Ma et al., 2023; Muñoz-Ortiz et al., 2023; Guo et al., 2023).

Stylometric approaches, particularly Burrows' delta and its modifications, combine methodological simplicity with interpretability and have a long tradition in authorship attribution (Burrows, 2002; Hoover, 2004; Rybicki & Eder, 2011; Eder et al., 2016). Initial attempts to adapt them for synthetic text detection indicate that they can differentiate between human-authored and LLM-generated texts in restricted settings (Rebora, 2023; Salnikov & Bonch-Osmolovskaya, 2023; Wang, S. et al., 2024). However, their performance has not been systematically tested in scenarios involving advanced generation techniques such as finetuning or few-shot prompting, and prior work has shown their vulnerability when models produce stylistically uniform outputs, for example in the case of misinformation (Schuster et al., 2020).

Table 1

Comparative Analysis of Detection Methods

Method Category	Model Access Needed	Labeled Data Required	Interpretability	Domain/Language Adaptability	Robustness to Finetuning/Few-Shot Prompting
Watermarking	Yes (owner-controlled)	No	Low	Low (requires pretraining integration)	Poor (cannot be applied post hoc)
Distribution Matching	Yes (requires logits)	No	Moderate	Limited to known models	Poor (fails with unseen or finetuned models)
Supervised Classifiers	No	Yes	Low (black-box)	Moderate (performance drops out-of-domain)	Poor (overfits and degrades on new models)
Unsupervised Feature-Based	No	No	High	Moderate (fragile across genres and domains)	Low (weak cross-domain performance)
Stylometry (e.g., Delta)	No	No	High	Historically applied across registers	Underexplored (particularly with finetuned and few-shot LLMs)

Taken together, these observations underscore a clear methodological gap. There remains a need for an unsupervised detection technique that is both interpretable and computationally efficient, yet at the same time robust to stylistic variation induced by different LLM generation strategies. To address this gap, the present study systematically evaluates stylometric delta with varying feature sets (unigrams, bigrams, trigrams) and distance metrics (Manhattan and cosine). The analysis is applied to human-written and synthetic diary entries generated through few-shot prompting and finetuning. This design makes it possible to assess the interpretability, scalability, and resilience of stylometry-based detection, thereby positioning it as a practical and transparent alternative to model-dependent or classifier-centric approaches.

METHOD

Research Design

The central hypothesis of the current study is that stylometric delta, when applied to the clustering of natural and synthetic texts, can serve as an effective basis for unsupervised detection of machine-generated content. To test this assumption, the experimental design was structured around systematic comparisons between human-authored diary entries and texts generated by large language models through different strategies, including few-shot prompting, finetuning with LoRA adapters, and Direct Preference Optimization (DPO). In addition, a finetuned transformer classifier, RuModernBERT, was employed as a supervised baseline in order to benchmark the performance of the proposed unsupervised approach.

The overall aim of the design was to assess whether stylometric delta can provide both efficiency and robustness in unsupervised detection tasks. By comparing across generation strategies and benchmarking against a supervised baseline, the study sought not only to evaluate the precision of the method but also to test its potential for scalability and generalization in broader applications of synthetic text detection.

Corpus and Data Preparation

The data for the study were extracted from the Prozhito corpus**,** a large archive of Russian diaries and ego-documents. To ensure diversity and representativeness, samples were constructed with consideration of authorial identity and subdomain characteristics. Randomized subsets of the corpus were divided into training and test partitions. The test sets were reserved as authentic examples of natural writing, while the training sets were used both for prompting and for the finetuning of generative models. Several Mistral-based LLMs were then trained on these subsets to approximate the stylistic properties of the original diaries. Using both pretrained LLMs and the finetuned Mistral variants, we generated synthetic corpora designed to mimic natural diaries across multiple stylistic domains.

Domain Choice

Russian diaries were chosen as the primary domain of analysis because of their distinctive linguistic and stylistic properties. Previous research has emphasized the heterogeneity of diary writing, which resists formal unification and encompasses considerable variation in length, narrative style, communicative function, and the cultural and social backgrounds of authors (Bogdanova, 2008). This inherent diversity renders the diary genre an especially challenging and therefore informative target for evaluating text detection methods. Restricting the study to diaries provided a clearly delimited genre while simultaneously allowing the examination of stylistic heterogeneity within that boundary.

Dataset Construction

In total, eleven datasets were constructed, each consisting of 500 texts and divided into three major categories: one category representing authentic human-written diaries and two categories representing synthetic texts. Within the natural category, three datasets were selected and further partitioned into training and test subsets. Training sets were used both for model training and for few-shot generation, whereas test sets were reserved exclusively for stylometric delta experiments as gold-standard examples of natural texts.

The three natural datasets included:

(1) PRISHVIN, consisting of diary entries authored by Mikhail Mikhailovich Prishvin;
(2) RANDOM, composed of entries from twenty randomly selected authors spanning a period of three centuries;
(3) WORK, derived from randomly chosen “work diaries,” which describe professional activities and are characterized by reduced emotional content (Vorobeva et al., 2024).

The WORK dataset was assembled using a classifier trained on materials available at GitHub repository[1]. This dataset includes contributions from more than one hundred authors, which substantially diminishes the strength of individual authorial signals and increases stylistic noise. Consequently, this domain was considered the most difficult for stylometric delta to discriminate. For the classification task, the rubert-tiny2 model was finetuned on one thousand texts randomly sampled across historical periods, achieving a classification precision of approximately 0.94.

[1] https://github.com/laidhimonthegreen/prozhito_thesis?utm_source=chatgpt.com

Synthetic Data Generation

Few-Shot Prompting

To collect synthetic datasets for the second category, we prompted four LLMs to write diaries as if they were Prishvin. We used a few-shot learning method, providing the model with two original Prishvin's diaries for each prompt (Table 2). This approach has already demonstrated better results than regular prompting, making it easier for models to copy the particular style. LLMs used for pseudo-Prishvin diaries generation are: ChatGPT (GPT-3.5), GigaChat (GigaChat-Pro), Vicuna-1.5, Saiga/Mistral.

Finetuning with LoRA and DPO

The last category of our data was composed of synthetic diaries generated by models trained on natural diaries of the first category. As a base model we used the original Mistral-7B (Jiang et al., 2023). First, we trained two models on Prishvin's diaries using two different approaches: LoRA-adapter (Hu et al., 2021) and Direct Preference Optimization (DPO) (Rafailov, 2024).

Every model was trained to write diary entries based on an author specified in the system prompt, an instruction to write a diary entry, and a date (Table 3).

For all generations the fixed set of parameters was used:

(1) maximum length = 1024 tokens,
(2) temperature = 0.7,
(3) top_p = 0.9,
(4) seed = 42.

Training LoRA-adapters proved to be a sufficient method, which helps to shift the model's distribution closer to the target domain, while requiring only some model's parameters (<1%) to be updated.

On the other hand, DPO is often used to explicitly set a model's preferences. This approach is similar to Reinforcement Learning from Human Feedback (RLHF) (Knox & Stone, 2011), used by OpenAI to train GPT, and thus requires at least a pair of inputs during training. In our case every training sample contained two inputs: "chosen" (an original Prishvin diary entry) and "rejected" (a randomly chosen

Table 2

Few-Shot Generation Example

Prompt	Напиши дневниковую запись за 10 января 1921 от лица Михаила Михайловича Пришвина: Святочный сон: в лесу на сосне привесил я портрет своей Козочки, волки стали собираться к портрету, а я их стрелять. Разрушитель не дикий человек, а недоучка, тот, кто, выкрав, напр., из физического кабинета спираль Румкорфа, знает, что ее можно приспособить для закуривания папирос. – 1 Р. Снег мелкий, но частый и сильный ветер. Ночью на 29-е – буря и + 5 Р. Напиши дневниковую запись за 16 января 1921 от лица Михаила Михайловича Пришвина: Искусство есть способность человека изображать предмет своей веры и любви (Христ.). (Возрождение – предмет наслаждений.) Вера без дел мертва, а вера без любви – зла и есть, кажется (надо подумать), основа величайших злодейств. Зло существует на кредит любви. Сатира, пародия – должники искусства. Напиши дневниковую запись за 26 января 1921 от лица Михаила Михайловича Пришвина:
Generation result (GPT 3.5)	Сегодня я прочитал долгое и принципиальное письмо от одного из моих адвокатов. Мне предстоит сразу же начать подготовку к защите своей позиции, поэтому я встал рано и занялся прочтением литературы по юридической теории. Я размышляю над тем, как использовать закон в мою пользу, чтобы достичь моей цели. Я верю, что правая юридическая аргументация поможет мне достичь моей цели. Да будет мне помощью мудрость и да будет мне поддержкой судьба!

Table 3

Example of Prompt Used for Finetuning

System Prompt	Ты - Михаил Михайлович Пришвин. Ты ведешь дневник.
Prompt	Напиши дневниковую запись за 4 марта 1942 года от лица Михаила Михайловича Пришвина:
Generation result	Продолжение «Весны» в духе Гёльцнера: «после революции» и так далее. Как надо заплатить налог: ...

diary entry from the RANDOM dataset). DPO was expected to change the model's preferences and allow it to better capture Prishvin's style by gradually distinguishing it from other authors. However, later experiments showed that the DPO model struggled to learn Prishvin's style, being easily detectable even by the simplest delta.

In contrast, the model trained with LoRA-adapter showed peculiar results. Despite the low quality of generated texts, classic Burrows delta was unable to distinguish them from their natural counterparts. Seeing such promising results, we decided to train another two models using the data from RANDOM and WORK datasets, respectively.

Stylometric Delta Methodology

As a primary method for generated text detection we chose stylometric delta. Stylometry is the application of linguistics designed to evaluate the individual style of an author. Stylometry methods proved to be efficient in authorship attribution of texts (Hoover, 2004; Craig & Kinney, 2009). As author attribution is similar to text classification as "natural" or "generated," stylometry methods could also be productive for our study.

Since the emergence of Burrows' delta in 2002, this method (and its variations) is frequently used for authorship attribution (Stamatatos, 2009). It uses z-scores of normalized word frequencies to calculate distances between texts. Variants of delta are usually made by altering distance measures or normalization procedures (Eder et al., 2016; Argamon, 2008). Delta is now the most established measure in authorship attribution (Rybicki & Eder, 2011).

Previous works (Rebora, 2023; Salnikov & Bonch-Osmolovskaya, 2023) showed that even classic Burrows' delta often demonstrates positive results in distinguishing generated texts. However, these studies are limited by reliance on zero-shot prompting, without LLM finetuning. We aimed to overcome this by employing few-shot prompting, LoRA-adapters, and DPO.

Each dataset was truncated to 6000 tokens, which is sufficient for delta to perform as expected. For each configuration, the top 1000 most frequent n-grams were extracted from the union of all datasets, following best practices (Kestemont, 2014; Evert et al., 2017). Burrows' delta was then calculated as the mean absolute (Manhattan) or cosine distance between vectors. Experiments were repeated for unigrams, bigrams, and trigrams.

Supervised Baseline

As a supervised baseline, we used RuModernBERT**,** a finetuned transformer classifier trained on one thousand randomly sampled texts. This allowed benchmarking of the stylometric delta approach against a state-of-the-art supervised model, ensuring robust evaluation of efficiency, scalability, and generalization potential.

RESULTS

Baseline Classifier Performance

The first stage of the evaluation focused on the performance of the finetuned RuModernBERT model in distinguishing natural diary entries from synthetic texts generated through few-shot prompting, LoRA finetuning, and Direct Preference Optimization (DPO). The classifier was trained in a supervised manner and was designed to assign each text to one of four categories: natural, few-shot, LoRA, or DPO. The quality of clustering, measured by the Adjusted Rand Index (ARI), reached only 0.28, which reflects a relatively weak correspondence between the predicted clusters and the true labels. A closer inspection of the results revealed frequent misclassification of Vicuna and GigaChat outputs as LoRA-generated texts, which illustrates the limited capacity of the model to capture structural distinctions among generation strategies.

Classic Delta (Manhattan and Unigrams)

In the next step, we applied an unsupervised approach based on the classic Burrows' delta. Hierarchical clustering was performed using Manhattan distances calculated over the one thousand most frequent unigrams, following z-score normalization. This configuration produced substantially more coherent clusters and was particularly successful in isolating few-shot outputs irrespective of their generative origin (see Dendrogram 1). Under this setting, the ARI increased to 0.53, which represents a notable improvement compared with the supervised baseline.

Cosine Delta with Unigrams

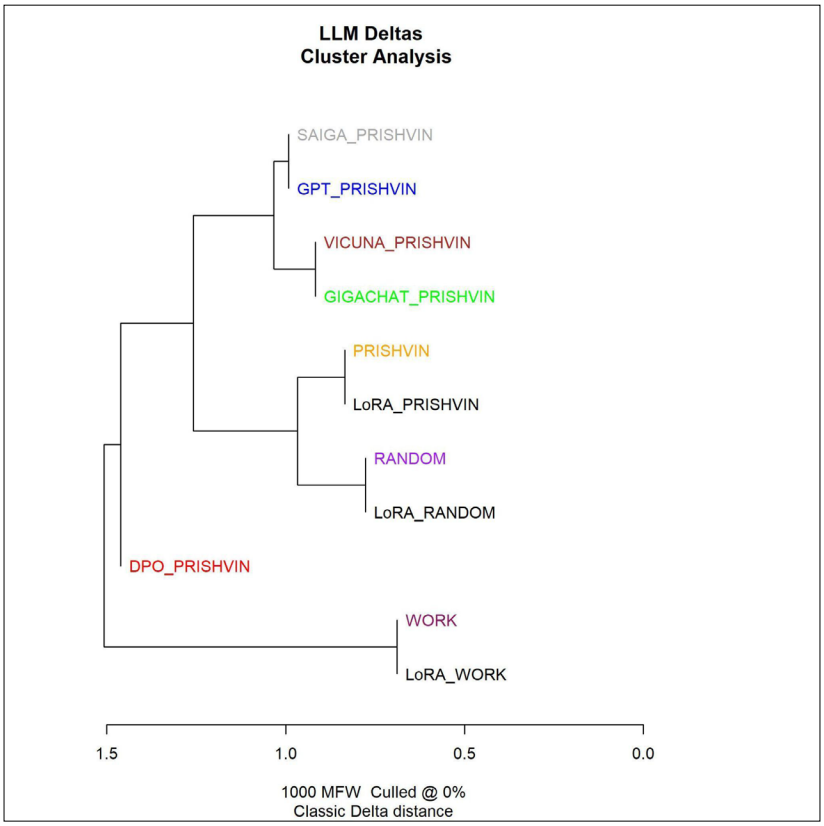
When Manhattan distance was replaced with cosine distance on the same unigram feature set, the overall results deteriorated. The Adjusted Rand Index dropped to approximately 0.12, indicating poor clustering performance. Although certain distinctions were visible, such as the partial separation between the PRISHVIN and LoRA_PRISHVIN datasets, the method failed to provide reliable clustering in more challenging domains. In particular, in the WORK dataset, LoRA-generated and natural texts were grouped together in an ambiguous manner (see Dendrogram 2).

Cosine Delta with Trigrams

To better capture stylistic nuances, we then employed cosine distance on the top 1000 trigrams. This configuration delivered the highest performance: the dendrogram clearly separated synthetic and human-authored texts across all domains (Dendrogram 3). Only a minor misclassification

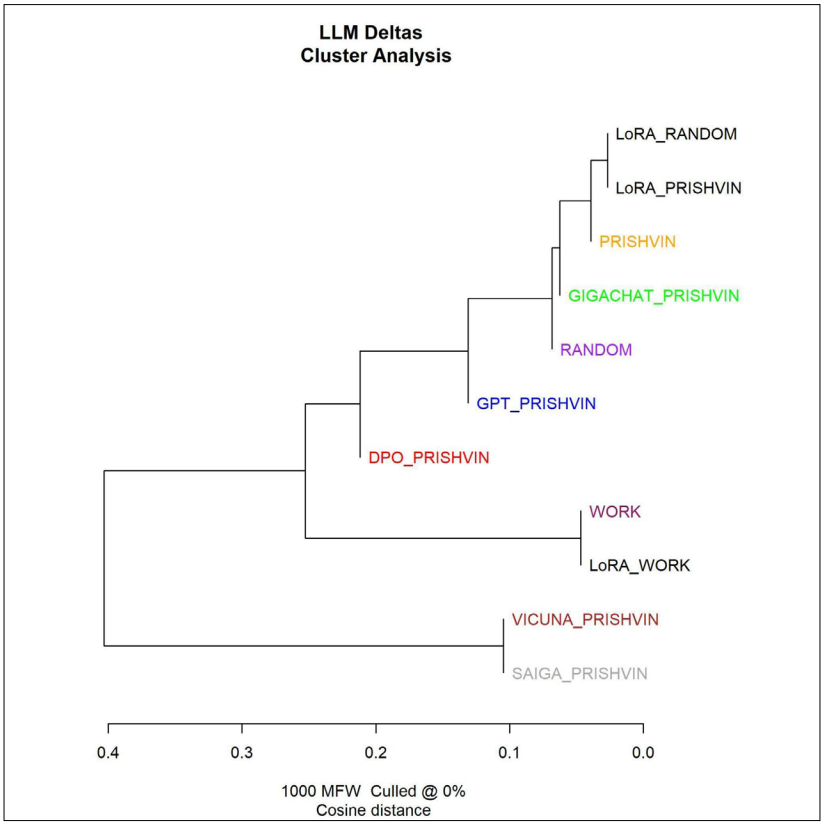
Dendrogram 1

Classic Delta (Manhattan Distance and Unigrams)



Dendrogram 2

Cosine Delta and Unigrams



occurred where LoRA_PRISHVIN clustered with fewshot outputs. Overall, ARI reached approximately 0.70, the highest of all methods.

Comparative Performance

Comparing ARI scores across methods shows a clear progression: the RuModernBERT classifier scored 0.28, Classic Delta with unigrams reached 0.53, Cosine Delta with unigrams dropped to 0.12, while Cosine Delta with trigrams achieved a peak score of 0.70. The Adjusted Rand Index values for all methods are presented in Table 4.

Out of Domain Evaluation

To further examine the general applicability of the trigram cosine delta, we conducted an evaluation using data from the RuATD 2022 generated text detection competition. This dataset included outputs from thirteen different language models, which allowed us to test the method in a more heterogeneous setting. The experiment was carried out in a multiclass configuration using only the validation subset. Since delta methods tend to perform more reliably on longer inputs, we aggregated all texts produced by each model, as well as the human-authored texts, into single datapoints. These datapoints were then compared using the one thousand most frequent trigrams extracted from the combined validation dataset.

The results are visualized in Dendrogram 4, which illustrates the clustering structure produced by the method. The dendrogram provides a clear separation between human-written and machine-generated texts and further demonstrates coherent subgrouping of language models according to their architectural families, such as the ruGPT3 and mT5 clusters. This outcome supports the potential of the proposed method to generalize across datasets with varying domains and stylistic characteristics.

DISCUSSION

Previous stylometry-based research has primarily concentrated on zero-shot detection of prompt-generated synthetic texts. Such studies demonstrated that basic stylometric measures, including Burrows' delta, are capable of distinguishing between human-authored and machine-generated texts, although often only at a superficial level (Rebora, 2023; Salnikov & Bonch-Osmolovskaya, 2023). However, these works did not explore the resilience of stylometry when applied to more advanced generation strategies, in particular finetuning, which is designed to approximate individual authorial style more closely. The results of the present study extend this line of research by demonstrating that stylometric detection remains effective even when LLMs are finetuned, provided that the method incorporates sufficiently complex features and an appropriate distance metric. Specifically, the use of trigram features in combination with cosine distance produced consistently enhanced discrimination.

At the same time, the evaluation carried out in this study has clear limitations. The analysis was restricted to Russian diary texts and to a relatively narrow set of generation techniques. Consequently, the extent to which the trigram cosine delta can be generalized to other genres, such as news reporting or academic writing, or to languages with markedly different syntactic structures, remains uncertain. Its performance against newer LLMs that are capable of producing highly polished and stylistically nuanced outputs also requires further validation. Earlier research has shown that stylometry may fail in contexts where machine-generated texts are deliberately homogenized, for example in the case of misinformation that is stylistically uniform (Schuster et al., 2019). For this reason, the findings presented here should be regarded as encouraging but preliminary, and they must be interpreted within the methodological and domain-specific constraints of the study.

A comparison with alternative methods further illustrates the advantages of the proposed approach. The BERT-based classifier struggled to capture subtle stylistic distinctions, whereas the classical delta method achieved moderate success, particularly in isolating few-shot text clusters. In contrast, the trigram-based cosine delta consistently produced robust separation between natural and synthetic texts, thereby confirming its utility and supporting the central hypothesis of the research. Qualitative inspection of the outputs further corroborates these results. Few-shot generations tend to display overly simplified phrasing and

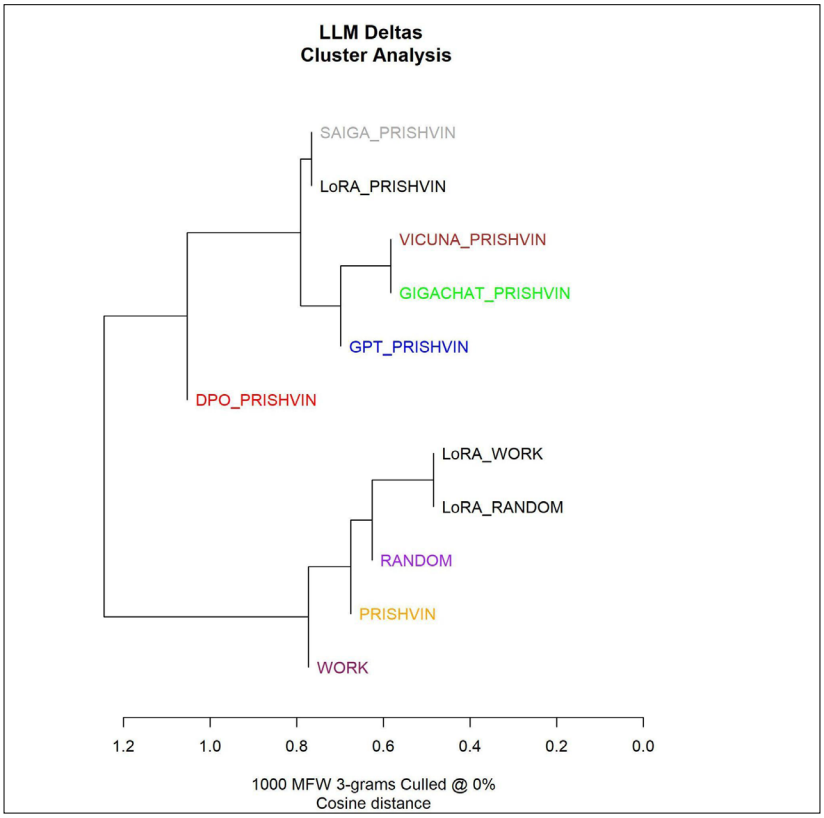
Table 4

Adjusted Rand Scores for Different Methods

Method	Adjusted Rand Score
RuModernBERT (Baseline)	0.2775
Classic Delta	0.5343
Cosine Delta and Unigrams	0.1204
Cosine Delta and Trigrams	0.6983

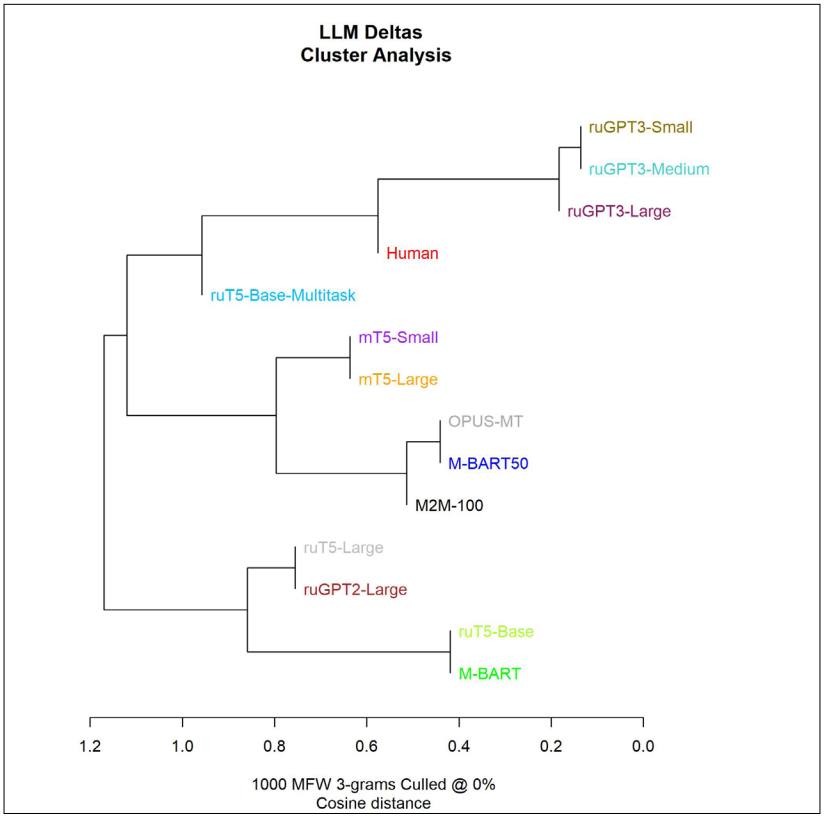
Dendrogram 3

Cosine Delta and Trigrams



Dendrogram 4

Cosine Trigram Delta and RuATD Data



weak stylistic alignment, while LoRA outputs more clearly attempt to replicate authorial style but often lack structural coherence (see Tables 5 and 6). The effectiveness of the trigram-based delta highlights the importance of capturing discourse-level cohesion, including connective structures and syntactic markers, as shown in Table 7. These elements play a critical role in accurate stylistic classification and appear to represent an area where LLMs still exhibit detectable shortcomings.

A central strength of the trigram-based approach lies in its capacity to capture discourse-level cohesion patterns that extend beyond surface fluency and reflect deeper syntactic complexity. In Russian, for example, repeated connective phrases such as «о том, что» function as markers of cohesion and are difficult for LLMs to reproduce consistently. This limitation is particularly evident in the case of finetuned models, which often generate texts that appear locally fluent but lack structural nuance at the global level (Muñoz-Ortiz et al., 2023). By incorporating trigrams, the method effectively encodes these cohesion cues, allowing the detection system to take advantage of weaknesses in the ability of LLMs to replicate authentic discourse-level style.

Equally important is the transparency of the proposed method. Because it relies on ranked and distance-based comparisons of n-grams, the approach aligns with broader research advocating for explainable linguistic techniques in the detection of synthetic text. This feature allows users to trace stylistic anomalies directly, rather than relying exclusively on opaque classification outcomes. Such interpretability enhances the method's practical applicability, especially in settings where auditability, adaptability to diverse domains, and methodological clarity are critical.

The findings of the present study confirm the core hypothesis that stylometric delta, when configured with trigram features and cosine distance, provides a reliable basis for the unsupervised detection of LLM-generated texts. The trigram–cosine configuration achieved an Adjusted Rand Index of approximately 0.70, which represents a substantial improvement over both the RuModernBERT baseline (ARI ≈ 0.28) and the classical unigram-based delta (ARI ≈ 0.53). The advantage was particularly pronounced in complex domains such as multi-author "WORK" diaries, where stylistic noise makes discrimination especially challenging.

Limitations

Nevertheless, our study has limitations. We tested only Russian diary texts, which limits generalization to other genres or languages. Similarly, emerging LLMs (GPT-4o/o1, LlaMA 3.1, Qwen 2.5) might better emulate individual style; our method's robustness against such models remains unproven.

Future work should evaluate wider text types – news, reviews, academic prose – in multiple languages, and measure sensitivity to variables like text length, authorial diversity, and intentional style copying. Testing against stronger foundation models will further assess the generality and lasting relevance of the trigram cosine delta approach.

Table 5

Examples of Texts Generated by Larger LLM Using Fewshot Learning (SAIGA_PRISHVIN) and Smaller LLM with fFnetuned LoRA-Adapter (LoRA_PRISHVIN).

SAIGA_PRISHVIN	LoRA_PRISHVIN
Сегодня я проснулся рано и решил посмотреть на восход солнца. Я вышел на улицу и увидел, что солнце еще не поднялось над горизонтом. Мне было интересно посмотреть, как оно будет взойти сегодня.	У нас доктор сословный, с его зелеными зубами, мертвым лицом, а смуты нет во мне. Я знал, что это он сам даст мне сестрам свое мнение и потом не позвонит по этому поводу: «Документов у нас нет».

Note. Despite producing mostly coherent text SAIGA_PRISHVIN fails to copy Prishvin individual style. On the other hand LoRA_PRISHVIN while also generating some nonsense, better captures Prishvin's individual style.

Table 6

Examples of original and synthetic texts of WORK domain

WORK	LoRA_WORK
Проходило оперативное совещание (по селектору). Энергосистемы повсеместно вышли на максимум энергонагрузок. К 1 января 1975 г. нагрузка будет (как обычно) снижаться, так как многие предприятия страны выполнили, а некоторые и перевыполнили свои планы.	Вышел в эфир первый выпуск телепредставления «Детские годы» о Горчакове. Длинное название ленты: «На смену литорине». Подобные темы очень не любят цензура и у внутрипартийной оппозиции.

Note. Due to domain's specifics (mostly simplicity) stylistic differences between original and generated texts are not very obvious.

Table 7

Randomly Picked Segments of Lemmatized Original and Generated Texts from WORK Domain

WORK	LoRA_WORK
нов заменить брусилов. володя вебер выслать соня 15 руб. и от (брат) колоть я сигара. * 15. * у я сильно заболеть правый бок. фрайнд приехать из крб. уговориться	прилепить полянка и клевер. панно делать и я поставить за она. пять колесо 16 колесо. сухой погода и добавлять зятка. яровый косово сносить. броса вода 12 бараболь по овес дать хуба 2 барабаль. выкорочный колесо 5. серени в прошед. гр. прох.
с он на случай надобность воспользоваться ссуда по 1200 акция по 250 р. и поручить он когда быть в берлин попытаться составить синдикат о покупка у я 10 т. акция. жалование і [половина] февра [аля] – 434. после урок бесконечный заседание – о архитектурный рисование – бродский, синайский, юон, я – против тырс. наконец получить согласие бродский на приглашение белкин. вечер телефон с он и с тата. (1) заканчивать прополка кукуруза. (2) гавриленко, тинавая, шаповалов – не работать . некрасово. событь весь день	< продолжение отменно стыдный. не оставить ничто никакой вывод ...> < первый сентябрь 1925 год > учить паренек овес спир 3 день. < один время перескакано, продолжиться следующий раз уже полный смысл ответ и возращение ... > петь вчера в поле попасть. пелед навеять. сирника в прошлый. гр. закончить 132 кубок. провотка и ножица. былед. гляз. одежда и рубашка. сено к 10 декабрь. быть весь день дома. много закурить. работать на пик

Note. Frequent trigrams are marked with color. The fragment «с он и с» is actually two frequent trigrams – «с он и» , «он и с». Trigrams of original text are much closer to connectors than those of generated text, thus embedding more syntactic information.

CONCLUSION

This study addressed the challenge of detecting synthetic text across outputs generated through finetuning and few-shot prompting by introducing a lightweight and interpretable detection method based on trigram cosine stylometric delta. The proposed approach achieved an Adjusted Rand Index of approximately 0.70, which substantially outperformed both the finetuned RuModernBERT classifier (ARI ≈ 0.28) and the classical unigram-based delta measure (ARI ≈ 0.53). These results provide strong confirmation of the central hypothesis regarding the effectiveness of higher-order n-gram features combined with cosine distance for unsupervised text detection.

Unlike classifier-centric or transformer-based detectors that depend on extensive labeled data and often operate as black-box systems, the present method relies on transparent and linguistically interpretable features. This characteristic not only enhances its suitability for academic scrutiny but also facilitates practical deployment in real-world applications. The findings further extend earlier stylometric research by demonstrating that trigram-based features capture discourse-level cohesion and syntactic complexity, and that these properties remain discriminative even in the context of finetuned models.

At the same time, the study has several limitations. The evaluation was restricted to Russian diary texts and a specific set of generation strategies, which constrains the scope of generalization to other genres, languages, and more advanced large language models. Future work should therefore test the approach across a wider range of textual domains, stylistic registers, and contemporary LLMs in order to validate its broader applicability.

In conclusion, this research contributes a scalable, transparent, and effective unsupervised technique for the detection of LLM-generated content. By combining interpretability with computational efficiency, the trigram cosine stylometric delta represents a meaningful advancement in the field of synthetic text detection and provides a foundation for further methodological development.

AI DISCLOSURE STATEMENT

During the preparation of this manuscript, the authors used GPT-4o and GPT-5 exclusively for language editing and stylistic refinement. All AI-generated suggestions were reviewed, revised, and approved by the authors. No AI tools were used in the design, analysis, interpretation, or presentation of the research data; all scientific conclusions are based on the authors' independent work, for which they take full responsibility.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTIONS

Egor Salnikov: conceptualization; software; formal analysis; methodology; investigation; visualization; writing – original draft; writing – review & editing.

Anastasiya Bonch-Osmolovskaya: conceptualization; data curation ; methodology; project administration; supervision; resources; writing – review & editing.

REFERENCES

Aich, A., Bhattacharya, S., & Parde, N. (2022). Demystifying neural fake news via linguistic feature-based interpretation. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 6586-6599). International Committee on Computational Linguistics

Antoun, W., Mouilleron, V., Sagot, B., & Seddah, D. (2023). Towards a robust detection of Language Model generated text: Is ChatGPT that easy to detect? In *Actes de CORIA-TALN 2023* (vol. 1, pp. 14–27). ATALA.

Argamon, S. (2008). Interpreting Burrows's delta: Geometric and probabilistic foundations. *Literary and Linguistic Computing, 23*(2), 131-147. https://doi.org/10.1093/llc/fqn003

Bakhtin, A., Gross, S., Ott, M., Deng, Y., Ranzato, M. A., & Szlam, A. (2019). *Real or fake? Learning to discriminate machine from human generated text.* arXiv:1906.03351. https://doi.org/10.48550/arXiv.1906.03351

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. https://doi.org/10.1145/3442188.344592

Bethany, M., Wherry, B., Bethany, E., Vishwamitra, N., & Najafirad, P. (2024*). Deciphering textual authenticity: A generalized strategy through the Lens of Large Language Semantics for detecting human vs. machine-generated text*. arXiv:2401.09407. https://doi.org/10.48550/arXiv.2401.09407

Bogdanova, E. (2008). Language features of diary genre. *Philological Sciences*, *1*(1), 28-33. https://doi.org/10.30853/filnauki

Vorobeva, V., Bonch-Osmolovskaya, A., Kryukov, A., & Podriadchikova, M. (2025). Distant reading of Soviet diaries. *Digital Scholarship in the Humanities, 40*(3), 956–968, https://doi.org/10.1093/llc/fqaf038

Burrows, J. (2002). 'Delta': A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing, 17*(3), 267-287. https://doi.org/10.1093/llc/17.3.267

Chhatwal, G. S., & Zhao, J. (2025). Multitask learning for authenticity and authorship detection. *Electronics, 14*(6), 1113. https://doi.org/10.3390/electronics14061113

Craig, H., & Kinney, A. F. (2009). *Shakespeare, computers, and the mystery of authorship.* Cambridge University Press.

Dong, Q., Li, L., Dai, D., Zheng, C., Wu, Z., Chang, B., Sun, X., Li, L., & Sui, Z. (2022). *A survey on in-context learning*. arXiv:2301.00234. https://doi.org/10.48550/arXiv.2301.00234

Eder M., Rybicki J., Kestemont M. (2016). Stylometry with R: A package for computational text analysis. *The R Journal, 8*(1), 107-121. https://doi.org/10.32614/rj-2016-007

Eder, M., Kestemont, M., & Rybicki, J. (2013). Stylometry with R: A suite of tools. *Digital Humanities 2013: Conference Abstracts* (pp. 487–489). University of Nebraska.

Emi, B., & Spero, M. (2024). *Technical report on the pangram ai-generated text classifier*. arXiv:2402.14873. https://doi.org/10.48550/arXiv.2402.14873

Crothers, E., Japkowicz, N., & Herna, L. (20023). Machine-generated text: A comprehensive survey of threat models and detection methods. *IEEE Access* (vol. 11, pp. 70977-71002). IEEE. https://doi.org/10.1109/ACCESS.2023.3294090

Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C., & Vitt, T. (2017). Understanding and explaining Delta measures for authorship attribution. *Digital Scholarship in the Humanities, 32*(suppl_2), ii4-ii16. https://doi.org/10.1093/llc/fqx023

Fraser, K. C., Dawkins, H., & Kiritchenko, S. (2025). Detecting AI-generated text: Factors influencing detectability with current methods. *Journal of Artificial Intelligence Research, 82*, 2233-2278. https://doi.org/10.1613/jair.1.16665

Fröhling, L., & Zubiaga, A. (2021). Feature-based detection of automated language models: Tackling GPT-2, GPT-3 and Grover. *PeerJ Computer Science, 7*, e443. https://doi.org/10.7717/peerj-cs.443

Gehrmann, S., Strobelt, H., & Rush, A. M. (2019). Gltr: Statistical detection and visualization of generated text. *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 111–116). Association for Computational Linguistics. https://doi.org/10.18653/v1/p19-3019

Gressel, G., Pankajakshan, R., & Mirsky, Y. (2024). Exploiting LLMs for scam automation: A looming threat. In *Proceedings of the 3rd ACM Workshop on the Security Implications of Deepfakes and Cheapfakes* (pp. 20-24). ACM. https://doi.org/10.1145/3660354.3660356

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). *How close is Chatgpt to human experts? Comparison corpus, evaluation, and detection*. arXiv:2301.07597.

Gurioli, A., Gabbrielli, M., & Zacchiroli, S. (2025). Is this you, LLM? Recognizing AI-written programs with multilingual code stylometry. In *2025 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)* (pp. 394-405). IEEE. https://doi.org/10.48550/arXiv.2412.14611

Hoover, D. L. (2004). Testing Burrows's delta. *Literary and Linguistic Computing, 19*(4), 453-475. https://doi.org/10.1093/llc/19.4.453

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng X., Qin, B., & Liu, T. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*. https://doi.org/10.1145/3703155

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). *Lora: Low-rank adaptation of large language models*. arXiv:2106.09685. https://doi.org/10.48550/arXiv.2106.09685

Hu, X., Ou, W., Acharya, S., Ding, S. H., D'Gama, R., & Yu, H. (2023). TDRLM: Stylometric learning for authorship verification by topic-debiasing. *Expert Systems with Applications, 233*, 120745. https://doi.org/10.1016/j.eswa.2023.1207

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. D. L., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux, M., Stock, P., Le Scao, T., Lavril, T., Wang, T., Lacroix, T., & Sayed, W. E. (2023). *Mistral 7B.* arXiv:2310.06825. https://doi.org/10.48550/arXiv.2310.06825

Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023). A watermark for large language models. In *International Conference on Machine Learning* (PMLR) (pp. 17061-17084). ML Research Press.

Knox, W. B., & Stone, P. (2011). Augmenting reinforcement learning with human feedback. In *ICML 2011 Workshop on New Developments in Imitation Learning* (vol. 855, No. 3). ACM. https://doi.org/10.1007/978-0-387-30164-8_714

Kumarage, T., Garland, J., Bhattacharjee, A., Trapeznikov, K., Ruston, S., & Liu, H. (2023). *Stylometric detection of AI-generated text in Twitter timelines*. arXiv:2303.03697. https://doi.org/10.48550/arXiv.2303.03697

Kumarage, T., & Liu, H. (2023). Neural authorship attribution: Stylometric analysis on large language models. *In 2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (pp. 51-54). IEEE. https://doi.org/10.1109/cyberc58899.2023.00019

Li, Y., Li, Q., Cui, L., Bi, W., Wang, L., Yang, L., Shi, S., & Zhang, Y. (2023). *Deepfake text detection in the wild*. arXiv:2305.13242. https://doi.org/10.48550/arXiv.2305.13242

Ma, Y., Liu, J., Yi, F., Cheng, Q., Huang, Y., Lu, W., & Liu, X. (2023). *Is this abstract generated by ai? A research for the gap between ai-generated scientific text and human-written scientific text*. arXiv:2301. 10416, https://doi.org/10.48550/arXiv.2301.10416

Maisto, A. (2025). Collaborative storytelling and LLM: A linguistic analysis of automatically-generated role-playing game sessions. arXiv:2503.20623. https://doi.org/10.48550/arXiv.2503.20623

Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023). Detectgpt: Zero-shot machine-generated text detection using probability curvature. In *International Conference on Machine Learning* (PMLR) (pp. 24950-24962). ML Research Press. https://doi.org/10.48550/arXiv.2301.11305

Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2024). Contrasting linguistic patterns in human and llm-generated text. *Artificial Intelligence Review, 57*, article number 265. https://doi.org/10.1007/s10462-024-10903-2

Opara, C. (2024). StyloAI: Distinguishing AI-generated content with stylometric analysis. In A.M. Olney, IA. Chounta, Z. Liu, O.C. Santos, & I.I. Bittencourt (Eds.), *Communications in Computer and Information Science, 2151*. Springer. https://doi.org/10.1007/978-3-031-64312-5_13

Przystalski, K., Argasiński, J. K., Grabska-Gradzińska, I., & Ochab, J. (2025). Stylometry recognizes human and LLM-generated texts in short samples. *Expert Systems with Applications, 129001*. https://doi.org/10.1016/j.eswa.2025.129001

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., & Finn, C. (2024). Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems, 36,* 53728-53741, https://doi.org/10.48550/arXiv.2305.18290.

Rebora, S. (2023). GPT-3 vs. Delta. Applying Stylometry to Large Language Models. In *La memoria digitale: Forme del testo e organizzazione della conoscenza*. Atti del XII Convegno Annuale AIUCD (pp. 292-297). AlmaDL. https://doi.org/10.6092/unibo/amsacta/7721

Roy, S. S., Thota, P., Naragam, K. V., & Nilizadeh, S. (2024, May). From chatbots to Phishbots?: Phishing scam generation in commercial Large Language Models. In *2024 IEEE Symposium on Security and Privacy* (pp. 221-221). IEEE Computer Society. https://doi.org/10.1109/SP54263.2024.00182

Rybicki, J., & Eder, M. (2011). Deeper Delta across genres and languages: Do we really need the most frequent words? *Literary and Linguistic Computing, 26*(3), 315-321, https://doi.org/10.1093/llc/fqr031

Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). *Can AI-generated text be reliably detected*? arXiv:2303.11156, https://doi.org/10.48550/arXiv.2303.11156

Sahoo, P., Meharia, P., Ghosh, A., Saha, S., Jain, V., & Chadha, A. (2024). A comprehensive survey of hallucination in large language, image, Video and audio foundation models. *Findings of the Association for Computational Linguistics* (EMNLP) (pp. 11709-11724). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.685

Salnikov, E., & Bonch-Osmolovskaya, A. (2023). Using stylometry to detect generated content. In *Proceedings of the DH Russia 2023 Conference* (pp. 176-182). Siberian Federal University.

Schuster, T., Schuster, R., Shah, D. J., & Barzilay, R. (2020). The limitations of stylometry for detecting machine-generated fake news. *Computational Linguistics, 46*(2), 499-510, https://doi.org/10.1162/coli_a_00380

Shamardina, T., Mikhailov, V., Chernianskii, D., Fenogenova, A., Saidov, M., Valeeva, A., Shavrina, T., Smurov, I., Tutubalina, E., & Artemova, E. (2022). *Findings of the the ruatd shared task 2022 on artificial text detection in Russian*. arXiv:2206.01583. https://doi.org/10.48550/arXiv.2206.01583

Smith, P. & Aldridge, W. (2011). Improving authorship attribution: Optimizing Burrows' delta method. *Journal of Quantitative Linguistics, 18*(1), 63–88, https://doi.org/10.1080/09296174.2011.533591

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, *60*(3), 538-556. https://doi.org/10.1002/asi.21001

Tang, R., Chuang, Y. N., & Hu, X. (2024). The science of detecting LLM-generated text. *Communications of the ACM, 67*(4), 50-59. https://doi.org/10.1145/3624725

Wang, S., Cristianini, N., & Hood, B. (2024). Stylometric comparison between ChatGPT and Human Essays. In *Workshop Proceedings of the 18th International AAAI Conference on Web and Social Media*, *Workshop: REAL-Info 2024: First Workshop on Reliable Evaluation of LLMs for Factual Information*. AAAI. https://doi.org/10.36190/2024.24

Wang, Y., Mansurov, J., Ivanov, P., Su, J., Shelmanov, A., Tsvigun, A., Whitehouse, C., Afzal, O. M., Mahmoud, T., Sasaki, T., Arnold, T., Aji, A. F., Habash, N., Gurevych, I., & Nakov, P. (2023). M4: Multi-generator, multi-domain, and multi-lingual black-box machine-generated text detection. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Vol. 1: Long Papers*, pp. 1369–1407). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.eacl-long.83

Wang, Y., Pan, Y., Yan, M., Su, Z., & Luan, T. H. (2023). A survey on ChatGPT: AI-generated contents, challenges, and solutions. *IEEE Open Journal of the Computer Society, 4*, 1-20. IEEE Computer Society. https://doi.org/10.1109/OJCS.2023.3300321

Weerasinghe, J., Seepersaud, O., Smothers, G., Jose, J., & Greenstadt, R. (2025). Be sure to use the same writing style: Applying authorship verification on Large-Language-Model-generated texts. *Applied Sciences, 15*(5), 2467. https://doi.org/10.3390/app15052467

Wu, J., Yang, S., Zhan, R., Yuan, Y., Chao, L. S., & Wong, D. F. (2025). A survey on LLM-generated text detection: Necessity, methods, and future directions. *Computational Linguistics*, *51*(1), 275-338. MIT Press. https://doi.org/10.1162/coli_a_00549

Zaitsu, W., & Jin, M. (2023). Distinguishing ChatGPT (-3.5,-4)-generated and human-written papers through Japanese stylometric analysis. *PLoS One, 18*(8), e0288453. https://doi.org/10.1371/journal.pone.0288453

Zhang, B., Liu, Z., Cherry, C., & Firat, O. (2024). *When scaling meets LLM finetuning: The effect of data, model and finetuning method.* arXiv:2402.17193. https://doi.org/10.48550/arXiv.2402.17193

Zhao, X., Li, L., & Wang, Y. X. (2022). *Distillation-resistant watermarking for model protection in NLP.* arXiv:2210.03312. https://doi.org/10.48550/arXiv.2210.03312

https://doi.org/10.17323/jle.2025.24343

Teacher-Student Dynamics in AI-Driven Language Education in the Post-Truth Era

Abdu Al-Kadi

University of Bergen, Bergen, Norway

ABSTRACT

Background: Artificial intelligence (AI) inaugurates a new educational era that compels a rethinking of established pedagogical paradigms. Within the epistemological uncertainty of the post-truth era, it has reconfigured teacher-student dynamics (TSD) in ways that challenge traditional assumptions about agency and authority in the classroom.

Purpose: The article addresses the disruption of TSD under the evolving force of AI, with particular attention to tensions in learner–teacher agency asymmetry in the context of AI-mediated language education.

Conceptual Contribution: Positioned within the genre of conceptual scholarship, the article introduces a model that delineates six interrelated dimensions of TSD disruption in the age of AI. The framework does not merely describe emerging shifts but systematizes them into an interpretive structure that traces the trajectory of TSD evolution. In doing so, it foregrounds the broader implications of these transformations for educational policy, pedagogical design, and research agenda in language education.

Implications: The analysis contends that AI realities not only govern but also reshape the human texture of pedagogical interaction. Preserving the integrity of the language classroom requires learning designs that foreground dialogic engagement and epistemic trust, while constructively integrating AI innovations.

KEYWORDS

Artificial intelligence (AI); ethics, integrity; teacher-student dynamics (TSD)

INTRODUCTION

In an era where machines can simulate instruction, the essence of education lies not in the delivery of information, but in the effective human dynamism between teacher and student. Machine learning and Artificial Intelligence (AI), which have been around for a while, have affected normal relations between teachers and the digital generation – or as some call it, a generation of automated education (Kooli, 2023; Sparrow & Flenady, 2025). This generation is shaped by AI tools, perhaps even more than human dialogue. While generative AI (GenAI) has been praised for enhancing efficiency in language learning (Moorhouse & Wong, 2025), it subtly alters pedagogical relationships in ways that are less visible but impactful. It weakens human-to-human connections and widens the gap in rapport between teachers and students (Ji et al., 2024; Montanucci & Peconi, 2024; Viberg et al., 2024). AI systems reshape the teacher-student dynamics (TSD) through automated systems. Such changes challenge the foundational principles of education, forcing a new way of thinking about what it means to teach and learn in an AI-driven era.

This article uses Paulo Freire's *dialogical pedagogy* and Gert Biesta's *subjectification* as touchstones to contemplate the new logic that AI has introduced into TSD. Biesta (2017) critiques the dominant view of teaching as knowledge transmission. To Biesta, teaching is a transformative and relational act that fosters student autonomy and forms responsible individuals who can maintain their agency. Similarly, Freire's idea of dialogical pedagogy challenges the 'banking model' of

Citation: Al-kadi, A. (2025). Teacher-student dynamics in ai-driven language education in the post-truth era. *Journal of Language and Education, 11*(3), 152-159. https://doi.org/10.17323/jle.2025.24343

Correspondence:
Abdu Al-kadi,
abdu.al-kadi@uib.no

Received: December 31, 2024
Accepted: September 18, 2025
Published: September 30, 2025

education in which the teacher 'deposits' information into students' minds, and students, in return, are expected to memorize the deposited information and retrieve it when they are asked to do so in exams. Freire's pedagogical legacy lays emphasis on mutual trust and critical thinking as a foundation for the co-construction of knowledge. This contrasts with AI-driven automation. The main argument this article advocates is that AI is not merely a pedagogical tool that reduces teaching to content delivery but a disruptive epistemic force that reconfigures TSD as it shifts authority, erodes dialogic trust, and redefines learner and teacher agency, and these shifts challenge Freire's vision of dialogical pedagogy and Biesta's call for subjectification in world-centered education.

Conceptual Tensions

One of the key challenges in AI-mediated education is navigating the post-truth era, where misinformation complicates knowledge validation. The term post-Truth Era denotes an age of doubts and distrust. It was selected as the Oxford Dictionaries Word of the Year 2016, and it has been used in published work to describe a period of uncertainty (Li & Chiu, 2024; Pratschke, 2024). It reflects a difficulty in discerning genuine information from fabricated content (Malcolm, 2021). In such an era, AI applications advance drastically, and so does the prevalence of deep-fake videos, images, and other digital content (Perkins & Roe, 2025). GenAI can be used nowadays to create fake images, videos, and audio recordings, which distort societal awareness and the teaching community. It impacts society and public opinion. Even if some of what we see is realistic, the growing awareness under the influence of AI makes us more skeptical and in want to distinguish truth from fake content (Pratschke, 2024). In this vein, Li and Chiu (2024) argued that 'truth' has evolved from being a static, absolute entity to a more dynamic and context-dependent construct. This shift, driven by AI, can generate multiple competing truths based on different perspectives and contexts.

In the post-truth era, AI has affected the trust that learners and educators place in knowledge sources. Trust operates on two levels: interpersonal trust, built through human relationships between teachers and students, and technological trust, which concerns confidence in AI systems, their reliability, and transparency. While the former fosters emotional connection and ethical guidance, the latter depends on how AI tools communicate their processes and limitations. Both are essential and intertwined in shaping effective TSD. Pratschke (2024) and Malcolm (2021) described the current era as both constructive and destructive. It has increased concerns about the originality of the content we have today because the latest AI-based technologies have largely dissolved and narrowed the boundary between the human species and machines. However, critics tend to question the trustworthiness of AI outcomes (Viberg et al., 2024), arguing that GenAI does not generate knowledge from scratch and AI's outcome is generally based on humans' knowledge, big data, and massive information fed into such language chatbots, which are trained on such inputs (Montanucci & Peconi, 2024).

TSD Reconfiguration Model

In language education, as in other fields, AI's transformative force has reshaped the nature of TSD, giving way to automated learning (Sparrow & Flenady, 2025) and epistemological uncertainty of the post-truth era (Li & Chiu, 2024; Malcolm, 2021). The new move can be represented as a conceptual model of six dimensions of TSD disruption. These dimensions, if I see it correctly, shape the transformation that this article brings to the fore. The model conceptualizes how TSD evolves under the influence of AI.

Criteria and Analytic Logic

The model theorizes TSD reconfiguration in the context of higher education. It was developed through a focused synthesis of literature, recent empirical studies, and analytical deduction. Systematic search was conducted across Scopus, Web of Science, ERIC, and major publisher platforms, covering the period from 2015 to 2025– a decade during which the post-truth era gained momentum, and the impact of AI tools are evident. The keywords that guided the search are *AI, education, trust, agency, assessment, curriculum, datafication, post-truth, Freire, and Biesta.* Only English-language sources relevant to language education were included. Of the initial records identified, only 42 studies were included after screening. Data extraction focuses on constructs (*trust, agency, assessment, and curriculum*), findings, and theoretical anchors. Evidence from the selected studies was analyzed using a thematic synthesis, combining deductive coding based on Freire's and Biesta's theoretical frameworks with inductive coding drawn from recent empirical findings. A constant-comparative method was used to refine categories and identify tensions across studies. Themes were elevated to dimensions when they appeared consistently across multiple sources as key factors of disruption, demonstrated conceptual clarity, and were supported by illustrative examples. Other dimensions, such as infrastructure or policy, were excluded to maintain focus on relational and epistemic shifts within classroom practice. Each of the six dimensions was defined with clear boundaries. For instance, *Dialogic Trust* was anchored in Freirean pedagogy. The resulting six dimensions, besides revising and extending existing conceptualizations of TSD, offer a critical lens that bridges educational theory and emerging AI realities. For instance, while Viberg et al. (2024) explored trust and demographic factors, and Zhai (2024) emphasized role refashioning, the model synthesizes these strands into a typology that foregrounds dialogic erosion, agency asymmetry, and curriculum flattening. These elements reveal imbalances in digital literacy and agency between teachers and students, both complementing and challenging the optimism expressed in Montanucci

and Peconi (2024). Quality screening and reflexivity were applied throughout, with noted limitations including language scope and variability in grey literature rigor.

Levels of TSD Reconfiguration

Traditionally, a classroom was viewed as a micro-learning community wherein social interaction and tasks with a clear purpose take place (Ji et al., 2024; Kramm & McKenna, 2023; Modhish & Al-Kadi, 2016; Montanucci & Peconi, 2024). In comparison, today's classrooms represent a space for wired and wireless gadgets (Ilic & Sato-Ilic, 2024) and humans (learners and teachers) using these high-tech gizmos as learning companions (Tolstykh & Oshchepkova, 2024). This has transformed the dynamics of traditional classroom settings. Educators are "no longer the sole authority or holder of knowledge, the written assignment is no longer viable as proof of learning, and the classroom is no longer the center of activity" (Pratschke, 2024, p. 2). Teachers, after a long history of 'sage on the stage,' no longer serve as a central source of knowledge and authority in the classroom, and knowledge is no longer defined as a transmission from teachers to learners.

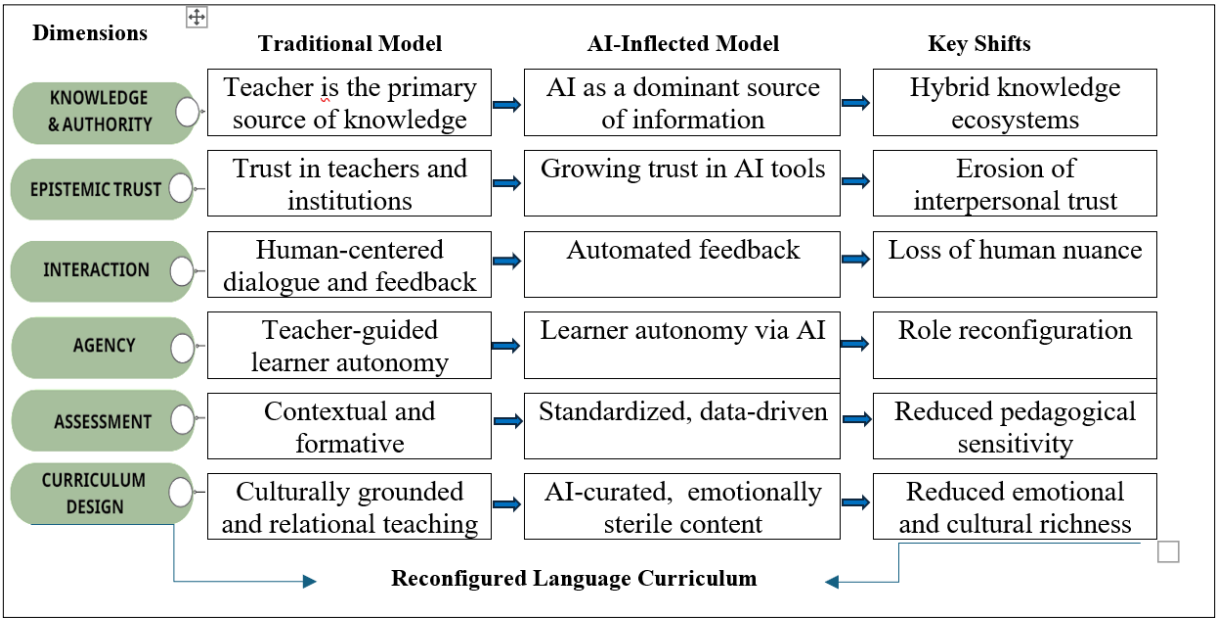
Figure 1 illustrates how AI, as a transformative force, is reshaping fundamental aspects of language education. In the first dimension, AI redefines traditional notions of knowledge creation and authority in education. The rapid advances of AI have blurred the lines between human and machine roles, raising concerns about authenticity and knowledge authority. In the new learning landscape, teachers' roles shift from being the main source of knowledge to minimal or even 'null' roles. Teachers, before GenAI tools and machine learning, used to imbibe knowledge from books and libraries, or they might have inherited it verbally from their ancestors, and students highly valued them as knowledge holders. In this light, teachers were touted as sources of knowledge that they then passed on to their learners. This TSD has been disrupted lately by the "development of machines that are intelligent [and] learn at a much faster rate than we could ever hope to" (Pratschke, 2024, p. 6). This shift decentralizes the teacher's traditional role as the sole authority and introduces AI as a co-instructor. It redistributes agency within the learning process.

AI-generated content sharply challenges the long-held ideas about who creates, owns, and shares knowledge. Recalling Pratschke's (2024) argument about who constructs and owns knowledge in this age of digitally generated content, it is fair to argue that teachers, learners, and smart machines are all agents of knowledge production and dissemination. At present, the new concept, encompassing both tacit and explicit knowledge, involves mutual engagement, negotiation of meaning, and interactions (Freire, 2017). This promotes a more holistic and interactive approach to learning wherein learners and teachers use GenAI tools to construct knowledge collaboratively. GenAI provides an array of tempting tools used for generating content and answering questions (Montanucci & Peconi, 2024) – tasks that were exclusively reserved for teachers. Most students, if not all, have AI systems now as their primary knowledge sources. Such tools shift the locus of authority from teachers to machines.

In the second dimension, the notion of 'trust,' traditionally grounded in human relationships, institutional credibility, and pedagogical transparency, leans toward AI more than humans in AI-empowered education. Students tend to rely on GenAI tools like ChatGPT for generating content or feedback, and teachers, likewise, find such tools useful and

Figure 1

A Conceptual Typology of AI-Driven Shifts in TSD



time-saving for scoring students' assignments and providing automatic feedback. When both learners and teachers use AI without acknowledgment, it casts uncertainty over how it is being used. So far, the epistemic foundations of GenAI tools (e.g., training data) are opaque, and this opacity demands a new form of *critical epistemic trust*– one that balances openness to machine-generated insights with skepticism about their provenance, bias, and authority. AI gives ground to cheating (Pikhart & Al-Obaydi, 2025), which contributes to a breakdown in trust between teachers and students. Students may use AI tools such as ChatGPT and its siblings to generate content in the form of essays or other assignments, passing them off as if they were their original work, and this, when it happens, undermines the originality of learning and brings about loss of higher-order thinking skills (Ogurlu & Mossholder, 2023). After the influx of such AI tools, a change in the quality or tone of a student's writing is noticed (Adams et al., 2022). When students use AI to cheat or produce fabricated content, teachers may become skeptical of all student work (Al-Kadi, 2025b), and this skepticism drives TSD transformations (Ogurlu & Mossholder, 2023; Tolstykh & Oshchepkova, 2024; Zhai, 2024). This status quo has led to a growing demand for originality detection tools (Giray et al., 2025), raising questions not only about trust between teachers and their learners (Al-Kadi, 2025b) but also the extent to which AI users trust knowledge generated by AI tools.

In the third dimension, the dialogic process that Freire (2017) identifies as essential for developing critical consciousness is being undermined. Interactions that used to be centered around human connection are now managed by machines. This shift degrades the co-construction model of education that Freire champions. That is, the reconfiguration of TSD through AI is not just a technological shift but a pedagogical rupture. It challenges the conditions under which emancipatory learning can occur. According to Freire (2017), true dialogue cannot exist unless the dialoguers (in this case, *teacher and students*) engage mutually in critical thinking. It is an act of humility and trust. This dialogical approach is an ethical stance toward education and human relationships. A shift of learning from traditional dialogic interaction to the AI-inflected model lacks emotional nuance and pedagogical sensitivity. It risks dehumanizing education (Al-Kadi, 2018) and erodes the relational foundation of education. Freire's pedagogical legacy emphasizes mutual trust and co-construction of meaning, and these two are diminished when AI replaces human interaction.

In the fourth dimension, the idea of learner agency, originated in learner-centered approaches and communicative teaching (Cong-Lem & Daneshfar, 2024), is amplified in AI-inflected education, wherein teachers represent one of the many learning resources that AI has made available at learners' fingertips. The term *agency* builds on, but also complicates, earlier learner-centered models. In AI-based language education, it stands for the ability of teachers and students to make meaningful, autonomous decisions, and this concept is an extension of the premise of learner-centeredness and CALL, wherein learners learn independently (Cong-Lem & Daneshfar, 2024; Pratschke, 2024) with or without formal support from their teachers or institutions (Modhish & Al-Kadi, 2016). In the context of Biesta's *subjectification*, agency involves how learners and educators position themselves in relation to knowledge, authority, and ethical responsibility. It is not just about acquiring skills autonomously but becoming subjects who can act with integrity and responsibility. To Freire, agency is cultivated through dialogue, where both teacher and student are transformed. Both views give ground for a student to be in the position of 'subject' rather than 'object'. Teaching, which has been dominated by the teacher and perceived as an act of control (Biesta, 2017), is shifting, thanks to AI, toward more learner-centered dynamics. Zhai (2024) observed that shifting teacher and learner roles foster greater autonomy, and Ilic and Sato-Ilic (2024) noted that AI enables learners to co-design and create content, often without direct teacher input (Montanucci & Peconi, 2024). The availability of digital tutors and language chatbots around the clock facilitates ongoing language learning beyond the confines of traditional settings (Montanucci & Peconi, 2024).

In the fifth dimension, assessment that Perkins and Roe (2025) viewed as a mechanism of power that shapes the student-teacher relationship, determining who gets to know, how, and under what conditions, is shifted now toward automated assessment. The new direction prioritizes scalability over nuance. It reduces opportunities for dialogic human assessment, which is essential for critical thinking and engagement. Freire's critique of the 'banking model' in which students passively receive knowledge resonates with concerns about AI-driven feedback systems, which deliver pre-structured responses without fostering much dialogue or critical reflection. As Moorhouse and Wong (2025) argue, AI-generated feedback lacks emotional nuance and pedagogical sensitivity. That is, GenAI has brought about "potential collapse of traditional assessments" (Perkins & Roe, 2025, p. 88), and in the post-collapse scenario, skills of analysis, recall, and, above all, writing are delegated to machines (Perkins & Roe, 2025). While the AI-inflected model is tied with consistency and bias reduction, it seems to narrow the scope of what counts as language proficiency, pragmatic competence, intercultural expertise, and critical language awareness.

The sixth dimension is about curriculum design. It has been traditionally a human endeavor grounded in cultural context, pedagogical values, and local relevance. In AI-curated content, learning materials are generated or selected with limited consideration for cultural, ethical, or educational aspects emphasized in traditional formal curriculum (Pikhart & Al-Obaydi, 2025). The force of AI undermines the relational core of education. As AI tools curate content, automate feedback, or suggest pedagogical strategies, they affect the

texture of human communication in the classroom. Such tools limit the space for intentional reflective actions. This shift flattens the multidimensional nature of language learning and reduces students' learning to measurable content. That is, it narrows the scope of formal curriculum, which is a broader pedagogical framework that goes beyond bite-sized content. It results in educational systems that lack pedagogical nuances, emotional and cultural depth. Raitskaya and Tikhonova (2025) contend that GenAI tools influence how students engage in higher-order thinking tasks, which is an overall goal of the curriculum. Apart from curriculum narrowing, automated input, which is emotionally sterile, lacks the nuance of human critique (Kooli, 2023; Luo, 2024). When education is reduced to a process of optimization, it tends to sideline the ethical and relational dimensions that are central to forming individuals capable of engaging with the world (Biesta, 2017) rather than merely adapting to existing systems.

Contextual Variability in TSD

The model contributes a context-sensitive, ethically grounded framework for understanding TSD in the post-truth era. It aligns with Montanucci and Peconi (2024) in recognizing that AI tools reshape TSD. However, the impact of these disruptions varies across educational contexts due to cultural and institutional factors. Each of the six dimensions interacts differently with local pedagogical norms. For instance, in Scandinavian contexts, where transparency and learner independence are embedded in the curriculum, AI tends to reinforce student agency and dialogic trust but may decentralize teacher authority. In East Asian systems, which emphasize hierarchy and standardized testing, AI may intensify automated assessment and surveillance, potentially undermining dialogical pedagogy and epistemic trust. In Arab educational settings, where teacher authority and rote learning are prevalent, AI's shift toward learner-centeredness and automated feedback may disrupt established norms. It requires careful integration to preserve relational trust and ethical guidance. In low-resource environments, limited infrastructure can hinder the adoption of AI technologies. However, the potential benefits, such as scalable assessment and efficient content delivery, may still hold strong appeal for educators and policymakers seeking cost-effective solutions.

Digital Literacy and Critical Thinking

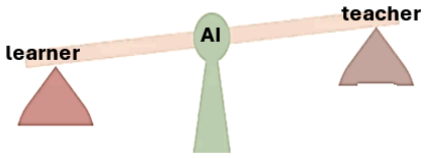
The model proposed in this article is context-sensitive yet implicative for change in the unstable context of the post-truth era. Two major factors shape how each dimension manifests: digital literacy and critical thinking. They should be considered when applying the model across diverse educational landscapes. In teaching contexts where teachers struggle with digital literacy, a disconnect arises between them and their students, affecting the dynamics of interaction between them. Adams et al. (2022) and Lucas et al. (2024) maintain that a lack of digital literacy contributes to teachers' misunderstanding of how learners do things with AI and, hence, assume plagiarism for everything their students do. Hence, teachers' digital literacy plays a key role in shaping their trust in AI tools and in their students' capabilities. Viberg et al. (2024) explored factors that affect teachers' trust in AI, including age, gender, level of education, cultural dimensions, and self-efficacy across six countries. In the study, teachers with higher self-efficacy and a better understanding of AI had more trust in its benefits and fewer concerns. According to the study, trust levels vary by country and culture, but demographic factors like age and gender do not significantly impact trust. These results are echoed in Lucas et al.'s (2024) findings in that teachers' familiarity with GenAI enables them to not only handle AI-related trust issues more capably but also view students' use of AI positively.

AI disproportionately empowers learners more than teachers with limited AI literacy, and this leaves teachers with a feeling of disempowerment. Learners are positioned in AI-based pathways to shape their learning experiences autonomously, construct and share knowledge independently of traditional teacher-led instruction (Lan & Chen, 2024; Yung, 2023). They gain more control over their learning through engagement with AI-powered tools and adaptive learning platforms. For example, ChatGPT and DeepSeek support language learning inside and outside the classroom. The imbalance in AI literacy between teachers and students can be described as a dynamic shift in pedagogical agency. Teachers need to reposition themselves closer to AI to restore balance within the AI-enabled learning spectrum illustrated in Figure 2. The closer teachers align themselves with AI (the key driver of change), the more they can maintain balance with learners in the AI-driven learning environment. This requires effective dialogue in which both sides change their positions on the swing to the extent of maintaining their balance. It also requires school-wide reforms: teaching methods, syllabi, content, assessment, and, above all, the mindsets.

Besides digital literacy, critical thinking is essential in an era rife with AI-based tools that facilitate the spread of junk digital content (Li & Chiu, 2024; Pratschke, 2024) and 'AI hallucination' (Crompton & Burke, 2024). Learning cannot simply be a process in which students receive course content as knowledge. It is about students attempting to become conversant with that body of content. Biesta's idea of subjectification helps us to resist AI control. Paulo Freire's critical pedagogy (2017) is inspirational to get into dialogue with learners to create a healthy learning atmosphere for discussing AI-generated content. Likewise, Raitskaya and Tikhonova's (2025) scoping review is insightful for rethinking critical thinking skills in AI-human interaction. It reinforces Sparrow and Flenady's (2025) argument for prioritizing hu-

Figure 2

Learners and Teachers on the AI Swing



man interaction and critical thinking within a more cautious and thoughtful approach to integrating AI into language learning, teaching, and assessment.

CONCLUSION

While AI offers opportunities for more efficient language learning, it challenges the theoretical foundations that underpin human-oriented education. The article brings to the foreground of AI-in-education research a speculation that unchecked influx of GenAI into language education not only reconfigures teacher-student roles but also destabilizes foundational principles of knowledge creation, authority, and trust. It indicates implications for policy, professional practice, and future research.

Policy Implications

The fact that students use AI in their assignments, and teachers, likewise, use it for assessing and evaluating these assignments, implies a necessity for institutional policies for transparent AI disclosure. Mandatory AI disclosure is a must for more transparency and cultivating epistemic trust. Instead of punitive or control-based approaches to managing AI use in education, we should be more open to dialogic use of AI innovations. Students and teachers should disclose how they employ AI tools in their work. International bodies should facilitate cross-cultural dialogue on AI's role in education, recognizing that its impact varies by context. Frameworks, such as the Ethical principles in Kooli (2023), along with the code of practice from Edinburgh University (2023) and the *European Ethical Guidelines for Educators on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning*, can be a reference point for enhancing benefits and addressing the challenges of AI in language education. These resources can be used for lawful mechanisms and strategies. They can readily inform policymakers about upholding ethical standards of AI in today's education. For example, the University of Edinburgh's AI Code provides a model for responsible AI use in higher education, emphasizing transparency, accountability, and human oversight. Such initiatives can serve as springboards for broader adoption and adaptation of AI in the TSD that this paper takes as its primary focus.

Another implication relates to digital literacy training for both teachers and students to foster agency, informed engagement with AI tools, and ethical and philosophical reflection on what it means to 'know' in an AI-mediated interaction. In the post-truth era now, AI drives us to behave like machines if we take its outcomes for granted. This is alarming for learners and teachers using AI. They should imbibe skills for filtering and questioning AI-generated content that they consume and share. Digital literacy should extend beyond technical proficiency to include pedagogical awareness of how AI reshapes classroom dynamics and feedback (Moorhouse & Wong, 2025). GenAI tools constantly improve to perform more complex tasks. This progression implies more teacher training and institutional support (Kamali et al., 2024). Training programs may adopt Li's (2024) strategies for ethical governance in education, for instance, to integrate such strategies for an ethically-aware academia.

When it comes to curriculum, classrooms should evolve into spaces of trustful human dialogue rather than mechanized control. For education to remain human-aligned, educators and policymakers should ensure that AI-driven learning aligns with ethical principles and genuine dialogue. Relational and negotiated agencies require deliberate curriculum design to preserve human discretion and dialogic engagement. Human oversight in AI-assisted assessment should be maintained to preserve dialogical engagement and critical thinking. Curriculum standards should ensure that AI-curated content reflects cultural relevance and pedagogical depth. In low-resource settings, policies should prioritize equitable access to AI tools, infrastructure support, and teacher training to prevent agency asymmetry and curriculum narrowing.

Implications for Teachers

There are implications for teachers to re-think their roles within the evolving educational systems. Teachers are still important for learning design (Lan & Chen, 2024), fostering learners' critical thinking, and the ethical use of AI (Kamali et al., 2024). They need to remain integral within AI-mediated learning. AI has repositioned teachers and learners along the route of learning, so teachers should play transformative roles in alignment with the mindset of today's generation. This study implies that teachers are repositioned as

co-learners in the AI-inflected context in which AI tools (e.g., ChatGPT) and adaptive learning platforms (e.g., Duolingo) offer alternative or additional sources of information. There should be a move beyond surveillance-driven approaches to partnering with AI in education. Students in a low-trust environment feel surveilled rather than supported, and this echoes the voice of Kramm and McKenna (2023) in that the prevalent focus on detecting AI in students' work overlooks the broader purposes of education.

Besides fostering a supportive learning environment in lieu of dictating outcomes, teachers should cultivate a dialogical space where students can think critically, pose questions, and participate in knowledge production. They may use AI to initiate human-led dialogue. They may also use structured protocols to help students engage with AI responses critically. While AI can assist with formative feedback, teacher-led assessments remain essential for evaluating creativity, critical thinking, and interpersonal skills. Reflective activities should also invite students to explain how they used AI and what they learned from it.

Implications for Research

The article gives way to empirical research into TSD under the influence of GenAI to illustrate, through case studies, how AI-mediated TSD unfolds in real-world educational settings. Also, inconsistent application of AI is particularly exacerbated by a lack of (a) a dichotomy of what is ethical and unethical, and (b) policies that govern these issues, which stimulates more critical reflection and suggests uncharted areas for further research. To test and refine the conceptual TSD model, empirical studies are needed to validate its six dimensions and explore their contextual variability. For example, cross-cultural case studies could examine how these dimensions manifest in different educational settings. Also, survey research could assess trust in AI-generated feedback, while experimental studies could compare student agency and learning outcomes in AI-mediated versus teacher-led environments.

In closing, this paper problematizes the shifting educational goals and roles resulting from AI's reconfiguration of TSD. AI shifts epistemic authority alters dialogic human-human pedagogy, and challenges traditional roles, and these shifts require a rethinking of educational frameworks. It penetrates the fabric of education, and the question is no longer about whether TSD will change, but how far the change will take education. In an era of uncertainty, the future of education is uncertain. We should dare to ask about the real purpose of today's education. When AI becomes a learning companion that outperforms students on exams, and teachers attend to AI for feedback and instructional support, what kinds of learners, citizens, and human beings do educational systems aim to cultivate? Freire's dialogical model and Biesta's concept of subjectification, which function as conceptual signposts, challenge us to ask whether we design AI tools to support genuine educational encounters or merely efficient delivery systems. How can we ensure that learners remain active participants in shaping their learning, rather than being shaped by invisible algorithms?

DECLARATION OF COMPETING INTEREST

None declared.

REFERENCES

Adams, C., Pente, P., Lemermeyer, G., Turville, J., & Rockwell, G. (2022). Artificial intelligence and teachers' new ethical obligations. *The International Review of Information Ethics, 31*(1). https://doi.org/10.29173/irie483

Al-Kadi, A. (2025b). Probing language teachers' complaints about ChatGPT and GenAI tools: A social media dataset study. *Issues in Educational Research, 35*(2).

Al-Kadi, A. (2018). Towards humanizing ELT: Revisiting the need for English in the medical context in Yemen. *Language Teaching and Educational Research, 1*(2), 121-138. https://dergipark.org.tr/en/download/article-file/612983

Biesta, G. J. J. (2017). *The rediscovery of teaching (1st ed)*. Routledge. https://doi.org/10.4324/9781315617497

Cong-Lem, N., & Daneshfar, S. (2024). Generative AI and second/foreign language education from Vygotsky's cultural-historical perspective. In H. P. Bui & E. Namaziandost (Eds.), *Innovations in technologies for language teaching and learning* (pp. 175-188). Springer. https://doi.org/10.1007/978-3-031-63447-5_10

Crompton, H. & Burke, D. (2024). The educational affordances and challenges of ChatGPT: State of the field. *TechTrends*, 68, 380–392. https://doi.org/10.1007/s11528-024-00939-0

Freire, P. (2017). *Pedagogy of the oppressed*. Penguin Classics.

Giray, L., Sevnarayan, K., & Ranjbaran Madiseh, F. (2025). Beyond policing: AI writing detection tools, trust, academic integrity, and their implications for college writing. *Internet Reference Services Quarterly, 1*(1), 1-34. https://doi.org/10.1080/10875301.2024.2437174

Ji, H., Han, I., & Park, S. (2024). Teaching foreign language with conversational AI: Teacher-student-AI interaction. *Language Learning & Technology, 28*(2), 91-108. https://hdl.handle.net/10125/73573

Kamali, J., Alpat, M.F. & Bozkurt, A. (2024). AI ethics as a complex and multifaceted challenge: decoding educators' AI ethics alignment through the lens of activity theory. *International Journal of Educational Technology in Higher Education, 21*(62). https://doi.org/10.1186/s41239-024-00496-9

Kramm, N., & McKenna, S. (2023). AI amplifies the tough question: What is higher education *really* for? *Teaching in Higher Education*, *28*(8), 2173-2178. https://doi.org/10.1080/13562517.2023.2263839

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability, 15*(7), 5614-5614. https://doi.org/10.3390/su15075614

Lan, Y.-J., & Chen, N.-S. (2024). Teachers' agency in the era of LLM and generative AI: Designing pedagogical AI agents. *Educational Technology & Society*, *27*(1), 1-18. https://www.jstor.org/stable/48754837

Li, Y., & Chiu, C. Y. (2024). Redefining truth in the context of the AI-truth era: A practice-led research of "From post-truth to AI-truth". In P. L. P. Rau (Ed.), *Cross-cultural design. HCII 2024. Lecture Notes in Computer Science* (pp. 245-257). Springer. https://doi.org/10.1007/978-3-031-60913-8_17

Ilic, P., & Sato-Ilic, M. (2024). AI-enhanced ecological learning spaces. In P. Ilic., I. Casebourne, & R. Wegerif (Eds.), *Artificial intelligence in education: The intersection of technology and pedagogy* (pp.17-37). Springer. https://doi.org/10.1007/978-3-031-71232-6_2

Lucas, M., Zhang, Y., Bem-haja, P. *et al.* (2024). The interplay between teachers' trust in artificial intelligence and digital competence. *Education and Information Technologies,* 29, 22991-23010. https://doi.org/10.1007/s10639-024-12772-2

Luo (Jess), J. (2024). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education,* 1-16. https://doi.org/10.1080/13562517.2024.2341005

Malcolm, D. (2021). Post-truth society? An Eliasian sociological analysis of knowledge in the 21st century. *Sociology*, *55*(6), 1063-1079. https://doi.org/10.1177/0038038521994039

Modhish, A. & Al-Kadi, A. (2016). Internet integration in EFL college instruction: Attitudes and perspectives. *International Journal on Studies in English Language and Literature, 4*(6), 52-62. https://doi.org/10.20431/2347-3134.0406008

Montanucci, G., & Peconi, A. (2024). Unlocking the potential of chatbots and intelligent tutors for language learning in online courses. In G. De Cristofaro, F. Silva, & B. Samu (Eds.), *Language MOOCs and OERs: New trends and challenges* (pp. 476-497). Perugia Stranieri University Press.

Moorhouse, B. L., & Wong, K. M. (2025). *Generative artificial intelligence and language teaching*. Cambridge University Press. https://doi.org/10.1017/9781009618823

Ogurlu, U., & Mossholder, J. (2023). The perception of ChatGPT among educators: Preliminary findings. *Research in Social Sciences and Technology, 8*(4), 196-215. https://doi.org/10.46303/ressat.2023.39

Perkins, M., & Roe, J. (2025). The end of assessment as we know it: GenAI, inequality and the future of knowing. In *AI and the future of education: Disruptions, dilemmas and directions: AI pedagogies, assessment and emerging educational futures*. Durham University Repository. https://durham-repository.worktribe.com/OutputFile/4512330

Pikhart, M., & Al-Obaydi, L. H. (2025). Reporting the potential risk of using AI in higher education: Subjective perspectives of educators. *Computers in Human Behavior Reports, 18*, 100693. https://doi.org/10.1016/j.chbr.2025.100693

Pratschke, B. M. (2024). *Generative AI and education digital pedagogies, teaching innovation, and learning design*. Springer

Raitskaya, L. & Tikhonova, E. (2025). Enhancing critical thinking skills in ChatGPT-human interaction: A scoping review. *Journal of Language and Education*, *11*(2), 5-19. https://doi.org/10.17323/jle.2025.27387

Sparrow, R., & Flenady, G. (2025). Bullshit universities: The future of automated education. *AI & Society*. https://doi.org/10.1007/s00146-025-02340-8

Tolstykh, O., & Oshchepkova, T. (2024). Beyond ChatGPT: Roles that artificial intelligence tools can play in an English language classroom. *Discover Artificial Intelligence, 4*(1)60. https://doi.org/10.1007/s44163-024-00158-9

Viberg, O., Cukurova, M., Feldman-Maggor, Y. et al. (2024). What explains teachers' trust in AI in education across six countries? *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-024-00433-x

Yung, K. W. H. (2023). Theorizing agency from a complex dynamic systems theory perspective: Evidence from language learner narratives in online shadow education. *The Modern Language Journal*. https://doi.org/10.1111/modl.12988

Zhai, X. (2024). Transforming teachers' roles and agencies in the era of generative AI: Perceptions, acceptance, knowledge, and practices. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-024-10174-0

https://doi.org/10.17323/jle.2025.27419

Applying the SAMR Model to AI-Enhanced Business Language Instruction: Comparative Insights from German and Spanish Contexts

Katrin Herget , Katty da Silva Ferreira

University of Aveiro/CLLC, Aveiro, Portugal

ABSTRACT

Background: The SAMR model offers a structured way to analyze how technology reshapes teaching and learning. In the context of Generative Artificial Intelligence (GAI), it provides a productive lens for reconsidering how language instruction - especially within Languages for Specific Purposes (LSP) - evolves in response to AI integration.

Purpose: This article examines the use of AI-driven tools in Business German and Business Spanish across two contrasting settings: higher education and corporate training. It aims to build a comparative perspective that highlights how pedagogical intentions and constraints differ across these environments and what that means for responsible and effective AI integration.

Conceptual Contribution: We propose that the SAMR model can be reframed specifically for Generative AI in Business LSP to clarify how instructional practices move between enhancement and transformation. Rather than reporting empirical findings, the article argues that AI-enabled personalization, collaboration, and authentic task design can be interpreted as occupying distinct SAMR levels. This conceptual reframing distinguishes between uses of AI that streamline existing practices and those that fundamentally rethink language learning experiences.

Implications: By repositioning SAMR as a flexible analytic tool rather than a descriptive taxonomy, this article contends that educators can better anticipate both opportunities and risks in professional language instruction. The argument advances a foundation for integrating GAI in ways that support pedagogical innovation while safeguarding equity, integrity, and quality in Business German and Business Spanish education.

KEYWORDS

Generative AI; SAMR model; Languages for Specific Purposes; Business German; Business Spanish

Citation: Herget, K., & da Silva Ferreira, K. (2025). Applying the SAMR model to AI-enhanced business language instruction: Comparative insights from German and Spanish contexts. *Journal of Language and Education, 11*(3), 160-167. https://doi.org/10.17323/jle.2025.27419

Correspondence:
Katrin Herget,
katrin.herget@gmail.com

Received: June 26, 2025
Accepted: September 18, 2025
Published: September 30, 2025

INTRODUCTION

What happens to established models of technology integration when Generative AI (GAI) no longer supplements language teaching but actively co-constructs it? The rapid uptake of Large Language Model (LLM)-based tools in professional language education exposes a conceptual gap: frameworks like SAMR, originally developed by Puentedura[1,2], were not designed for technologies that generate discourse, simulate interactional scenarios, and redistribute agency between teachers, learners, and machines.

GAI is already reshaping language instruction by acting as tutor, collaborator, and content generator. Pratschke (2024, p. 53) argues that GAI "requires a complete rethink of how educators deliver teaching," insisting on structured scaffolding to ensure pedagogical coherence. Concerns raised by Floridi (2024)

[1] Puentedura, R. R. (2006, November 28). *Transformation, technology, and education in the State of Maine* [Web log post]. http://www.hippasus.com/rrpweblog/archives/2006_11.html

[2] Puentedura, R. R. (2013, May 29). *SAMR: Moving from enhancement to transformatio*n [Web blog post]. Retrieved from http://www.hippasus.com/rrpweblog/archives/000095.html

and Alonso-Rodríguez (2024) highlight the risks of opaque decision-making, data misuse, and the amplification of social inequality. Kohnke et al. (2024) further draw attention to the phenomenon of "technostress," a form of cognitive overload experienced by educators using unfamiliar and fast-evolving AI tools.

Despite SAMR's widespread adoption in educational technology discourses, it has been criticized for its linear structure and tendency to oversimplify complex pedagogical change (Hamilton et al., 2016). Its application to GAI in professional language education has so far been limited and unsystematic. Boateng et al. (2024, p. 44) note that "the SAMR model can be seen as a catalytic method in the transformation of educational technology, providing a guide for transformative learning experiences for both students and teachers," yet this catalytic potential has not sufficiently been adapted to GAI-mediated LSP contexts.

We position this article as a conceptual perspective rather than an empirical study. Our aim is to stimulate debate by re-examining SAMR in light of GAI's affordances and constraints for Business LSP. To illustrate our argument, we draw on examples from Business German in higher education and Business Spanish in corporate training.

Literature Review

The integration of GAI into language education has drawn increasing scholarly attention because of its potential to reshape how languages are taught, learned, and applied in professional communication. In Business Language Education and Languages for Specific Purposes (LSP), GAI enables more personalized, context-sensitive instruction adopted to workplace discourse demands. Skrabut (2023, p. 131) observes that GAI can "support multiple stages of the learning process by adapting to individual learner needs". Its ability to replicate realistic professional contexts makes it particularly relevant for Business LSP. Pack and Maloney (2023, p. 5) reinforce this view, arguing that "the potential affordances generative AI may offer language educators are often overlooked, and more attention needs to be given to the work of educators and researchers who explore the positive potentials of these technologies." Chatbots provide consistent linguistic input and simulate authentic conversational settings (Wan & Moorhouse, 2024), a function especially relevant in business-oriented LSP contexts where learners must adhere to pragmatic norms and genre conventions. Son, Ruzic, and Philpott (2023) identify automated writing evaluation, adaptive feedback, and intelligent tutoring systems as emerging practices that support real-time monitoring and individualized instruction. Chen et al. (2025) highlight five digital affordances of ChatGPT, including opportunities for practice, immediate feedback, and learner autonomy. In a systematic review, Weng and Fu (2025) report that GAI can boost inclusivity and motivation by addressing diverse proficiency needs in language classrooms.

Taken together, these studies demonstrate that GAI can increase access, personalization, and interactivity in language learning. However, for Business LSP, unresolved questions remain about whether such affordances adequately address domain-specific communication challenges. These limits signal the need to revisit existing models of technology integration, particularly SAMR.

The SAMR model offers a widely adopted framework for categorizing technology integration into four levels: Substitution, Augmentation, Modification, and Redefinition. Boateng et al. (2024, p. 44) describe it as "a catalytic method in the transformation of educational technology, providing a guide for transformative learning experiences for both students and teachers." However, the model has been criticized for its linearity and oversimplification of complex pedagogical dynamics. Ethical and practical concerns complicate the integration of AI tools. Floridi (2024) and Alonso-Rodríguez (2024) warn of opaque decision-making processes, potential misuse of learner data, and the amplification of social inequities. Yan et al. (2023, p. 3) caution that "despite the growing empirical evidence of LLMs' potential in automating a wide range of educational tasks, none of the existing work has systematically reviewed the practical and ethical challenges of these LLMs-based innovations," adding that training on unfiltered data can reproduce "biased and toxic knowledge (e.g., gender and racial biases)." Dabis and Csáki (2024, p. 1) emphasize institutional accountability, asserting that "student assignments must reflect individual knowledge (…) with human individuals retaining moral and legal responsibility." Caines et al. (2023) call for transparency and fairness in AI-mediated instruction, while Sharples (2023, p. 7) argues that effective use of GAI "requires building GAI to follow fundamental human rights, respect the expertise of teachers and care for the diversity and development of students." Aoun (2017) proposes a model of "humanics" that balances digital innovation with ethical, creative, and cognitive development - an approach especially relevant in professional language learning where judgment and accountability cannot be automated. Kohnke et al. (2024) add another concern: "technostress," the cognitive overload educators experience when adapting to rapidly evolving digital tools.

Within this broader context, the application of AI to LSP, particularly in professional and business communication, emerges as a promising yet under-explored area. As Grib et al. (2024, p. 1) argue, "the one-size-fits-all approach of conventional curricula does not effectively address the varied learning styles, cultural context, and educational background of a global student body". AI allows instruction to be more responsive, flexible, and aligned with individual learner goals.

Nevertheless, effective AI integration in LSP instruction depends on several critical conditions. First, digital infrastructure and access remain uneven across educational

institutions, with disparities that particularly affect under-resourced learners[3]. Second, while GAI can promote the linguistic accuracy and efficiency of communication training, it cannot replicate the emotional intelligence and cultural sensitivity that are essential for professional discourse. As such, human oversight remains vital to ensure that LSP instruction remains pedagogically sound and culturally responsive.

The integration of AI into language education must be guided by a combination of ethical frameworks, regulatory standards, and pedagogical competencies. International and European policy documents, such as *The Artificial Intelligence Act*, European Commission[4]; *Artificial intelligence, human rights, democracy, and the rule of law - A primer* (Leslie at al., 2021), among others, offer a comprehensive structure for aligning AI use with educational values such as equity, transparency, and inclusion. When embedded thoughtfully into curriculum design, AI technologies have the potential to not only increase language proficiency but also to contribute to learners' success in multilingual and digitally mediated work environments. This broader landscape sets the stage for examining how existing pedagogical frameworks account for GAI's impact on learning design in LSP.

Debates around technology use in education have long drawn on models such as SAMR. Rather than rehearsing its structure, this article argues for a reinterpretation of SAMR in light of how GAI reshapes pedagogical decision-making in Business German and Business Spanish. Although widely adopted, the SAMR model has been critiqued for suggesting linear progression, and for reducing analysis to tool replacement rather than learning design (Hamilton et al., 2016). Studies in AI-supported language learning (e.g., Son et al., 2023; Wan & Moorhouse, 2024) confirm that LSP learners do not passively receive content from AI; they engage in generative co-production. This suggests that the pedagogical question is not "What level are we at?" but "What kind of redesign is taking place when AI intervenes in professional discourse tasks?" The SAMR model still offers value here, but only when treated as a heuristic for examining shifts in task structure, rather than a checklist for technology adoption.

Application of SAMR Business German and Spanish Instruction

In Business Language Education, SAMR offers a way to distinguish between AI uses that simply digitize existing practices and those that enable qualitatively new forms of

Table 1

SAMR-Based Mapping of AI-Supported Business German and Spanish Tasks

Context	Task	AI Tool	SAMR Level	Learning Goal	Assessment Signal	Ethical Checkpoint
Business German	Drafting a *Mahnung bei Lieferverzug* (delay notification)	AI Email Writer	Substitution	Practice formal register and business correspondence conventions	Accuracy of syntax, register adherence	Human review of sensitive content to prevent overreliance on automated outputs
Business Spanish	Preparing minutes of a procurement meeting	AI Grammar Checker / Summarizer	Augmentation	Enhance lexical precision and coherence in professional genres	Improved cohesion and consistency in terminology	Transparency about AI assistance to preserve authorship integrity
Business German	Collaborative CSR project design	Miro AI (brain-writing + prompts)	Modification	Develop critical reasoning, intercultural awareness, and CSR-related domain language	Quality of collaborative synthesis and use of specialized terminology	Ensuring equitable participation across learners with different digital literacies
Business Spanish	Simulated negotiation with international partners	Custom-built AI roleplay	Redefinition	Engage in dynamic, pragmatic interaction requiring mitigation strategies	Ability to manage ambiguity, adapt register, and achieve negotiation outcomes	Safeguarding cultural authenticity, preventing stereotyping in AI-generated roles

[3] United Nations Children's Fund and International Telecommunication Union (2020). *How many children and young people have internet access at home? Estimating digital connectivity during the COVID-19 pandemic*. UNICEF.

[4] European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (COM/2021/206 final). European Comission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

learning. Rather than assuming transformation is always superior, the framework helps instructors align technological use with genre-specific communicative goals in German and Spanish for professional contexts. The focus here is not on tool novelty but on how AI reshapes or maintains the structure of core tasks.

To make these distinctions concrete, Table 1 presents a SAMR-based overview of tasks common to Business German and Spanish training. Each example links an instructional objective to a genre-bound communicative form, an AI tool, an assigned SAMR level, and an ethical checkpoint. The entries are illustrative and serve to anchor the interpretive analysis that follows.

As illustrated in Table 1, Substitution tasks in Business German, such as drafting a *Mahnung bei Lieferverzug* (*reminder for delayed delivery*) with an AI Email Writer, replicate existing tasks, without altering their fundamental pedagogical objective. In this case, the learning outcome - mastery of formal register and specialized terminology (*Lieferverzug*, *Mahnung*, *Zahlungsziel*) - remains intact, while the technology simply facilitates more efficient drafting and editing. Similarly, in Business Spanish, drafting a procurement-related message with AI support would maintain the same communicative goals, namely grammatical accuracy, register adherence, and genre-specific conventions. Although innovation is minimal at this stage, Substitution represents a meaningful step toward the digitization of conventional business communication tasks.

As shown in Table 1, Augmentation tasks enrich traditional activities by introducing functional improvements that scaffold comprehension and vocabulary use. In the context of Business Spanish, learners can use *Mapify* (*mapify.ai*, s.d) to generate a mind map on *Responsabilidad Social Corporativa (RSC)* (Figure 1).

In a lesson focused on Corporate Social Responsibility (CSR), the mind map (Figure 1) presents a well-structured overview of the topic, dividing it into clearly defined branches such as *Definición de RSC*, *Principios*, *Estrategias de implementación*, *Áreas de impacto*, and *Ejemplos de buenas prácticas*. This visual representation enables learners to see how key CSR concepts interconnect thematically and hierarchically. For example, the map outlines core CSR principles, including ethical behavior, transparency, and environmental sustainability, and links these to concrete practices such as reducing carbon footprints or promoting the circular economy. It also visualizes implementation strategies, like defining measurable goals, involving all organizational levels, and publishing sustainability reports.

As Amelina et al. (2023, p. 52) argue, mind maps support a wide array of learning activities. They assist in organizing information, visualizing relationships among ideas, and planning presentations or projects. These features are especially beneficial in the LSP context, where structured thinking, topic-specific vocabulary, and clarity of expression are key to communicative competence.

The Modification stage enables substantial redesign of learning tasks. In this phase, learners move beyond basic enhancements to engage with content in entirely new ways. In a project on CSR, digital platforms like *Miro AI* (*Miro AI*, s.d.) facilitate collaborative, inquiry-driven learning. For example, students begin the unit with a brain-writing activity on a Miro board, individually posting their ideas about what CSR means and how companies contribute to social and environmental goals (Figure 3).

Students use the Miro board to organize their findings, including definitions, strategies, and impact areas, as identified in the Spanish-language mind map (Figure 1). Each group selects a company and critically evaluates its CSR efforts, including successes and potential gaps. AI-generated prompts within Miro help scaffold language production, encouraging students to use domain-specific vocabulary and connect arguments logically. This redesigned learning experience transforms CSR from a theoretical topic into a collaborative, action-oriented investigation that integrates critical thinking, intercultural awareness, and language production.

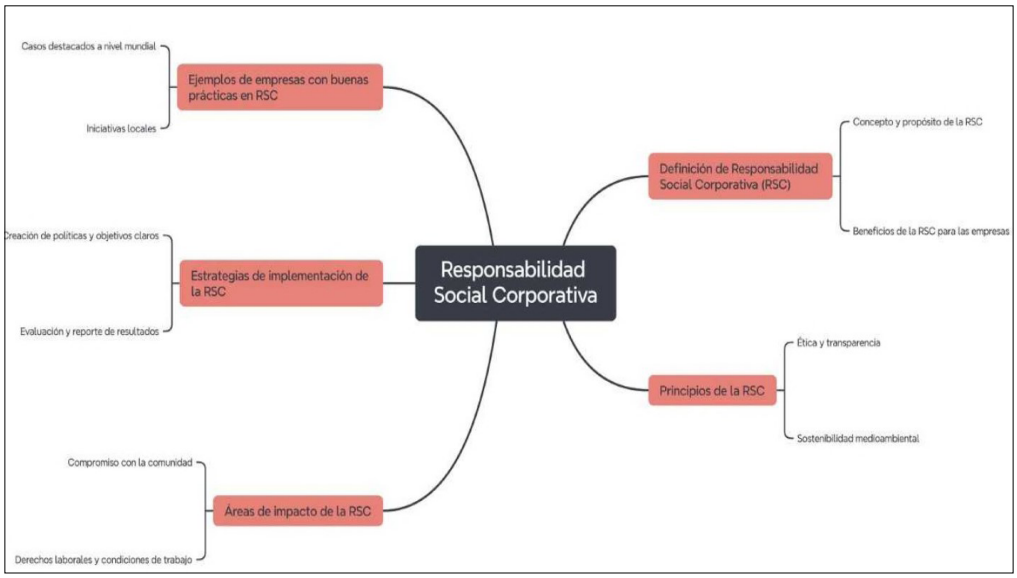
At the Redefinition stage of the SAMR model, GAI enables the design of interactive, collaborative language tasks that would have been logistically or pedagogically impractical without advanced technology. As Mollick & Mollick (2024, p.1) argue, "The transformative power of AI has put educators in the position of builders and creators, potentially democratizing the development of educational technology. Instead of having to choose from pre-built role-play experiences, you can more easily develop practice spaces and interactive solutions that better suit your learners". This shift allows instructors to custom-build AI-supported simulations.

For example, rather than merely reading scripted dialogues, learners can participate in dynamic simulations where they role-play as entrepreneurs pitching a new product to an international investor, or as marketing consultants negotiating contract terms with German or Spanish-speaking clients.

Peer review and AI feedback further augment the reflective and strategic dimensions of the task, supporting both linguistic improvement and metacognitive development. This redefined practice encourages learners to become not just consumers of AI-driven exercises, but co-designers of authentic communicative experiences in professional German and Spanish contexts.

Figure 1

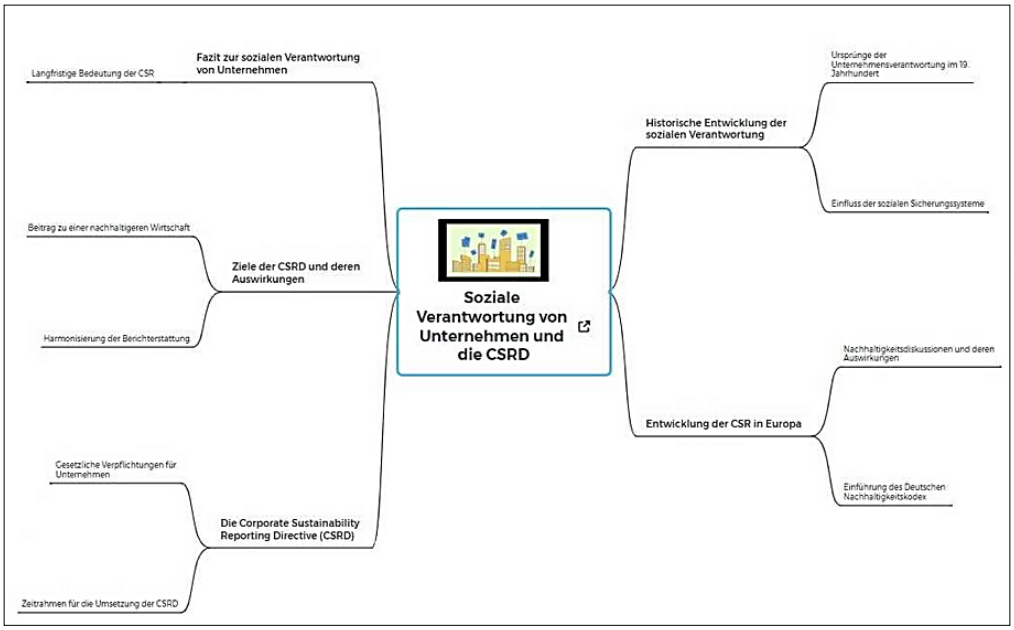
Digital Mind Map on Responsabilidad Social Corporativa



Note. Source: Mapify, auto-generated from Spanish input.

Figure 2

Mind Map on „Soziale Verantwortung von Unternehmen"



Note. Source: Mapify auto-generated from spoken video content.

DISCUSSION

AI tools frequently enact features from more than one SAMR stage simultaneously, which complicates linear interpretations of the model and underscores the need for a task-oriented approach. In this sense, the present paper aligns with Hamilton et al.'s (2016) critique and adapts SAMR as a heuristic rather than a hierarchy. Their position informs three design principles that can guide AI-supported LSP instruction: P1 process over product, P2 reflective scaffolding, and P3 collaborative inquiry.

Operationalising the principles in practice clarifies their relevance across SAMR stages. Under P1, learners are required to revise or justify AI-generated outputs rather than submit them unaltered, ensuring that drafting and correction remain cognitively active processes. Under P2, instructors integrate mandatory checkpoints—such as commentary on

Figure 3

Collaborative Brain-Writing Activity about CSR

Note. Source: Miro AI board used in CSR project.

AI suggestions—to foreground reflection rather than automation. Under P3, activities incorporate peer interaction or co-construction, whether in brainstorming, redrafting, or simulations, so that AI does not replace human dialogue but supports it.

These principles manifest differently across the model. At Substitution and Augmentation, grammar checkers, writing suggestions, and concept-mapping tools provide immediate feedback and structure, but learners must diagnose and adjust outputs, maintaining agency. At Modification, collaborative platforms such as Miro AI facilitate dialogic knowledge-building and distributed authorship, enabling students to integrate domain-specific content from authentic sources. At Redefinition, AI simulations generate professional scenarios that combine intercultural expectations, strategic language use, among others.

Two boundary conditions shape any flexible use of SAMR. First, uneven access to advanced systems limits the scalability of AI-supported task design. Second, domain-specific constraints in business communication require learners to cope with not only accuracy but also register, appropriateness, and cultural nuance. Assessment must therefore align with communicative outcomes rather than technology use, emphasizing move-structure completeness, mitigation or directness strategies.

This contribution is necessarily provisional. It offers a conceptual reframing supported by illustrative examples rather than empirical validation, and it is situated within a rapidly evolving technological landscape. Nevertheless, by grounding instructional design in Hamilton et al.'s principles and acknowledging infrastructural and pragmatic limits, this paper recasts SAMR as a flexible framework that can promote reflective, collaborative, and linguistically grounded task design in Business German and Spanish. Rather than classifying tools, it foregrounds learner agency, intercultural specificity, and process-oriented use of AI in professional language education.

CONCLUSION

SAMR must be reinterpreted in the context of GAI - not as a hierarchy of technological progress, but as a heuristic for aligning tools with the pragmatic, cognitive, and genre-specific demands of LSP instruction. As the application of AI in language education becomes more sophisticated, the boundaries between the SAMR stages tend to blur. This study illustrates that AI tools frequently operate across multiple levels of the model simultaneously, challenging the assumed linearity of progression. Therefore, rather than viewing SAMR as a rigid structure, it is more productive to interpret it as a flexible guide that encourage educators to

align technological tools with pedagogical goals, cognitive demands, and communicative contexts.

Through the case studies of Business German and Business Spanish instruction, it becomes evident that meaningful AI integration depends not on the tool itself but on how it is embedded within intentional instructional design. From supporting linguistic accuracy through automated feedback, to promoting collaborative inquiry and simulating authentic communication, GAI applications have demonstrated their capacity to transform LSP teaching into dynamic and context-rich learning environments.

To move meaningfully toward redefinition, educators must go beyond tool adoption and engage in pedagogical reimagination. This involves designing AI-integrated tasks that also cultivate metacognitive awareness, intercultural communication, and ethical engagement with AI systems.

When applied critically and contextually, the SAMR model offers a valuable perspective through which to evaluate the pedagogical affordances of GAI in business language education.

Across all SAMR levels, responsible AI use depends on three baseline practices: learners must revise and interrogate AI outputs rather than accept them wholesale; instructional design must make reflection visible through prompts, annotations, or peer review; and simulations or co-constructed tasks must preserve human agency and linguistic intentionality. Ethical engagement becomes part of instructional design, not compliance: disclosure of AI involvement, documentation of prompts and revisions, and scrutiny of generated language for bias or register drift should be standard.

The SAMR model retains its relevance in AI-mediated LSP only when detached from linear progression and aligned with communicative intent, task authenticity, and learner agency.

ACKNOWLEDGMENTS

The authors would like to thank the Centre for Languages, Literatures and Cultures at the University of Aveiro for the support of this research.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTIONS

Katrin Herget: conceptualization; data curation; formal analysis; funding acquisition; methodology; project administration; visualization; writing – original draft; writing – review & editing.

Katty da Silva Ferreira: formal analysis; investigation; methodology; resources; software; supervision; writing – original draft.

REFERENCES

Alonso-Rodríguez, A. M. (2024). Towards an Ethical Framework for Artificial Intelligence in Education. *Revista Interuniversitaria*, *36*(2), 79-98. https://doi.org/10.14201/teri.31821

Amelina, S. M., Tarasenko, R. O., & Semerikov, S. O. (2023). Enhancing foreign language learning with cloud-based mind mapping techniques. *Proceedings of the VIII International Workshop on Professional Retraining and Life-Long Learning using ICT: Person-oriented Approach* (vol. 3535, pp. 48-60). CEUR. https://doi.org/10.31812/123456789/8484

Aoun, J. E. (2017). *Robot-proof: Higher education in the age of artificial intelligence.* MIT Press.

Bax, S. (2018). MOOCs as a new technology: Approaches to normalising the MOOC experience for our learners. In *Flipping the blend through MOOCs, MALL and OIL – new directions in CALL* (pp. 9-16). Research-Publishing.net.

Boateng, S., & Kalonde, G. (2024). Exploring the synergy of the SAMR (Substitution, Augmentation, Modification, and Redefinition) model and technology integration in education: The key to unlocking student engagement and motivation. *Proceedings of the International Conference on Advanced Research in Education, Teaching, and Learning*, *1*(1), 37- 46. https://doi.org/10.33422/aretl.v1i1.185

Caines, A., Benedetto, L., Taslimipoor, S., Davis, C., Gao, Y., Andersen, Ø., Yanakoudakis, H., Nicholls, D. (2023). On the application of large language models for language teaching and assessment technology. *arXiv Preprint*. https://doi.org/10.48550/arXiv.2307.08393

Chen, Z., Wang, L., & Huang, W. (2025). Generative AI technology and language learning: A sequential mixed-method study of emotional self-reflection in ChatGPT-assisted language practice. *Interactive Learning Environments.* https://doi.org/10.1080/10494820.2025.2511248

Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications, 11*(1), Article 1006. https://doi.org/10.1057/s41599-024-03526-z

Floridi, L. (2024). The ethics of Artificial Intelligence: Exacerbated problems, renewed problems, unprecedented problems. *American Philosophical Quarterly*. http://dx.doi.org/10.2139/ssrn.4801799

Gallacher, A., Thompson, A., & Howarth, M. (2018). - My Robot is an Idiot! - Students' perceptions of AI in the L2 classroom. *Research presentation at EUROCALL International Convention*. https://doi.org/10.14705/rpnet.2018.26.815

Grib, M., Kirste, L., & Holtbrügge, D. (2024). The Use of AI in teaching international business: Making online lecturing more inclusive and interactive. *AIB Insights*, 25(1). https://doi.org/10.46697/001c.127712

Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) model: A critical review and suggestions for its use. *TechTrends*, *60*, 433–441. https://doi.org/10.1007/s11528-016-0091-y

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning - Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237-257. https://doi.org/10.1111/jcal.12610

Koehler, M.J., Mishra, P., Kereluik, K., Shin, T.S., Graham, C.R. (2014). The technological pedagogical content knowledge framework. In J. Spector, M. Merrill, J. Elen, M. Bishop (Eds.), *Handbook of research on educational communications and technology*. Springer. https://doi.org/10.1007/978-1-4614-3185-5_9

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70.

Koehler, M.J., Mishra, P., & Cain, W. (2013). What is Technological Pedagogical Content Knowledge (TPACK)? *Journal of Education*, *193*(3), 13-19. https://doi.org/10.1177/002205741319300303

Kohnke, L., Moorhouse, BL., & Zou, D. (2023). Using Chat GPT for language teaching and learning. *RELC Journal*, 54(2), 537-550. https://doi.org/10.1177/00336882231162868

Kohnke, L., Zou, D., & Moorhouse, B.L. (2024). Technostress and English language teaching in the age of generative AI. *Educational Technology & Society*, *27*(2), 306-320. https://doi.org/10.30191/ETS.202404_27(2).TP02

Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., & Briggs, M. (2021). *Artificial intelligence, human rights, democracy, and the rule of law: A primer*. Council of Europe.

Mollick, E. R., & Mollick, L. (April 22, 2024). *Instructors as innovators: A future-focused approach to new AI learning opportunities.* The Wharton School Research Paper. http://dx.doi.org/10.2139/ssrn.4802463

Pack, A., & Maloney, J. (2023). Potential affordances of generative AI in language education: Demonstrations and an evaluative framework. *Teaching English with Technology, 23*(2), 4–24. https://doi.org/10.1177/10.56297/buka4060/vrro1747

Pratschke, M. (2024). *Generative AI and education. Digital pedagogies, teaching innovation and learning design*. Springer.

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *arXiv Preprint*. https://doi.org/10.48550/arXiv.2306.10063

Skrabut, S. (2023). *80 Ways to use ChatGPT in the classroom.* Stanley A. Skrabut. Son, J.-B., Ruzic, N., & Philpott, A. (2023). Artificial intelligence technologies and applications for language learning and teaching. *Journal of China Computer-Assisted Language Learning.* https://doi.org/10.1515/jccall-2023-0015

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, *6*(1), 1-10. https://doi.org/10.37074/jalt.2023.6.1.17

Wan, Y., & Moorhouse, B. L. (2024). Using Call Annie as a generative artificial intelligence speaking partner for language learners. *RELC Journal*. https://doi.org/10.1177/00336882231224813

Weng, Z., & Fu, Y. (2025). Generative AI in language education: Bridging divide and fostering inclusivity. *International Journal of Technology in Education*, *8*(2), 395-420. https://doi.org/10.46328/ijte.1056

Yan, L., Sha, L., Zhao, L., Li, Y., MartinezMaldonado, R., Chen, G., Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *arXiv Preprint*. https://doi.org/10.48550/arXiv.2303.13379

https://doi.org/10.17323/jle.2025.27821

Visible Learning: The Sequel: A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement: A Book Review

Taoufik Boulhrir

Fordham University, USA

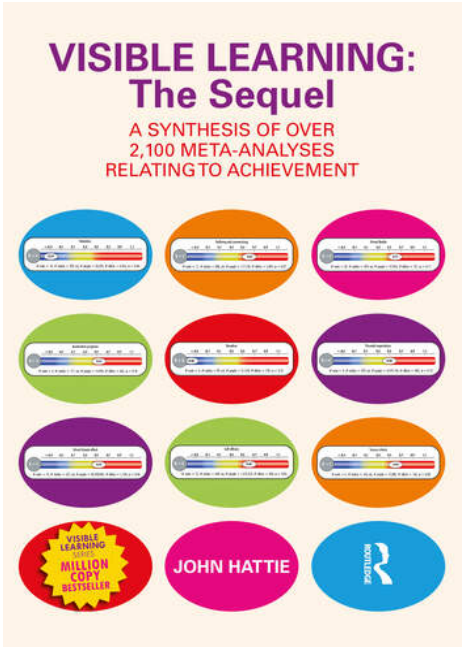
Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement, Written by John Hattie, London: ROUTLEDGE, 2023, 512 pp., (e-Book), ISBN: 9781003380542

INTRODUCTION

John Hattie's Visible Learning: The Sequel is a follow-up of his earlier research (Hattie, 2008). This update builds on the massive success of the earlier work by synthesizing evidence from over 2,100 meta-analyses, representing over 130,000 studies and over 400 million students worldwide. It does more than just update the original; it revises and refines the underlying concepts of VL, addresses the successes and limitations of the original and responds to some of the criticism it has received since 2008. This review therefore explores Hattie's latest meta-analysis to provide a summary of his findings and what they mean for practice. The sequel challenges the status quo and provides a practical framework for continuous improvement and this review aims to provide researchers and education practitioners with a comprehensive look at the key arguments, strengths and weaknesses of the book and how the findings can be applied to practice and future research. Hattie also significantly expands the database to include domains like digital technologies, motivation, and socio-emotional learning. The chapters, hence, are reviewed thematically rather than sequentially; chapters from different parts of the book may be synthesized together if they share a common theme and/or purpose.

Introducing VL and Responding to Criticism

The book is organized into 16 chapters, the first of which Hattie uses to introduce VL and its development since 2008. He emphasizes the shift from knowing "what works" to "what works best" (p.3) and responds to key criticisms, such as Terhart's (2011) skepticism about the validity of learning as the dependent variable and the model's challenge to traditional European views of teaching. Rømer (2018) similarly critiques VL's constructivist basis and its reductive notion of feedback. The chapter begins with outlining the problems educators face in improving student achievement. Hattie says that traditional education focuses on what teachers do rather than what students learn, and that we need an evidence-based approach to education that focuses on identifying and implementing the best teaching and learning strate-



Citation: Boulhrir, T. (2025). [Review of the book Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement by J. Hattie]. *Journal of Language and Education, 11*(3), 168-172. https://doi.org/10.17323/jle.2025.27821

Correspondence:
Taoufik Boulhrir,
boulhrir@fordham.edu

Received: July 31, 2025
Accepted: September 18, 2025
Published: September 30, 2025

gies. It also responds to critiques who think the model is too simplistic or that it ignores context.

Hattie explains, "a mistake of many critics is to overly focus on the one book in 2008, as I have expanded, clarified, and explored many of the ideas in VL1 in these other sources" (p.1). Here, he indicates that the model is not a one-size-fits-all solution but a framework for understanding the factors that contribute to student achievement. He also acknowledges the need to consider context as key for effective implementation of the principles. Although this chapter sets the scene for the rest of the book, an overview of the VL model, its key principles, the challenges, and criticisms of implementation; readers may benefit from a more detailed overview of the book's structure and the specific educational interventions to be discussed including how theoretical concepts apply to educational factors or interventions. His revised synthesis may therefore be understood not only as an academic contribution through the aggregation of empirical findings, but also as a methodological intervention into the politics of educational evidence. This positioning places VL within a broader academic and policy-driven efforts to formalize educational improvement through empirical aggregation and ranked effectiveness.

The VL project is a global initiative that engages with evidence-based educational policy designed and developed by organizations such as the What Works Clearinghouse (WWC) and the Organization for Economic Co-operation and Development (OECD). While these organizations have somewhat distinct scopes and methodological emphases, they share a common goal of being dedicated to syntheses and aggregation of empirical evidence primarily aimed at enriching teaching and better learning outcomes. Hattie's methodological approach favors statistical generalization through meta-analysis over narrative or primary study review, aligning with the WWC's emphasis on evidence from randomized controlled trials and quasi-experimental designs (Slavin, 2020). In contrast to the OECD's focus on large-scale structural determinants of learning such as national socio-economic and political systems, Hattie's framework prioritizes intra-school variables, particularly teacher-led practices (Grey, 2020).

Theoretical Foundations

Chapters 2 and 3 collectively lay down the theoretical background that enables the understanding and application of the VL model in its entirety. Chapter 2 focuses on the methodology of meta-analysis as a statistical technique used to synthesize the results of several studies and, particularly, explore the effects of different educational interventions. This process often uses statistical indicators like Cohen's d to measure effect size, where values above 0.40 are generally considered to indicate a meaningful impact. Hattie adopts this 0.40 benchmark to signal influences that exceed the 'average zone of desired effects.' Moreover, the R-value is introduced to gauge the replicability and robustness of these findings across contexts. Here, Hattie explains how meta-analysis provides estimates of the overall impact of these interventions and dispels criticism that it is too simplistic and contextually blind. He goes further to illustrate how VL has developed an approach that keeps refining itself with ongoing research and feedback.

Hattie revisits the VL model in Chapter 3 and categorizes what influences student achievement into eight explicit areas: the student, the home, the classroom, the school, the curriculum, the teacher, teaching strategies, and learning strategies. In addition, the chapter includes concrete examples for educators about how the model can help in understanding students' struggles and putting effective interventions in place. It must be noted, however, that the second chapter addresses researchers, who are statistically inclined, more than educators who might be overwhelmed by the technical details. Although the third chapter delves into practical details, it responds more to criticism than explains how schools and teachers can effectively implement the VL model.

Implementing Visible Learning

Chapters 4 and 11 concretize the real-world structure of how to implement and optimize the VL model in school environments. Chapter 4, «*Implementing the Intentional Alignment VL Model*,» describes steps necessary for the alignment of teaching and learning strategies to produce specific intended learning outcomes. It argues for clarity in understanding goals of learning and instructional strategies aligned to meet those goals. It also addresses the roles of various stakeholders, including teachers, students, leaders, and parents, with some useful practical advice on how to involve each of these groups in a cohesive way that ensures focus on common objectives. Yet, the actual implementation of purposeful and evidence-based teaching strategies is covered in Chapter 11»*Teaching with Intent*». It highlights that setting clear learning objectives, choice of appropriate teaching methods, and consistent monitoring of student progress by use of formative assessments are important in adapting instruction to student needs. Speaking of which, student feedback is reported here to be the backbone of instructional adaptation. These chapters complement each other and placing them sequentially in the book would enhance the reader's understanding by pairing practical applications with a focus on intentional teaching. Although they write more to the academic community, teachers should not expect specific implementation examples; the chapters simply link the theoretical constructs to practical strategies with the hope of improving student learning and their own teaching environment. To facilitate implementation, teachers can consider experimenting with strategies that Hattie identifies as being very high impact, such as formative assessment, teacher clarity, and classroom discussion. These

could take the form of concrete practices like exit tickets, explicit learning intentions, or scaffolding peer discussion.

Learners and Learning Environment

To discuss the learners and learning environment, Hattie provides a comprehensive explanation of the critical factors influencing student achievement in chapters 5-7 and 15 which explore how various internal and external factors interact to shape student achievement. Chapter 5, «students,» is focused on personal characteristics such as motivation, self-efficacy, prior-knowledge, and learning style. Hattie stresses the value of responsive instruction to students demonstrating self-confidence and adaptive learning, and he indicates the value of the growth mindset, which «puts students in a position in which they can look at difficulties as opportunities for growth.» Chapter 6, «The Home and Family,» focuses on family and socio-economic drivers of learning, arguing that «active parents, supportive home environments, and adequate access to resources significantly boost student achievement.

Chapter 7, "School and Society," extends this analysis to institutional and community supports, stressing that "a positive school environment and quality teacher influence, with community support," contribute meaningfully to outcomes. Hattie also identifies extracurricular and real-world learning experiences as valuable supplements to classroom instruction. Notably, these ideas reappear in Chapter 15, "Whole-School and Out-of-School Influences," where school leadership and broader environments are positioned as levers for educational equity. As he notes, "School-wide initiatives and out-of-school influences [...] are to reduce the achievement gaps among groups of students or increase the impact on students with specific learning needs" (p. 411).

From the reviewer's perspective, while these chapters collectively set out the ecological richness of student learning, their internal sequence is incoherent. Chapter 7's break with 5–6 discontinues the thematic sequence, and the conceptual duplication with Chapter 15 dilutes the strength of both. More significantly, while the evidence base is robust, the text makes very few specific suggestions as to how teachers could carry these findings forward into school-level interventions. The discussion remains highly abstract, and practitioners need to bridge the gap between concept and practice themselves.

Classroom Practices

Chapters 8, 9, 12, and 13 collectively address classroom practice in terms of instruction and how this influences student achievement. Rather than individually describing each component, these chapters emphasize how classroom climate, teacher effectiveness, and student-centered approaches to instruction operate and impact success. Hattie stresses the importance of teacher clarity, positive classroom management, and high expectations in inclusive and heterogeneous learning environments. In the reviewer's view, even if these claims are in line with pre-existing literature on quality teaching, treatment of such topics is shallower in terms of practical scaffolding.

Chapters 12 and 13, for example, stress metacognition, self-regulation, and self-motivation as critical to academic success but have little to say about how these are implemented in practice classrooms. The book appears to assume a professional level of sophistication that many may not possess, especially in the taking up of abstract strategies without practical examples. This limits the application of Hattie's otherwise solidly grounded conclusions to those practitioner teachers in search of pragmatic approaches based on the realities of curriculum design and time constraints.

Curriculum

Curriculum is addressed in chapter 10 as a mediating structure between teaching, learning, and assessment. Hattie names four dimensions, clarity, coherence, relevance, and challenge, as essential to a high-quality curriculum. Clarity requires both students and teachers to share a clear, unwavering understanding of content and intent of the curriculum. Coherence requires systematic relationships between concepts and topics, and relevance entails consistency with what students are experiencing and concerned about. Challenge, on the other hand, is the level of mental effort to sustain engagement and enable development. These are presented as dependent dimensions which require intentional design and regular scrutiny.

From a practical perspective, Hattie's point to assert that curriculum development should be done collaboratively and rooted in classroom realities is valid, in the sense that he maintains that teachers ought to be at the center of this exercise, considering that they have special insights into the feasibility of instruction and the needs of students. This focus, however, differs from practices in most education systems, where policy makers and administrative centers end up commanding curriculum design. The book works this gap by advocating additional teacher participation in decision-making, but not to the point of offering structural recommendations or case examples demonstrating how it would work in practice. The argument, therefore, is inspirational rather than a practical one. Including more specificity in this chapter through actual models for participatory cur-

riculum planning would make it of more value to practitioners as well as to proponents of reform.

Technology Integration

Technology integration is discussed in Chapter 14 in terms of how technology supports student learning, including its ability to enhance good teaching instead of replace it. Hattie provides guiding principles for the application of educational technology, noting the necessity of connecting tools with good pedagogy. The chapter discusses the ways technology can improve formative feedback, working together, and access to rich learning environments, particularly through mediums such as social media or systems that provide feedback.

According to Hattie, technology in the classroom is relatively a question of how its impact on student learning and developing effective teaching practices are integrated together. Although he does not explain how different technological advances can be used to improve learning, he focuses on several key principles in using technology effectively: it should enhance, not deliver, content; it must complement a diverse range of teaching strategies and by no means replace them; and teachers need proper training to use technology effectively.

While the inclusion of this technology takes in the new scope of Hattie's synthesis, the coverage remains conceptual. The chapter lacks concrete examples of the effective use of specific technologies in varying classroom contexts. Educators who read the book, from a reviewer's point of view, would benefit from clearer illustrations of how to integrate tools in alignment with the high-impact strategies identified elsewhere in the book, such as metacognition or formative assessment. There is also minimal attention given to the challenges of technology adoption, such as training, infrastructure, and digital equity, which may limit the practical utility of the chapter for teachers and school leaders seeking guidance.

Style and Language in VL: The Sequel

The language used in the book is clear and concise which enhances accessibility to the diverse audience, including researchers, educators, and policymakers. Occasionally, anecdotes, case studies, and humor are used which help to make it readable and engaging. Although Hattie takes on a very direct and engaging style of writing that avoids jargon and technical terms so that content is approachable for readers who are less familiar with educational research, there are several instances where the style is burdened with technical explanations of data that may not address most practitioners' needs. Furthermore, although the organization of the chapters could benefit from better sequence and order in the book for enhanced continuity between them, each chapter's structure is well-organized. Each chapter is fitted with *"conclusions"* and *"reflections"* which are found to be direct and straightforward a variety of readership for an in-a-nutshell glimpse or main takeaways of the chapters. The use of headings, subheadings, bullet points, and lists makes it easy to navigate and locate information». Such well-thought-out presentation not only makes the book user-friendly but also shows its practical relevance.

The fact that the book provides a wealth of information that is helpful to a variety of audiences, including researchers, education leaders, teachers, and policymakers, generally makes it somehow challenging to navigate. Depending on the purpose of the book, the reader may need a 'How To Use This Book' section to guide different readerships to their respective needs (e.g., researchers vs. teachers). Such a section might direct researchers to the chapters on meta-analytic methodology and evidence classification, while guiding teachers toward practice-oriented sections on classroom effects, learning strategies, and implementation models.

The book's strengths are its massive scope, derived from over 2,100 meta-analyses of studies involving over 400 million students, and its thematic consistency, with chapters addressing rational categories of learners, teaching, and curriculum. It constructively builds on criticisms and develops the VL model to include new fields like digital technology, socio-emotional learning, and school-level factors. Although terminology is more accessible to academic readers than to education practitioners, the inclusion of case studies and bullet points helps to improve overall clarity. In the meantime, the book has several limitations. It is not overly forthcoming with pragmatic implementation strategies for practitioners. Certain sections, particularly those dealing with statistical thresholds, are likely to alienate non-specialist readers. There are a few structural discontinuities that disrupt the flow of the whole, and the emphasis on «what works» is not always followed through with sufficient interpretive depth regarding how or why particular strategies operate in varied educational settings.

This Research-to-practice gap is evident in sections where high-level models are introduced with limited procedural elaboration. For instance, the "intentional alignment" framework includes a five-phase cycle: Discover, Design, Deliver, Double-back, and Double-up, but its implementation in concrete classroom routines and instructional practices remains largely implicit. While the "Double-back" phase emphasizes that all stakeholders (i.e. students, teachers, and school leaders) need to systematically and deliberately mon-

itor and evaluate the impact of all the factors on the learning journeys, such guidance remains abstract and undeveloped at the level of practical enactment. Hattie's endorsement of teacher mind frames is another example; the statement "I explicitly inform students what successful impact looks like from the outset" (p. 47), is not accompanied by sample protocols, diagnostic tools, or illustrative enactments. These instances exemplify the broader pattern in The Sequel, where strong empirical syntheses are not always matched by equally robust interpretive or operational scaffolds to support implementation across varied educational contexts. Hence, this work is strongest in its synthesis of evidence, but its practical utility would be enhanced by more explicit attention to the classroom-level constraints typically navigated by mediating actors such as schools, educational leaders, and teacher preparation programs, for whom additional scaffolding within the model would be valuable support.

CONCLUSION

The book revisits and updates key conclusions drawn in the original VL model, as it reflects on the 14-year journey of VL and its significant impact on education practice. It updates initial conclusions with key findings from the experience of over 10,000 schools implementing the VL approach. Hattie's recent deliberations emphasize that, for achievement in students, it is essential to have positive and caring learning environments, to keep expectations high, and to develop critical thinking. Although the book presents a detailed overview of the findings, there is a lack of sequential organization in some chapters, potentially hindering readers from following the evolution of ideas in a coherent manner. Furthermore, most chapters are found to be devoid of real-life examples of implementable strategies and their fulfillment, which would go a long way in bridging the theory-practice or research-teaching and learning divide. Although Hattie expounds on his previous advocated principles by moving from 'what works' to 'what works best,' which is useful, talking about 'how what works best works' could potentially be more enlightening. A «How To Use This Book» section would also be helpful in guiding different types of readers on how to utilize the material in the book based on their specific requirements (e.g., researchers vs. teachers). This book review is addressed primarily to education practitioners, for it outlines the main arguments and raises several limitations, particularly in addressing the practical interests of classroom teachers who very often need help with specific applications of research findings.

DECLARATION OF COMPETING INTEREST

None declared.

REFERENCES

Grey, S. (2020). [A book review: World class: how to build a 21st-century school system: Strong performers and successful reformers in education: by A. Schleicher]. *Comparative Education, 56*(2), 305–307. https://doi.org/10.1080/03050068.2020.1741792

Hattie, J. (2023). Visible Learning: The Sequel: A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement (1st ed.). Routledge. https://doi.org/10.4324/9781003380542

Hattie, J. (2008). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement (1st ed.). Routledge. https://doi.org/10.4324/9780203887332

Terhart, E. (2011). Has John Hattie really found the holy grail of research on teaching? An extended review of visible learning. *Journal of Curriculum Studies, 43*(3), 425–438.https://doi.org/10.1080/00220272.2011.576774

Rømer, T. A. (2018). A critique of John Hattie's theory of Visible Learning. *Educational Philosophy and Theory*, *51*(6), 587–598. https://doi.org/10.1080/00131857.2018.1488216

Slavin, R. E. (2019). How evidence-based reform will transform research and practice in education. *Educational Psychologist, 55*(1), 21–31. https://doi.org/10.1080/00461520.2019.1611432

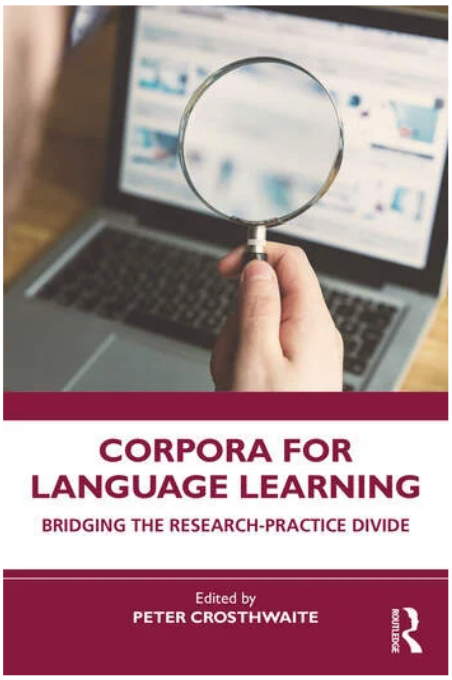
https://doi.org/10.17323/jle.2025.26958

Corpora for Language Learning: Bridging the Research-Practice Divide: A Book Review

Nguyen Huu Chanh

University of Health Sciences, Viet Nam National University Ho Chi Minh City

Corpora for language learning: Bridging the research-practice divide by Peter Crosthwaite, New York: Routledge, 2024, 288 Pp., (Paperback), ISBN: 9781032537221



Corpus linguistics has gained considerable attention in applied linguistics over recent decades, particularly for its potential to inform and improve language teaching and learning pedagogy (Conrad, 2005; Stoykova, 2014; Yan & Ma, 2025). However, much of the existing literature tends to focus on theoretical or experimental research, with less attention on how data-driven approaches are systematically integrated into everyday classroom practice. To fill this gap, *Corpora for Language Learning: Bridging the Research-Practice Divide* by Peter Crosthwaite, is committed to examine the challenges of applying corpus-based methods in practical teaching contexts. It provides a detailed account of various corpus tools, methods, and case studies that aim to support language educators, researchers, and practitioners in utilizing corpus linguistics more effectively. The volume contributes to the field by offering concrete examples of how corpus data can inform language acquisition and instruction, especially in vocabulary development, academic writing, and argumentation skills.

This volume provides a detailed examination of corpus-assisted language learning and its practical applications within language pedagogy, with a focus on narrowing the divide between research and classroom implementation. It addresses a broad spectrum of topics, including foundational concepts of Data-Driven Learning, the use of specific corpus analysis tools, and the incorporation of corpus-based methods across diverse educational settings such as teacher training, assessment, and language-specific contexts. The chapters collectively emphasize the potential of corpus linguistics to inform language teaching practices, while also discussing the roles of technology, open educational resources, and promoting learner autonomy.

Several chapters focus on corpus tools that support Data-Driven Learning (DDL) in classrooms. Laurence Anthony's AntConc remains a foundational, accessible concordancer, while Ana Frankenberg-Garcia's ColloCaid specifically targets academic collocations, a crucial skill for advanced learners. Tatyana Karpenko-Seccombe offers practical corpus-based activities to develop argumentation skills, expanding pedagogical uses of corpora. Tony Berber Sardinha's exploration of multimodal corpora in-

Citation: Huu Chanh, N. (2025). [Review of the book Corpora for language learning: Bridging the research-practice divide by P. Crosthwaite]. *Journal of Language and Education, 11*(3), 173-175. https://doi.org/10.17323/jle.2025.26958

Correspondence:
Nguyen Huu Chanh,
nguyenhuuchanh06@gmail.com

Received: May 08, 2025
Accepted: September 18, 2025
Published: September 30, 2025

troduces innovative perspectives by integrating visual elements, though this approach may require additional classroom scaffolding. Overall, these chapters provide valuable resources but would benefit from deeper discussion on practical classroom integration to fully address both research and practice.

A core theme is the application of corpus linguistics to enhance language competencies. Luciana Forti expands DDL beyond English, addressing an important research gap. Clarence Green highlights corpus-informed curriculum decisions for vocabulary acquisition through extensive reading and viewing, while Reka Jablonkai and Adriane Orenha-Ottaiano focus on corpus-supported teaching and phraseological competence development. Paula Tavares Pinto connects DDL with global education goals, adding a socially relevant dimension. This group showcases strong pedagogical innovation, translating corpus research into actionable teaching strategies, though more attention to learner engagement and challenges would strengthen their practical relevance.

This section addresses the crucial role of educators in adopting corpus-based methods. Qing Ma advocates equipping pre- and in-service teachers with corpus literacy effectively. Agnieszka Leńko-Szymańska offers corpus-based approaches for more rigorous vocabulary assessment. Elen Le Foll emphasizes open science and education to enhance accessibility and sustainability of corpus resources, while Tove Larsson and Douglas Biber call for methodological transparency to ensure linguistic interpretability. These chapters highlight systemic factors essential to bridging the research-teaching gap, combining theory and ethics, though more practical examples would enhance their classroom impact.

The final theme broadens DDL beyond formal settings into informal and lifelong learning contexts. Peter Crosthwaite and Alex Boulton set the stage by emphasizing collaboration and accessibility. Pascual Pérez-Paredes introduces Broad Data-Driven Learning (BDDL), reflecting the rise of digital literacies in informal environments. Vander Viana promotes autonomous corpus use in English for Academic Purposes, acknowledging current limitations but advocating for learner empowerment. These chapters push DDL's boundaries and highlight adaptability as key to closing the research-practice gap, though further guidance on supporting informal learners would strengthen these forward-looking insights.

This volume makes a significant contribution to the field of corpus-assisted language learning by exploring practically and comprehensively how corpus linguistics can be integrated into language teaching. One of its greatest strengths lies in thoroughly applying corpus-based methods to address concrete classroom challenges such as vocabulary acquisition, academic writing, and argumentation skills. The volume excels in showcasing concrete tools, such as AntConc (e.g., Chapter 2 by Laurence Anthony), ColloCaid (e.g., Chapter 6 by Ana Frankenberg-Garcia), and Lextutor Concordancer (e.g., Chapter 9 by Tatyana Karpenko-Seccombe), and providing case studies and practical activities that make corpus linguistics accessible and actionable for educators. By linking theory with real-life classroom applications, the volume helps address the often-cited research-practice gap, offering educators tangible ways to enrich their teaching with authentic language data.

Further strength of the volume is its detailed attention to corpus tools and resources that support various aspects of language learning. Chapters dedicated to Data-Driven Learning (DDL) approaches (e.g., Chapter 16 by Paula Tavares Pinto) emphasize not only the functionality of tools such as AntConc for corpus analysis but also how these resources can be embedded into lesson plans to enhance learning outcomes. The ColloCaid tool's role in improving collocational accuracy in academic writing and the use of extensive reading and viewing to support vocabulary growth also demonstrate the volume's capacity to address diverse language skills with corpus technology. Including the practical guidance tailored to different learner profiles and teaching contexts, such as English for Specific Purposes, makes these resources both relevant and versatile for a range of language educators.

Moreover, the volume stands out for its focus on the pedagogical implications of corpus linguistics, particularly its potential to foster learner autonomy (e.g., Chapter 8 by Reka R. Jablonkai) and evidence-based instruction (e.g., Chapter 12 by Agnieszka Leńko-Szymańska). By encouraging learners to engage directly with language patterns and corpus data, the volume advocates for a more active and informed approach to language learning. This learner-centered orientation is well supported by chapters that explore not only tool usage but also instructional design and task development, allowing teachers to create tailored, needs-based lessons. Emphasizing transparency in linguistic research methods and the promotion of open science and education also contribute to building a sustainable framework for ongoing teacher development and learner engagement with corpus linguistics.

Despite these considerable and valuable contributions, the volume has some minor limitations that affect its overall applicability and scholarly rigor. To begin with, some chapters (e.g., Chapter 2 by Laurence Anthony and Chapter 11 by Elen Le Foll) tend to focus primarily on their contexts with reliable technological infrastructure, offering little discussion on how corpus methods can be adapted to diverse teaching environments, especially for the educators with restricted resources or institutional constraints. This narrow contextualization reduces the volume's practical value for many educators working in less technologically equipped or lower-resource settings. Future edition could provide sample activities, scaffolding methods, and frameworks for educators unfamiliar with corpus methods.

Furthermore, while some chapters provide strong empirical support through well-documented case studies, others (e.g., Chapter 4 by Alex Boulton) rely mainly on descriptive accounts without much sufficient data to substantiate their claims, which weakens the evidence base for certain pedagogical recommendations. Hence, future editions could offer more data illustrations and teaching tools that advocates the conversations closer to the educators' lived experiences.

Additionally, although the theoretical depth and pedagogical applicability are addressed throughout the volume, the highly technical characteristics of some chapters (e.g., Chapter 10 by Tove Larsson and Douglas Biber) may alienate teachers without specialized training, and some chapters can benefit from a more rigorous theoretical grounding. Therefore, upcoming editions could consider pedagogical summaries or simplified instances to illustrate how the claims and critiques can translate into daily teaching and material developments. Solving these limitations in the following edition would provide a more balanced, nuanced, and globally relevant resource for corpus-assisted language learning.

Overall, *Corpora for Language Learning: Bridging the Research-Practice Divide* makes a valuable contribution to the ongoing dialogue in corpus linguistics and Data-Driven Learning by explicitly focusing on addressing the persistent gap between research and classroom practice. Unlike earlier foundational works such as Johns (1990), which primarily introduced corpus methods into academic language teaching, or Boulton (2010), who addressed accessibility barriers to DDL through innovative delivery formats, this volume consolidates these insights while expanding their application across diverse teaching contexts and learner needs. Moreover, it responds to recent calls for deeper theoretical engagement, as advocated by O'Keeffe (2021), by combining empirical case studies with pedagogical reflection and open science principles.

What distinguishes this volume is its practical and comprehensive orientation, providing educators, from novices to seasoned professionals, practitioners, educators, and researchers with concrete tools, resources, and evidence-based strategies that integrate meaningful corpus into in everyday teaching. Rather than simply revisiting existing knowledge, it pushes the field forward by emphasizing collaborative, transparent, and sustainable research-practice partnerships and by embracing digital literacies and broader educational agendas such as internationalization and lifelong learning. In doing so, the volume offers a comprehensive resource that not only consolidates established DDL methodologies but also charts new pathways for their evolution within modern language education.

ACKNOWLEDGMENTS

I would like to thank the anonymous reviewers for their thoughtful comments and valuable feedback, which greatly improved the clarity and depth of this review.

DECLARATION OF COMPETING INTEREST

None declared.

REFERENCES

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning, 60*(3), 534-572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393–410). Lawrence Erlbaum.

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, *10*, 14–34.

O'keeffe, A. (2021). Data-driven learning–a call for a broader research gaze. *Language Teaching, 54*(2), 259-272. https://doi.org/10.1017/S0261444820000245

Stoykova, V. (2014). Teaching corpus linguistics. *Procedia-Social and Behavioral Sciences*, *143*, 437-441. https://doi.org/10.1016/j.sbspro.2014.07.513

Yan, J., & Ma, Q. (2025). Theory-supported corpus pedagogy for ESL pre-service teachers: using Parallel EAP Corpora for language learning. *Journal of China Computer-Assisted Language Learning*. https://doi.org/10.1515/jccall-2024-0016

CONTENT

EDITORIAL

RESEARCH PAPERS

OPINION PAPERS

BOOK REVIEWS