ISSN 2411-7390

jle

JOURNAL OF LANGUAGE & EDUCATION

Volume 9 Issue 3, 2023

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

EDITOR-IN-CHIEF

EDITORIAL BOARD

EDITORIAL TEAM

CONTENT

EDITORIAL

RESEARCH ARTICLES

OPINION PAPERS

https://doi.org/10.17323/jle.2023.18119

ChatGPT: Where Is a Silver Lining? Exploring the realm of GPT and Large Language Models

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

[1] National Research University Higher School of Economics

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

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

ABSTRACT

Introduction: the JLE editors analyse the scope and depth of the subject area of ChatGPT and related topics based on the Scopus database. The Scopus statistics prove a skyrocketing rise in the number of publications in the field in question during 2023. The major alarming themes cover authorship and integrity related to AI-assisted writing, threats to educational practices, medicine, and malevolent uses of ChatGPT.

Keywords Explained: the key terminology is defined, including generative pre-trained transformers (GPT); ChatGPT; artificial intelligence (AI); AI chatbots; natural language processing (NLP); large language models; Open AI; large language model (LLM).

International Research on ChatGPT: as of September 24 2023, the Scopus database has indexed 1,935 publications, with "ChatGPT" in the title, abstract, or keywords. A skyrocketing rise in the number of research has been reported since the early days of 2023. 1,925 indexed publications out of 1,935 were published in 2023. Most of them came from the USA, India, the UK, and China. The number of documents indexed in the Scopus database as well as PubMed, arXiv and others are exponentially rising.

ChatGPT in Education: the academic community has been actively discussing the challenges education will face in the era of ChatGPT in the context of the fundamental threats posed to the educational system. The latter include assessment procedures, information accuracy, and skill devaluation. As many complex technologies, generative pre-trained transformers are ambivalent in nature, providing a great potential for learning and education at large, including new approaches based on critical thinking and awareness of the pros and cons of AI.

ChatGPT in Science: great prospects for text generation and improvements in language quality adjoin to dubious authorship and potentially inconsistent and erroneous parts in the AI-produced texts. Publishers and journals are working out new publishing policies, including publishing ethics towards AI-assisted or AI-improved submissions.

Conclusion: JLE is planning to revise its editorial policy to address the new challenges from AI technologies. JLE editors welcome new submissions of research articles and reviews as well as special issues on ChatGPT and related themes, with potential applications of chatbots in education, innovative approaches to writing assignments, facilitating personalized learning, academic integrity issues related to AI-supported writing, etc. in focus.

KEYWORDS

generative pre-trained transformers (GPT), ChatGPT, artificial intelligence (AI), AI chatbots, natural language processing (NLP), large language model (LLM)

Citation: Tikhonova E., & Raitskaya L. (2023). ChatGPT: Where Is a Silver Lining? Exploring the realm of GPT and large language models. *Journal of Language and Education, 9*(3), 5-11. https://doi.org/10.17323/jle.2023.18119

Correspondence:
Elena Tikhonova,
etihonova@hse.ru

Received: September 01, 2023
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

The world witnesses that AI-generated writings are spreading across various fields. AI assistants have been used for the recent years across education and science. The most popular and efficient AI tools encompass Grammarly[1],

[1] Grammarly. https://www.grammarly.com

Jasper AI[2], JenniAI[3], Hemingway Editor[4], QdouillBot [5], and others. Their use improves writing, checks for errors, corrects spelling. Some of them assist with citations (e.g. QuillBot), others check for plagiarism (e.g. Grammarly). All AI tools of the kind do not produce text. Their usefulness is obvious. In November 2022, an advanced technology of generative artificial intelligence was launched by OpenAI, an American California-based laboratory, showing great performance and spreading at lightning speed. ChatGPT reached one million users only within five days as compared with 300 days for Facebook, 75 days for Instagram to reach the same audience (Firat, 2023, p. 58).

The popularity of ChatGPT can be easily explained. Writing forms an integral and important part of education and work elsewhere. Educational systems of assessment are essentially based on writing. Professional requirements widely imply good-to-perfect writing skills and skills of writing communication. For instance, in the USA, 872 occupations relate to writing skills (Steele, 2023). Authors, journalists, and researchers are frontrunners in writing. They ought to possess most elaborate writing skills. Not surprisingly, the spheres they are engaged in are likely to be influenced most by rapidly developing large language model (LLM) chatbots. The recent ubiquity of the advanced AI technologies replicating human language patterns has led to a discussion of their pros and cons. The challenges, or rather threats and advantages, are considered to have some potential implications for education and various professional fields. The perceptions of the brand-new technologies range from negative or even alarming to positive and enthusiastic.

Even before the arrival of ChatGPT 4.0, its previous version was successfully applied in medical education and practice. ChatGPT is good at "interpreting clinical information" (Ho, Koussayer, & Sujka, 2023), giving full and correct answers to all questions that students of medicine may get at an examination, diagnosing complicated cases, and consulting on treatments. In the same vein, journals on medicine and nursing became the frontrunners who introduced a new stand on ChatGPT's participation in writing research.

Some of medical journals stick to an editorial policy allowing AI-generated text incorporation but subject to a statement of the way ChatGPT was used. Authors are required to indicate where and how this technology was applied. The sections of the submissions covering this information may vary, but most sources single out the methods section or the acknowledgements section as the most appropriate. But all agree that it may be any section but for the information on the authors.

Some researchers assume that artificial intelligence chatbot may pose a threat to the very pillars of education, including assessment of students' educational outcomes (Rudolph et al., 2023), accuracy and credibility of information, skills devaluation (Steel, 2023). The technologies may bring ethical threats and academic integrity concerns, wider exposure to misinformation and fake news in the media (Tewari et al., 2021). Different malevolent uses are likely to influence other human activities (Alasadi, & Baiz, 2023; Fyfe, 2023; Firat, 2023; Illia et al., 2023; Yeo, 2023; Lund, & Wang, 2023).

The JLE editors in their review aim to consider the scope of the emerging field and outline some implications of the technology for scholarly publishing and education as well as the most essential directions of research.

Keywords Explained

Generative pre-trained transformer (GPT) is a large language model serving as a framework for generative artificial intelligence. Such transformers are pre-trained on big sets of text. They generate human-like texts.

ChatGPT is an AI-powered language model developed by OpenAI (Los Angeles, California). On November 30 2023, Open AI launched ChatGPT that opened new opportunities for text production. At present, ChatGPT3.5 and ChatGPT4 (or ChatGPT Plus) are available on the market. The former was freely released as a research preview. ChatGPT4 is distributed to paid subscribers.

Artificial intelligence (AI) may be defined as the intelligence of software, mainly high-profile applications such as advanced web search engines, natural language understanding, generative tools, recommendation platforms, driverless cars, strategic games. AI became an academic field and discipline in 1956.

AI chatbots represent a software application initially called chatterbots, mimicking human conversation. AI chatbots are based on text or voice interactions.

Natural language processing (NLP) is a subfield of computer science and linguistics. It aims to enable computers to understand and generate human languages based on natural language datasets in the form of corpora (both text and speech).

Open AI is an American artificial intelligence laboratory founded in 2015. In 2020, Open AI presented GPT3, a large language model trained on big datasets. In late 2022, GPT3.5 chatbot was launched. In March 2023, GPT4 entered the market.

[2] Jasper AI. (https://www.jasper.ai/)

[3] JenniAI. (https://jenni.ai/)

[4] Hemingway Editor. (https://hemingwayapp.com)

[5] QdouillBot. (https://quillbot.com)

Large language model (LLM) is a model of natural language processing that uses big data to be trained to generate human texts. The model is based on billions of parameters that help it to mimic human languages.

International Research on ChatGPT

We searched the Scopus database for the key word "ChatGPT" and found 1,935 indexed documents on September 24, 2023. Almost all publications (n=1,925) came out in 2023. The prevailing publication type is the "article" (n=827). There are quite many letters (n=350), editorials (n=204), and notes (n=164). Conference papers account for 211 papers. 123 reviews were published during 2023. The most prolific authors include A.Kleebayoon (n=35), V.Wiwanitkit (n=29), and P.P.Ray (n=22). The most highly cited publication in the area is headlined "ChatGPT is fun, but not an author" and has 233 citations as of September 24, 2023 (Thorp, 2023).

Most publications came from the USA (n=611), India (n=192), the UK (n=161), and China (n=154) (see Fig.2). Medicine (n=797), Computer Science (n=493), and Social Sciences (n=472) top the breakdown by subject area (Figure1).

To analyse the speed at which the field had been rising, we compared the readings of the above search and the one made as of April 1, 2023 (Liu et al., 2023). The latter brought 194 papers mentioning ChatGPT on arXiv. A search on the keyword "ChatGPT" identified 186 articles in the PubMed as of April 3, 2023, as compared with only 36 publications on February 23 2023 (Misra & Chadwar, 2023). We expect that much more research articles and other publications will add to the field in the near future.

The most cited article in our search dwell upon the issues of authorship (Thorp, 2023). Other popular directions of research cover priorities related to ChatGPT for researchers (Stokel-Walker, 2023), challenges and implications of ChatGPT in research (Qasem, 2023), practice and policy, ChatGPT performance in the US medical licensing examination and its implications for medical education, the quality of writing (articles, abstracts, essays, etc.) by ChatGPT, potential for education (Crompton, & Burke, 2023; Ivakhnenko, & Nikolskiy, 2023; Fuchs, 2023; Kikalishvili, 2023; Su, & Yang, 2023; Rudolph, Tan, & Tan, 2023), ethical challenges for publishing, ChatGPT and assessment in education (Rudolph, & Tan, 2023), AI-based bot impact on libraries (Lund, & Wang, 2023), ChatGPT in journalism, the future of education, a new academic reality (Lund et al., 2023), academic integrity (Perkins, 2023), science communication (Schäfer, 2023), etc.

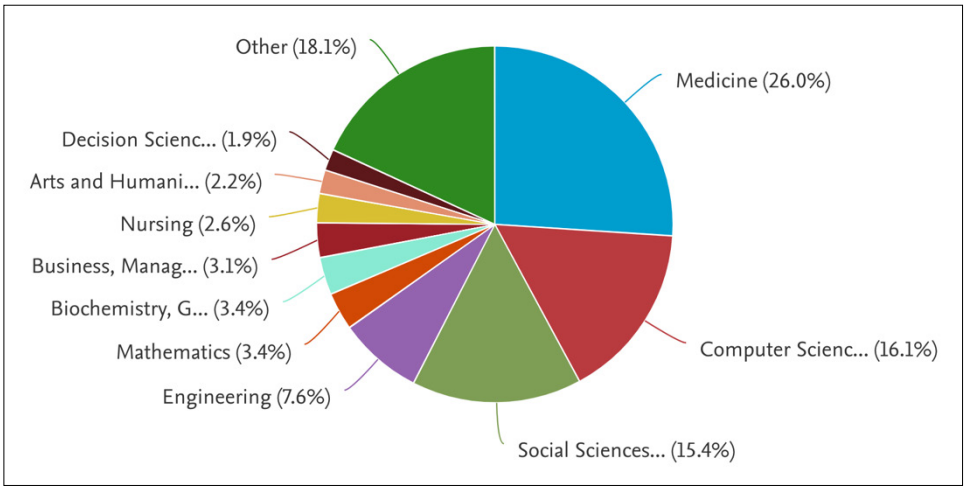
ChatGPT in Education

ChatGPT makes educators, teachers, faculty, professors, and lectures revise traditional educational practices. Assessment is an essential part of education at all educational levels as it gives feedback and outlines the educational outcomes. Traditionally, writing is a predominant way of evaluation. As it takes time, it is often applied out of class (writing essays, reports, and other tasks). An easy access to ChatGPT and similar technologies may tempt students to outsource their tasks to AI tools (Steele, 2023). Two important educational failures are likely to follow: wrong assessment and devaluation of skills.

GPT bots may lure students into accumulating the information they need, ignoring other more reliable sources. Reliability can be easily sacrificed in favour of availability of information via ChatGPT. Educators will have to work out tasks based on critical thinking for students to evaluate any i nformation they get. Traditional writing forms of assessment may be limited to digital-free class (Kikalishvili, 2023).

Figure 1

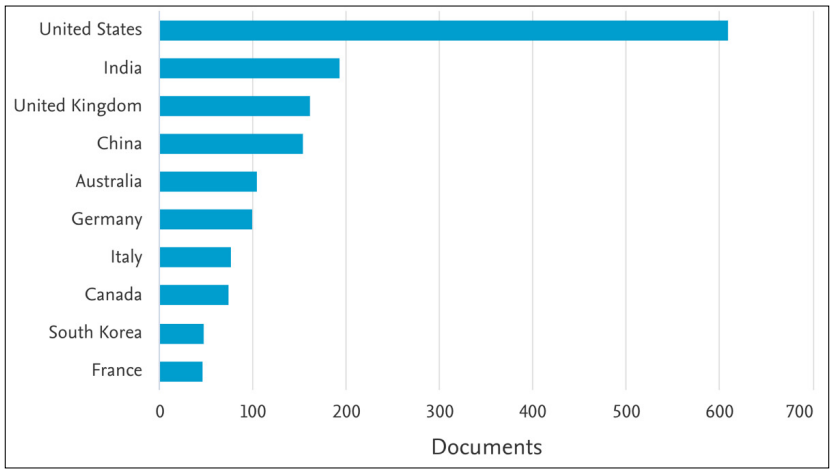
Scopus-Indexed Research on ChatGPT: Breakdown by Subject Area



Note. Source: Scopus Database as of September 15 2023.

Figure 2

Scopus-Indexed Research on ChatGPT: Breakdown by Country or Territory



Note. Source: Scopus Database as of September 24 2023.

As previous threats to education (for instance, a calculator some forty years ago), ChatGPT may prove boon. A similar paradox (Steele, 2023) will cause major revisions in the measurements of knowledge and skills. A new AI-supported workplace will require employees with adequate skills. The latter are to be reevaluated to meet the emerging demands. Given mental awareness and critical perception of information, any new technologies may be adapted for educational purposes and turned into supporting tools. Alarming rhetoric of educators is largely caused by prudence and conservativeness of education as a social institution. But Covid-19 pandemic and the related pedagogy of emergency have increased the adaptivity of educational systems. Today, they are more or less prepared for the sweeping changes associated with AI.

JLE board members and editors are looking forward to new research submissions that will shed light on the looming educational landscape where AI plays on educators' and students' side. The research agenda covers "further exploration of the ethical implications of AI for education, the development of strategies to manage privacy concerns, and the investigation of how educational institutions can best prepare for the integration of AI technologies" (Firat, 2023, p.57). The recently published literature reviews and research outline some potential lines of research that we also see as promising and essential for the academic community at large. They embrace potential applications of chatbots in education, including innovative methods and writing assignments, shifting the focus on skills and competencies (Firat, 2023); facilitating personalized learning and consequently academic achievements, engagement, and self-efficacy (Fuchs, 2023); academic integrity issues related to online examinations (Huber et al., 2023; Fyfe, 2023); ChatGPT mediating role in assessment practices (Farazouli et al., 2023), etc.

ChatGPT in Science

The academic community is stirred by consequences of potential uses of ChatGPT-generated content (Tang, 2023). As the texts produced by AI may successfully follow language patterns typical of the academic writing style and mimic research articles, there is a growing concern that unscrupulous researchers may be tempted to generate partially or more extensively texts, using ChatGPT, and deceitfully pass off AI-created texts as their writings. They may incorporate incomplete, inconsistent, or fallible pieces of the LLM-based texts into their submission (Tools such as ChatCPT, 2023).

No doubt, the technologies are advantageous for non-native English-speaking authors or even native speakers as they may avoid weaknesses in their submissions related to the language quality. But can such a text be totally attributed to the researcher? The plagiarism detecting tools can tell the ChatGPT-generated texts from human writings, though ChatGPT-produced texts are considered as original or newly produced. Special tools detecting AI-generated content are already available with more work in progress (Misra & Chandwar, 2023). The way AI presence is found is connected to regular patterns and algorithms any AI-generated text is based on. Researchers may select to play around with the ChatGPT's help throughout the writing, or only in some chunks of the article.

Authorship of such texts as submissions is raising doubts (Hufton, 2023). Academics and researchers express worries as AI does not bear any responsibility for the produced information (Tang, 2023). In late 2022 and early 2023, several preprints and publications were released, with ChatGPT indicated as a co-author. It led to a heated discussion of AI's authorship. In the wake of ChatGPT launch, Springer Nature was nearly the first to develop new technologies spotting LLM-generated output. The publishers also supported those researchers who disapprove of "citing the bot as an author"

(Tollefson, 2023). The debate is still on the rise on the role of the AI tools in producing scientific literature. LLM tools cannot be accepted as a credited author as "any attribution of authorship is connected to responsibility" (Tools such as ChatCPT, 2023) that sounds senseless if applied to AI.

Many journals are revising their editorial policy regarding their authors' use of AI in their submissions. They tend to disallow crediting ChatGPT or other artificial intelligence language models as a co-author. In early 2023, a few preprints and submissions, mainly in medicine, turned out to contain information on AI authorship[6]. It launched a discussion on the possibility of AI authorship. Consequently, medical journals pioneered the revision of roles of authors and contributors, specifying the disclosure procedure of artificial intelligence-assisted technology in the production of any submission.[7] Elsevier was among those publishers who pioneered new policies related to AI-assisted tools. Elsevier in their journals expects their authors to make a statement on the use of AI-assisted tools. In other publishing houses or journals, researchers should seek permission from their publisher or editor in case they use AI in any part of their submission (but for the information on the authors that is generally prohibited) or specify the sections where they used AI.

Elsevier's Practices

In February 2023, updates on the use of artificial intelligence tools in the submissions were introduced into Elsevier's authorship policy (Hufton, 2023). According to Elsevier policies and guidelines, authors, editors, and reviewers are to follow Publishing Ethics where the use of generative AI and AI-assisted technologies (ChatGPT, NovelAI, Jasper AI, Rytr AI, DALL-E, etc.) in scientific writing, in the journal peer-review and editorial process is described[8].

For Elsevier authors: The policy regarding AI-based technologies exclusively refers to the writing process barring the research process. Authors may improve readability and language of their submission without reservations. General oversight and editing are the author's responsibility. If AI is applied, the author is to make a statement. "Authors should not list AI and AI-assisted technologies as an author or co-author, nor cite AI as an author." [9]

For Elsevier editors and reviewers: As any submitted manuscript is confidential, no part of it may be uploaded into a generative AI tool. The latter may infringe the author's confidentiality and data privacy rights. As correspondence with authors contains personal data, editors cannot upload it into a generative AI tool either. Reviewers should not use AI-assisted tools in the scientific review as peer review is based on critical thinking that is missing from such tools. Moreover, generative AI technologies may produce incorrect or biased conclusions.

The academic community is unanimous that any content produced by AI tools should be "screened and edited for accuracy and appropriateness before dissemination" (Misra & Chadwar, 2023). JLE editors cannot but share the stance of Elsevier and other publishers on the AI-related publishing ethics.

CONCLUSION

ChatGPT has been changing the realities in education, academia, media, and communication. At present, it is impossible to foresee the speed, depth, and scope of transformations. New and unexpected implications of ChatGPT may arise soon. It is high time for journals to revise their notions related to authorship, integrity, and use of AI at large in research and scholarly writing. Following this editorial, JLE is planning to include a provision regulating AI-supported and AI-generated writing in the JLE guidelines for authors and reviewers.

As this new emerging field of study is rising fast, JLE editors welcome any initiatives on special issues, new submissions of research articles and reviews on ChatGPT and associated themes.

AUTHORS CONTRIBUTIONS

Elena Tikhonova: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, other contribution.

Lilia Raitskaya: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, other contribution.

[6] King, M.R., & ChatGPT. (2023). A conversation on artificial intelligence, Chatbots, and plagiarism in higher education. Cellular and Molecular Bioengineering, 16, 1–2. https://doi.org/10.1007/s12195-022-00754-8

[7] International Committee of Medical Journal Editors. Defining the Role of Authors and Contributors. Artificial Intelligence-Assisted Technology. https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html (accessed September 19, 2023).

[8] Elsevier. Publishing Ethics. https://beta.elsevier.com/about/policies-and-standards/publishing-ethics?trial=true

[9] Ibid., The use of generative AI and AI-assisted technologies in scientific writing.

REFERENCES

Alasadi, E.A., & Baiz, C.R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, *100*(8), 2965-2971. https://doi.org/10.1021/acs.jchemed.3c00323

Crompton, H., Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1), 22. https://doi.org/10.1186/s41239-023-00392-8

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023): Hello GPT! Goodbye home examination? An exploratory study of AIchatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2023.2241676

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching*, *6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: is ChatGPT a blessing or a curse? *Frontiers in Education*, *8*, 1166682. https://doi.org/10.3389/feduc.2023.1166682

Fyfe, P. (2023). How to cheat on your final paper: Assigning AI for student writing. *AI and Society*, *38*(4), 1395-1405. https://doi.org/10.1007/s00146-022-01397-z

Huber, E., Harris, L., Wright, S., White, A., Raduescu, C., Zeivots, S., Cram, A. & Brodzeli, A. (2023). Towards a framework for designing and evaluating online assessments in business education. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2023.2183487

Hufton, A.L. (2023). No artificial intelligence authors, for now. *Patterns*, *4*, 14, 2023. https://doi.org/10.1016/j.patter.2023.100731

Ho, W.L.J., Koussayer, B. & Sujka, J. (2023). ChatGPT: Friend or foe in medical writing? An example of how ChatGPT can be utilized in writing case reports. *Surgery in Practice and Science*, *14*, 100185. https://doi.org/10.1016/j.sipas.2023.100185

Illia, L., Colleoni, E., & Zyglidopoulos, S. (2023). Ethical implications of text generation in the age of artificial intelligence. *Business Ethics, Environment and Responsibility*, *32*(1), 201-210. https://doi.org/10.1111/beer.12479

Ivakhnenko, E. N., Nikolskiy, V. S. (2023). ChatGPT in higher education and science: A threat or a valuable resource? *Vysshee obrazovanie v Rossii = Higher Education in Russia*, *32*(4), 9-22. https://doi.org/10.31992/0869-3617-2023-32-4-9-22

Kikalishvili, S. (2023). Unlocking the potential of GPT-3 in education: Opportunities, limitations, and recommendations for effective integration. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2023.2220401

King, M. R., & ChatGPT. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering*, *16*, 1–2. https://doi.org/10.1007/s12195-022-00754-8

Lund, B.D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*, *40*(3), 26-29. https://doi.org/10.1108/LHTN-01-2023-0009

Lund, B.D., Wang, T., Mannuru, N.R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, *74*(5), 570-581. https://doi.org/10.1002/asi.24750

Misra, D.P., & Chandwar, K. (2023). ChatGPT, artificial intelligence and scientific writing: What authors, peer reviewers and editors should know? *Journal of the Royal College of Physicians of Edinburgh*, 1-4. http://doi.org/10.1177/14782715231181023

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, *20*(2), 7. https://doi.org/10.53761/1.20.02.07

Qasem, F. (2023). ChatGPT in scientific and academic research: Future fears and reassurances. *Library Hi Tech News*, *40*(3), 30-32. https://doi.org/10.1108/LHTN-03-2023-0043

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, *6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, *6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Schäfer, M.S. (2023). The Notorious GPT: Science communication in the age of artificial intelligence. *Journal of Science Communication*, *22*(2), Y02. https://doi.org/10.22323/2.22020402

Steele, J.L. (2023). To GPT or not GPT? Empowering our students to learn with AI. *Computers and Education: Artificial Intelligence*, *5*, 100160. https://doi.org/10.1016/j.caeai.2023.100160

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*, 613 (7945), 620-621. https://doi.org/10.1038/d41586-023-00107-z

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, *6*(3), 355-366. https://doi.org/10.1177/20965311231168423

Tewari, S., Zabounidis, R., Kothari, A., Bailey, R., & Alm, C.O. (2021). Perceptions of human and machine-generated articles. *Digital Threats: Research and Practice*, *2*(2), 12. https://doi.org/10.1145/3428158

Tang, G. (2023). Academic journals cannot simply require authors to declare that they used ChatGPT. *Irish Journal of Medical Science* (1971 -). https://doi.org/10.1007/s11845-023-03374-x

Thorp, H.H. (2023). ChatGPT is fun, but not an author. *Science*, *379*(6630), 313. http://doi.org/10.1126/science.adg7879

Tollefson, J. (2023). The plan to "Trump-proof" US science against meddling. *Nature*, *613*(7945), 621-622. http://10.1038/d41586-022-03307-1

Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. (2023). *Nature*, *613*(7945), 612. https://doi.org/10.1038/d41586-023-00191

Yeo, M.A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal*, *14*(3), e716. https://doi.org/10.1002/tesj.716

https://doi.org/10.17323/jle.2023.17302

The Learning Potential of a TV Series in Promoting L2 Incidental Learning of Idiomatic and Non-Idiomatic Phrasal Verbs

Hassan A. Alshumrani

King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

Background: The bulk of past studies, which have shown that audiovisual materials are potential sources for phrasal verb learning, have focused on short materials. However, the incidental learning of idiomatic and non-idiomatic phrasal verbs through extensive viewing of a complete season of a TV series has remained underexplored.

Purpose: The main aim of the present study is to explore the learning potential of viewing an entire season of a TV series in incidental learning of idiomatic and non-idiomatic phrasal verbs.

Method: The study recruited 75 second language (L2) learners and placed them randomly into an experimental group and a control group. Data were gathered through the updated vocabulary levels test and two vocabulary tests. Over one month, the experimental group viewed an entire season of a TV series, and the control group followed their regular learning routine. Immediately after the end of the eight viewing sessions, the experimental and control groups completed the form and meaning post-tests.

Results: The results indicated that viewing a TV series contributed to the development of phrasal verb knowledge. The experimental group developed both form and meaning knowledge of the target phrasal verbs, and greater gains were made at the form recognition level. The research also revealed that although repetition of the target phrasal verbs in the series significantly correlated with the learning gains reported from both the form and meaning tests, its role in meaning was greater.

Implications: The study provides further valuable insights into how watching a TV series affects the learning of idiomatic and non-idiomatic phrasal verbs. It also advances our understanding of how repetition impacts phrasal verb uptake.

KEYWORDS

audiovisual input, incidental vocabulary learning, phrasal verbs, second language learning, vocabulary knowledge aspects

Citation: Alshumrani H. (2023). The Learning Potential of a TV Series in Promoting L2 Incidental Learning of Idiomatic and Non-Idiomatic Phrasal Verbs. *Journal of Language and Education, 9*(3), 12-23. https://doi.org/10.17323/jle.2023.17302

Correspondence:
Hassan A. Alshumrani,
halshumrani1@kau.edu.sa

Received: May 31, 2023
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

Vocabulary plays a crucial role in L2 development, even more so than in first language development (Schmitt, 2010). However, L2 vocabulary learning is a daunting undertaking (Dang et al., 2022) because L2 learners must master thousands of words to reasonably understand written and spoken discourse (Nation, 2006). Yet, intentional instruction is often not practical for L2 vocabulary learning due to limited classroom time and the sheer number of words that requires covering (Pujadas & Muñoz, 2019). Therefore, vocabulary researchers and scholars have strived to identify other ways to encourage L2 learners to acquire vocabulary incidentally. Over the past decade, the accessibility of streaming services has greatly increased, creating ample chances for L2 learners to boost vocabulary learning. This and the increasing popularity among young people of viewing audiovisual materials (Puimège & Peters, 2019) have prompted vocabulary researchers to study the effect of audiovisual resources for incidental vocabulary development. In this respect, a large body of research has established that audiovisual materials, which combine aural and visual input, are beneficial

for word uptake (Pujadas & Muñoz, 2023) to their activation of both the verbal and visual memory processing channels (Mayer, 2005).

Although there is a growing interest in examining word learning through audiovisual resources, the majority of earlier research has focused almost exclusively on individual lexical items (e.g., Cekic, 2022; Feng & Webb, 2020; Fievez et al., 2021; Hsieh, 2020; Peters & Webb, 2018; Teng, 2022). Relatively little research has looked at the learning of multiword items incidentally through audiovisual materials (e.g., Majuddin et al., 2021; Pattemore & Muñoz, 2023; Puimège, et al., 2021; Puimège & Peters, 2019; 2020). These few studies have shown that incidental learning of various types of multiword items is possible through viewing audiovisual materials. It has also been reported that learning is moderated by a range of variables (e.g., cognateness word relevance, and frequency of repetition).

Nonetheless, phrasal verbs (a category of multiword items), which are said to be highly ubiquitous in both spoken and written registers (Garnier & Schmitt, 2016), are considered to be important linguistic element of achieving native-like fluency (Gardner & Davies, 2007). Despite their widespread presence in the English language, they are perceived as being a major source of difficulty and bewilderment for many L2 learners (Yasuda, 2010), who tend to avoid using them in their production (AbdulRahman & Abid, 2014; Liao & Fukuya, 2004). Phrasal verb learning through audiovisual materials has been examined in some studies (e.g., Kim, 2019; Spring, 2019). These studies have suggested that visual representations were effective for "making connections between metaphorical extensions of particle meanings", which in turn, can aid the teaching and learning of phrasal verbs (Spring, 2019, p. 108). While these studies have contributed to our knowledge regarding the positive impacts of audiovisual materials on the learning of phrasal verbs, they were carried out under intentional learning conditions. However, the effect of audiovisual input on phrasal verb learning through incidental learning conditions is still an underexplored topic.

A TV series can serve as a potential source of authentic L2 input to allow L2 learners to continue their learning outside classrooms over an extended period of time (Kobayashi, 2017). However, there is a paucity of studies that have examined the effectiveness of viewing a complete season of a TV series on the incidental learning of phrasal verbs. There is reason to believe that this type of viewing has the potential to promote vocabulary knowledge of the most frequent words, as well as the less frequent ones (Rodgers, 2018; Webb, 2015). Since viewing a whole season of a TV series has proved useful for fostering the incidental learning of single lexical items (Fievez et al., 2021), it is worth examining whether its effects can extend to the incidental learning of phrasal verbs. In addition, given that repetition through viewing a TV series could contribute to the incidental learning of single-word and various multiword vocabulary items, it is plausible that a similar effect would be observed for phrasal verbs. However, no previous studies have shown this specifically. Accordingly, this study bridges the identified literature lacunas by exploring the potential of viewing a TV series on the incidental learning of phrasal verbs. Additionally, it examines the effect of repetition on phrasal verb uptake, and addresses the below research questions:

(1) Does viewing an entire season of a television series result in the incidental learning of idiomatic and non-idiomatic phrasal verbs?

(2) If so, how is the incidental learning of phrasal verbs through a television series affected by the frequency of repetition?

THEORETICAL BACKGROUND

Phrasal verbs, the most prevalent type of multiword items (Liu & Myers, 2020), have attracted considerable research attention, but they have also been at the centre of unabated definitory and classification controversies for a long time. Different studies have adopted different definitions of phrasal verbs. However, for the purpose of this study, a phrasal verb is defined as "a structure that consists of a verb proper and a morphologically invariable particle that functions as a single unit lexically and syntactically" (Liao & Fukuya, 2004, p. 73). Phrasal verbs are regarded as a vital linguistic element (Gardner & Davies, 2007; Schmitt & Redwood, 2011), and therefore failing to use them correctly could lead to unnatural and non-idiomatic English usage (Garnier & Schmitt, 2015). There is an abundance of evidence for their close relationship with language proficiency and fluent language use (Crowley et al., 2023).

Phrasal verbs are ever-present in our everyday language (Chen, 2013), and according to some estimates, in every 150 English words that L2 learners may meet in an English text, there is at least one phrasal verb (Gardner & Davies, 2007). The ratio of phrasal verbs regularly increases because of the constant emergence of new phrasal verbs in use (e.g., chill out) (Schmitt & Redwood, 2011). In the field of L2 research, it is well established that the mastery of phrasal verbs enables L2 learners to produce native-like discourse (Garnier & Schmitt, 2016). Yet, their peculiar syntax and semantic complexity make them a major challenge for the majority of L2 learners, especially those with different first languages to English (Haugh & Takeuchi, 2022; Kovács, 2011).

Research dealing with phrasal verbs has classified phrasal verbs in different ways (Liao & Fukuya, 2004). While some researchers have focused on the syntactic properties of phrasal verbs, the degree of the association between the verb proper and the particle, others have strived to offer insights into the semantic intricacies of phrasal verbs. On the semantic side, Dagut and Laufer (1985) distinguished between

three types of phrasal verbs based on levels of idiomaticity: (1) literal phrasal verbs that have obvious meanings that can be determined from their constituent parts; (2) idiomatic phrasal verbs whose meaning is less obvious and cannot be guessed from the analysis of the meanings of their individual parts; and (3) completive phrasal verbs whose particles indicate the result of the action. By the same token, Ke (2017) divided phrasal verbs into three groups: (1) literal phrasal verbs; (2) semi-transparent phrasal verbs; and (3) idiomatic phrasal verbs. However, Armstrong (2004) argues that the classification of phrasal verbs should be based on the role the particle plays in the meaning of the verb-particle combination, and therefore he has classified phrasal verbs into three types: (1) the directional type, which refers to a group of phrasal verbs whose particles have directional meanings; (2) the aspectual type, which includes a group of phrasal verbs whose particles contribute aspectual meanings to the verb-particle combinations; and (3) the idiomatic type, which consists of phrasal verbs whose verbs and particles are non-transparent.

Moreover, publications that concentrated on phrasal verb learning have adopted two different semantic approaches. The traditional semantic approaches, on the one hand, postulate that the verb-particle formations of phrasal verbs are "arbitrary" and "non-compositional" in the sense that the particle makes no contribution to the meaning of a phrasal verb (Al-Otaibi, 2019; Fraser, 1976). It is thus assumed that the figurative meanings of idiomatic phrasal verbs cannot be explained by the analysis of the meanings of their constituent parts (Al-Otaibi, 2019). Cognitive approaches, on the other hand, hypothesise that figurative phrasal verbs have some degree of compositionality and analyzability and thus "their meanings are not arbitrary but motivated, in the sense that the speakers recognize a few basic relationships between the words in the idioms and their overall figurative interpretations" (Yasuda, 2010, p. 254).

Unlike the traditional approaches which assume that particles make no systematic contribution to the meanings of phrasal verbs, cognitive linguists posit that "particles are orientational metaphors" (Yasuda, 2010, p.252) that carry more importance than the verb constituents in supplying the imagery contents of the meanings of a phrasal verb. In this case, raising learners' awareness of the orientational metaphors in particles would enable them to mentally generalize meanings across unknown phrasal verbs (Boers, 2000). It is argued that the different meanings of phrasal verbs are often due to the fact that their particles have multiple meanings, and therefore focusing on particle meanings when learning phrasal verbs is more effective than learning the entire phrasal verbs as units (Side, 1990; Spring, 2018). Researchers working withing the cognitive semantic paradigm have suggested that the cognitive approaches, such as metaphor awareness (Boers, 2000; Yasuda, 2010) and event conflation (Spring, 2018), facilitate the learning of the figurative meanings of unknown phrasal verbs. However, the effect of this metaphor awareness has been reported to be moderated positively (Boers, 2000) or negatively (Yasuda, 2010) by typological similarities and dissimilarities between learners' first and second languages.

Incidental Phrasal Verb Uptake

The process of acquiring lexical items without intending to do so, while being engaged in activities such as reading, listening, or watching audiovisual materials is called incidental vocabulary learning (Hulstijn, 2001). Learning is regarded as incidental when learners are not instructed to learn word items (e.g., phrasal verbs) or informed in advance about the administration of the subsequent retention test. Following similar previous studies (Dang et al., 2022; Puimège & Peters, 2019; 2020), this study operationalized incidental learning with regards to the absence of the direct instruction to learn and the prior test announcement. It is now well documented that audiovisual input, which is an effective input source (Kim, 2019; Spring, 2019; Webb et al., 2023), is beneficial for incidental word learning (see, Montero Perez's, 2022 for a review of research on the effectiveness of audiovisual input for L2 learning). In this regard, Rodgers (2016) argues that audiovisual materials have a potential for vocabulary development as they fulfil the following conditions of suitable input suggested by Nation (2007): audiovisual materials exist in large quantities; they are engaging and comprehensible; they provide contextual cues that help language learners to pick up the language; and they are familiar to learners. The advantage of audiovisual input centres on the fact that it incorporates auditory and visual input modes. It stands to argue that the use of imagery in audiovisual input is useful for establishing direct cognitive links between pictorial and auditory information (Mayer, 2005), thus aiding viewers to infer the meanings of novel lexical items (Rodgers, 2018).

The Effects of Audiovisual Resources

So far, most published studies that has investigated the impact of viewing audiovisual materials has focused on single vocabulary item learning. For example, Peters and Webb (2018) run two experiments to look at the learning of 64 single lexical items through a one-hour TV documentary program. The researchers reported that their participants were able to incidentally pick up about four new words from the audiovisual material. One of the few studies that has utilized extensive viewing of audiovisual input is Pujadas and Muñoz (2019); they used a TV series to examine the incidental learning of 120 items by young secondary school learners. The findings showed that the learners picked up some words incidentally after extensively viewing the TV series. Further support for the positive impact of extensive viewing of audiovisual materials on the incidental learning of individual lexical items has been provided by Fievez et al. (2021). In their experiment, the researchers have investigated how a set of 78 individual words were learned by watching a complete season of a French series, using form and meaning

recall tests. Their findings indicated that the form of around 14 words and the meaning of 13 words were recalled, after watching the entire season.

The incidental learning of multiword items by viewing audiovisual input has been far less studied. One of the few studies was that by Puimège and Peters (2019), which explored the potential benefits of viewing an excerpt of a TV program on the incidental learning of 20 different kinds of multiword items, including four phrasal verbs, using a range of vocabulary measures. Analysis showed that incidental learning of the multiword items occurred. Similarly, Majuddin et al. (2021) examined L2 students' learning of 20 multiword items (seven phrasal verbs) by viewing one episode of a TV series. Significant gains were reported, indicating that viewing an audiovisual program had positively impacted the learning of diverse kinds of multiword items, including phrasal verbs.

In sum, the foregoing literature review points to the potential effects of viewing different kinds of audiovisual resources for the learning of different kinds of multiword items. However, none of the former studies investigated exclusively the incidental learning of the phrasal verbs. This is surprising given the significance of phrasal verbs in developing nativelike English skills. In addition, all the above-cited studies have used either an excerpt of a TV program or only one episode of a TV program. To date, there has been no research examining the impact of watching a whole season of a television series on incidental learning of phrasal verbs. Further research on the effects of viewing a whole season of a TV series for the development of phrasal verbs is, therefore, needed.

The Role of Repetition

In the field of L2 research, it is now documented that frequency of repetition impacts incidental word learning. Provided that learners encounter the target items frequently in a meaning-focused input context, learning can take place (Uchihara et al., 2023; Zhang, 2022). A substantial body of studies have furnished evidence for this assertion, indicating that higher frequencies are linked with greater learning gains via written input (e.g., Pellicer-Sánchez & Schmitt, 2010), spoken input (e.g., Vidal, 2011), and audiovisual materials (e.g., Peters & Webb, 2018). Though viewing research has devoted considerable efforts to explore the impact of repetition on the learning of individual lexical items, only two viewing studies Dang et al. (2022) and Majuddin et al. (2021) to date have studied the effect of repetition on incidental learning of multiword items and both have reported a positive impact of repetition on learning gains. The present study operationalized repetition as repeated meetings with the target phrasal verbs in the chosen TV series. In summary, most former studies have found that repetition has an impact on the target word development via different input modes. However, there is still a need to explore whether this effect could extend to the learning of phrasal verbs through viewing a complete season of a TV series.

METHOD

Participants

90 L2 first-year students at a key Saudi university participated in this study, but 15 were removed from the final data set because of absenteeism, resulting in 75 final participants. They were undertaking a pre-intermediate English course as a compulsory course in the preparatory year. The participants had learned English for almost nine years and their proficiency level could be considered as lower-intermediate based on their performance in the Cambridge English Placement Test (CEPT). The participants' ages ranged from 18 to 21 years old. The participants received extra credits in their English course in exchange for their time. They belonged to three intact classes. Two intact classes formed the experimental group (n = 48) and one intact class comprised the control group (n = 27).

Materials

An entire season of the animated series *Jurassic World* was adopted as viewing input for the experimental group. It contained eight episodes. The length of each one was 24 minutes, totalling 192 minutes of viewing. The series was about a group of teenagers who had to band together to survive in an adventure camp when the dinosaurs broke out of captivity. The series was watched in the classroom via the streaming service *Netflix*. To ascertain that the chosen series was watchable regarding the speed and clarity of the speech and that it would sustain the participants' interest, a pilot study with 20 first-year university students who shared similar characteristics with the students in the main study was conducted. The outcomes of the pilot study showed that the chosen series was appropriate. The lexical text analysis performed by Heatley's et al. (2002) RANGE software found that there were 17,024 tokens and 1,259 types in the input[1]. A large majority (92.13%) of the running lexical items in the input were from the most frequent 2000-word families. Puimège and Peters (2020) suggested that about a 91% lexical coverage cut-off point was acceptable for achieving adequate viewing comprehension. Given the learners' results on the CEPT and their prior vocabulary knowledge as tested by the Updated Vocabulary Levels Test (UVLT) of Webb et al. (2017), this percentage (92.13%) was deemed acceptable for comprehension.

[1] Heatley, A., Nation, P., & Coxhead, A. (2002). Range: A program for the analysis of vocabulary in texts. http://www.vuw.ac.nz/lals/staff/paul-nation/ nation.aspx.

Target Phrasal Verbs

The target phrasal verbs were selected on the basis of: 1) the unlikelihood of them being already known by the target subjects; 2) the unlikelihood of encountering them outside the experimental; and 3) their frequency of repetition in the selected audiovisual series. The selection of the target phrasal verbs contained several steps. Firstly, an initial list of possible candidates (45 phrasal verbs) was compiled. Secondly, the list was piloted with a group of 30 first-year university students who had a similar L2 proficiency level to the actual participants of the study (from the same B1 level course). Following former viewing studies (e.g., Peters & Webb, 2018), items known to 20% of the piloting participants were removed. Thirdly, two experienced L2 instructors at the same university as the main participants were consulted regarding the chances of encountering the chosen phrasal verbs in the concurrent L2 classes and materials. Phrasal verbs that were likely to be encountered outside the experiment were excluded. It was decided not to add any distractor test items to avoid test fatigue. The final list consisted of 21 phrasal verbs and had 3.0 or higher mutual information scores.

Instruments

Vocabulary Size Test

To check whether or not the participants in the experimental and control groups had similar prior vocabulary knowledge, the UVLT was administered. The test, which has five levels, measures the test-taker's receptive knowledge of the most frequent 5000 vocabulary items. There are 30 items in each level. Each item consists of three meanings and six words (three keys and three distractors). The job of the participants is to tick the correct meaning of the relevant words. The participants were given instructions in Arabic on how to attempt the test. This test has been selected as it has been used in similar former viewing research on L2 learners (e.g., Dang et al., 2022).

Vocabulary Tests

Form recognition and meaning recall tests were adopted to assess the students' knowledge of the chosen phrasal verbs. Both tests were deployed as pretests and posttests and they used the same lexical items, but the target items were ordered differently in the two tests to avoid the risk of a testing effect. The participants first attempted the form recognition test and then took the meaning recall test. This sequence minimized the chance of the testees remembering some of the presented forms of the phrasal verbs (Schmitt & Redwood, 2011). To minimize the chance of random guessing, an 'I don't know' option was used in both tests. In the recognition test (see Appendix A, for an example of this test), the correct forms of the phrasal verbs that were heard in the series had to be chosen from among three options (the target item and two distractors). In the meaning recall test, the participants were given the target forms of the phrasal verbs and had to recall the meaning. The testees were told to supply the meanings (e.g., synonyms, L2 explanations, or L1 translations) of the target forms of the phrasal verbs in English or Arabic as they appeared in the series. (see Appendix B, for an example of this test),

The participants' answers on the form recognition test and the meaning recall test were marked dichotomously. Zero was given to an incorrect answer and 'I don't know', and one point was awarded to a correct answer. An experienced Arabic-speaking L2 teacher was given 20 random exam sheets of the meaning recall test to mark. An initial high interrater reliability of 98% was reached. The remaining inconsistencies were also discussed by the two assessors until an agreement was reached.

Procedure

The study involved three main phases that were conducted over a period of six consecutive weeks of the first academic term in the 2022 academic year. In the first phase, which took place over two sessions, prior to signing the consent form, both groups of subjects were informed about the nature of the experiment (without telling them about the true purpose of the study) and they completed the UVLT. In the following session, they undertook the pretests. In the second phase, the experimental group watched the eight episodes of the series. The participants watched two episodes each week. The control group had their normal English classes and did not watch the series. In the last stage (week six), the experimental and control groups completed the two vocabulary post-tests, which lasted for about 30 minutes each, over two sessions. At the end of the experiment, all participants were given a full explanation of the real aims of the study.

RESULTS

Participants' Reported Vocabulary Knowledge from the UVLT

To check whether or not the participants in the experimental and control groups had similar prior vocabulary knowledge, the UVLT was run. Table 1 presents the participants' mean scores and standard deviations on the UVLT. The results of the series independent samples t-tests showed that the scores on the UVLT of the two groups of the subjects were not significantly different ($p > 0.05$ in all cases), which may suggest that the vocabulary levels of the two groups were similar.

Table 1

Mean Scores and (Standard Deviations) on the UVLT of the Two Groups of Participants

Levels of UVLT	Experimental group (n = 48)	Control group (n = 27)
	Mean (SD)	Mean (SD)
1K	29.07 (3.86)	29.20 (2.08)
2K	22.94 (3.81)	22.44 (6.82)
3K	16.19 (7.04)	16.56 (4.90)
4K	13.90 (4.41)	14.00 (6.82)
5K	8.93 (1.5)	8.77 (3.59)
Overall	91.03 (15.12)	90.97 (13.49)

Note. Max of each level = 30, Total Max = 150

Table 2

Descriptive Statistics of the Form Recognition Test

groups	Pre-test		Post-test	
	mean	SD	mean	SD
experimental (n = 48)	.46	.622	9.41	4.57
Control (n = 27)	1.19	1.24	4.11	3.01

Table 3

Two-Way ANOVA Results for Form Recognition Test

Source	*df*	SS	Mean Square	*F*	Sig
Groups	1	205.1	205.1	22.2	.001
Testing Times	1	1352.5	1352.5	146.7	.001
Groups* Testing times	1	218.6	218.6	23.7	.001
Error	146	1345.6	9.21		

Viewing Effects on Phrasal Verb Learning

Table 2, which presents the participants' scores on the form recognition pretest and posttest, shows low scores obtained on the pretest by both the experimental and control groups. This may indicate that they had little knowledge of the form of the chosen phrasal verbs before the experimental. Their mean scores improved on the posttest. On average, the participants' mean scores increased by 8.95 and 2.92 points in the experimental and control groups, respectively. This suggests that the form recognition knowledge of the target phrasal verbs was developed between the two test administration times. As depicted in Table 2, the participants in the two groups performed differently in the form recognition pre-and-post-tests. To examine the effects of testing times (pre and post) and the two groups (experimental and control) on the scores on the form recognition test, a two-way ANOVA was conducted (reported in Table 3). Significant main effects for groups, F (1, 146) = 22.2, p <.001, η^2 =.13 and for testing times, F (1, 146) = 14.0, p < .001, η^2 =. 57 were found, with the former having a medium effect size and the latter having a large effect size. Similarly, a significant interaction with a medium effect size was found between the effects of testing times and groups, F (1, 146) = 23.7, p < .001, η^2 =.14.

The data in Table 4 shows that the participants in both groups learned some of the target lexical items between the two testing times. In the experimental group, the knowledge of the meaning of the phrasal verbs increased by 3.96 points between the two testing times. This could indicate that they learned around four phrasal verbs on average. A two-way ANOVA was run to determine the effects of the test administration times (pre and post) and the two groups (experimental and control) on the scores on the meaning recall test (displayed in Table 5). The ANOVA results revealed that

there was a significant main effect for groups, F (1, 146) = 39.9, p <.001, η^2 =.21. and for test administration times, F (1, 140) = 30.7, p < .001, η^2 =. 17, with large effect sizes. A significant interaction between the testing times and groups, F (1, 146) = 24.3, p < .001, η^2 =.14 was also found, albeit with a medium effect size.

Role of Repetition

The relationship between repetition and phrasal verb incidental learning was assessed using the Spearman's rho correlation (the data violated the assumptions of the parametric correlation). These analyses only included data from the experimental group. Any target phrasal verbs that were known at the pretest by the participants were excluded from the analyses. As shown in Table 6, a moderate positive correlation between repetition and learning gains reported at the form recognition test was found, r (19) = .44, p = .045. A strong positive correlation was also found between repetition and meaning recall, r (19) = .62, p = .003. These results may suggest that though repetition played a positive role in the incidental learning of both form and meaning, its role in meaning was greater.

DISCUSSION

The findings of this study indicate that viewing a TV series facilitates form and meaning uptake of phrasal verbs. On average, after eight viewing seasons, the experimental participants recognized the form of around nine (42%) new phrasal verbs, and they recalled the meaning of around four (19%) of the 21 target phrasal verbs. On the other hand, the control group recognized about three (13%) new forms of the phrasal verbs and recalled the meanings of.23 (1%) of the items. The control group's scores could be attributed to the testing effect, which is assumed to be very common in incidental vocabulary research (Dang et al., 2022). In fact, this echoes the importance of the recruitment of a control group, as it allows a safe attribution of the learning gains to the intervention employed (Webb & Chang, 2015). However, significant differences in the learning rates were observed in the control and experimental groups; in the two posttests, the experimental group gained significantly higher scores. This suggests that viewing a complete season of a TV series is useful for the learning of phrasal verbs. The present outcomes support the findings provided by a large body of studies regarding the positive impact of watching audiovisual resources on incidental word learning (e.g., Fievez et al., 2021; Majuddin

Table 4

Descriptive Statistics of the Meaning Recall Test

groups	Pre-test		Post-test	
	mean	SD	mean	SD
Experimental (n = 48)	1.17	1.48	5.13	3.26
Control (n = 27)	.70	.99	.93	1.38

Table 5

Two-Way ANOVA Results for Meaning Recall Test

Source	df	SS	Mean Square	F	Sig
Groups	1	182.7	182.7	39.9	.001
Testing Times	1	140.6	140.6	30.7	.001
Groups* Testing times	1	111.3	111.3	24.3	.001
Error	146	668.3	4.57		

Table 6

Correlation Analysis between Vocabulary Tests and Repetition

Variables	Frequency of repetition
Form recognition	422*
Meaning recall	.621**

Note. **p<.01, *p<.05.

et al., 2021; Puimège & Peters, 2019; 2020). They also provide empirical evidence for Webb's (2015) recommendation of extensive viewing, which could be a useful complement to formal language instruction, especially in contexts where there is little exposure to aural L2 input.

On the pretest multiple-choice word form recognition test, the mean score of .46 increased to 9.41 after the intervention and the mean score of 1.17 on the pretest meaning recall test rose to 5.13 on the posttest. This translates into learning rates of approximately 42% and 19% for form recognition and meaning recall, respectively, after 192 minutes of viewing. The learning rate of meaning recall is comparable to that of Puimège and Peters (2019), who reported a learning gain of 3.6 items at the meaning recall mastery level. However, this comparison is little problematic due to differences in the number of phrasal verbs and the length and of input adopted in both studies. Unlike the present study, which targeted 21 phrasal verbs through extensive viewing of an audiovisual series, Puimège and Peters' (2019) used a 30-minute clip of a TV program and tested 20 different types of multiword items, only four of which were phrasal verbs. However, the learning uptake on the meaning recall test (about four items) in the present study is smaller than that reported in the follow-up investigation of Puimège and Peters (2020) – on average, their participants learned about seven new target items.

Nonetheless, the vocabulary uptake reported in the present study might underestimate the actual vocabulary gains that could have occurred through the experiment. Firstly, phrasal verbs that occurred only once (e.g., kick out, cut off) in the series were not included as target items. These items might have been learned but not assessed, given that a TV series presents both verbal and visual forms of input, and the chances of acquiring new items increase with one exposure (Rodgers, 2016). Also, the participants may have developed other knowledge dimensions of the target phrasal verbs that have not been tested. For these reasons, it can be reasonably assumed that the approach of incidental word learning through a TV series is beneficial for phrasal verb development.

The findings of the second research question draw attention to the role repetition plays in word learning through audiovisual resources. Specifically, the results demonstrated that repetition had an impact on the target phrasal verb uptake. Better recall and recognition patterns were observed for the items that have been met most often. Since this study is the first to adopt multiple viewing sessions to take the repetition of phrasal verbs into account, it is difficult to compare its results to those of previous research examining multiword items. Dang et al. (2022) and Majuddin et al. (2021) are the only researchers who have studied the effect of repetition on multiword item uptake by viewing audiovisual materials. These studies, however, only examined the effects of repetition in a one-off learning session, and so they reveal little about the role repetition plays in multiword item learning by viewing multiple episodes.

Nonetheless, the present study's findings align with preceding viewing research examining individual word learning, which found that the frequency of repetition positively affected new word learning (Peters & Webb, 2018). Thus, the emerged findings of this study add to the scarce amount of research investigating the effect of repetition. It seems that when L2 learners frequently meet L2 words, they are likely to notice them and acquire them. One interesting result that emerged from this study was that repetition influenced the learning of form recognition and meaning recall differently. In particular, the findings indicated that meaning recall was impacted slightly more by repetition than form recognition. Fewer encounters were needed to develop the knowledge of form. Conversely, the participants needed multiple repetitions to develop knowledge of meaning. These conflicting findings may be attributed to the narrow range of frequency. Incidental vocabulary research has not reported that frequency has an effect when words are close in their frequency (Webb & Chang, 2015). The results seem to suggest that different knowledge dimensions might need different amounts of repetition, which supports the same findings reported in some reading and listening studies (e.g., Pellicer-Sánchez & Schmitt, 2010; Van Zeeland & Schmitt, 2013).

Limitations and Future Directions of Research

Despite the distinct contributions that the present research makes to the literature concerning how watching a complete season of a TV series influences the learning of phrasal verbs, a range of limitations need to be addressed in future studies. Firstly, the number of target phrasal verbs (n = 21) was limited. A larger number of the target items is desirable to better understand the effects of a TV series on phrasal verbs. More research is also needed to investigate how viewing of a long TV series can develop incidental multiword item knowledge. Another limitation pertains to the limited number of the examined word knowledge types. The study only assessed the form and meaning of the phrasal verbs, which may underestimate the extent of the learning that may have occurred. Future research could make use of multiple measures to assess the learning of other essential levels of item knowledge, such as form recall, meaning recognition, and use.

CONCLUSION

This study provides some useful insights into how watching a TV series affects the learning of phrasal verbs. Firstly, the results demonstrate that viewing an entire season of a TV series contributes to incidental learning gains of phrasal verbs. Particularly, the experimental group acquired around nine novel phrasal verbs at the form knowledge mastery level and approximately four items at the level of meaning recall. The research findings support the recommendation of extensive viewing as a potential source for enhancing L2 learners' lexical item learning and contribute addition-

al evidence that assumes that viewing a TV series is beneficial for incidental learning of phrasal verbs. In addition, the study provides additional affirmative evidence for the role of frequency of repetition in phrasal verb learning via a TV series. While the findings show that phrasal verb gains demonstrated in both the form recognition and meaning recall tests correlate significantly with repetition, the role of repetition in developing meaning knowledge is greater.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

AbdulRahman, Z., & Abid, R. (2014). Rarity or non-existence of phrasal verbs in the written discourse of Omani student-teachers of English. *SAGE Open, 4*(4). https://dx.doi.org/ 10.1177/2158244014556988

Al-Otaibi, G. M. (2019). A cognitive approach to the instruction of phrasal verbs: Rudzak-Ostyen's Model. *Journal of Language and Education, 5*(2), 4–15. https://doi.org/10.17323/jle.2019.8170

Armstrong, K. (2004). Sexing up the dossier: A semantic analysis of phrasal verbs for language teachers. *Language Awareness, 13*(4), 213–224. https://doi.org/10.1080/09658410408668808

Boers, F. (2000). Metaphor awareness and vocabulary retention. *Applied Linguistics, 21*(4), 553–571. https://doi.org/10.1093/applin/21.4.553

Cekic, A. (2022). Incidental L2 vocabulary learning from audiovisual input: The effects of different types of glosses. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2062004

Chen, M. (2013). Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics, 18*(3), 418–442. https://doi.org/10.1075/ijcl.18.3.07che

Crowley, K., Haugh, S., & Spring, R. (2023). An examination of correlations between multiword expression interpretability and general proficiency test scores. *APU Journal of Language Research, 8*, 46–61. https://doi.org/10.34409/apujlr.8.1_47

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition, 7*, 73–79. https://doi.org/10.1017/S0272263100005167

Dang, Y., Lu, C., & Webb, S. (2022). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition, 1*(29), 708–736. https://doi.org/10.1017/ S0272263121000474

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing. *Studies in Second Language Acquisition, 42*(3), 499–523. https://doi.org/10.1017/S0272263119000494

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2021). Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions. *Computer Assisted Language Learning, 36*(1-2), 26-51. https://doi.org/ 10.1080/09588221.2021.1899244

Fraser, J. B. (1976). *The verb-particle combination in English*. Academic Press.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly, 41*(2), 339–359. https://doi.org/10.1002/j.1545-7249.2007.tb00062.x

Garnier, M., & Schmitt, N. (2015). The PHaVE list: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research, 19*(6), 645–666. https://doi.org/10.1177/1362168814559798

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System, 59*, 29–44. https://doi.org/10.1016/j.system.2016.04.004

Haugh, S., & Takeuchi, O. (2022). Phrasal verb usage and L1 typology: A corpus-based study on Chinese and Japanese EFL learners. *The Journal of Asia TEFL*, *19*(4), 1200–1218. https://doi.org/10.18823/asiatefl.2022.19.4.4.1200

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning, 33*(5-6), 567–589 https://doi.org/10.1080/09588221.2019.1577898.

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press.

Ke, Y. (2017). A bi-axis model for profiling English phrasal verbs for pedagogic purposes. *TESOL Quarterly, 51*(4)*,* 972–984. https://doi.org/10.1002/tesq.345.

Kim, S. (2019). The effects of using images and video clips for vocabulary learning: Single words vs phrasal verbs. *STEM Journal, 20*(2), 19–42. https://doi.org/10.16875/stem.2019.20.2.19

Kobayashi, T., (2017). Five possible advantages of watching TV series over movies for autonomous learners of English. Teaching English through Movies*. ATEM Journal, 22,* 43–57. https://doi.org/10.24499/atem.22.0_43

Kovács, É. (2011). The traditional vs. cognitive approach to English phrasal verbs. *Journal of Linguistics*, *1*(2), 141-160.

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning, 54*(2), 193–226. https://doi.org/10.1111/j.1467-9922.2004.00254.x

Liu, D., & Myers, D. (2020). The most-common phrasal verbs with their key meanings for spoken and academic written English: A corpus analysis. *Language Teaching Research*, *24*(3), 403–424. https://doi.org/10.1177/1362168818798384

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition, 43*(5), 985–1008. https:// doi.org/10.1017/S0272263121000036

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching, 55*(2), 163–192. https://doi.org/10.1017/S0261444821000501

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching, 1*(1), 2–13. https://doi.org/10.2167/illt039.0

Pattemore, A., & Muñoz, C. (2023). The effects of binge-watching and spacing on learning L2 multi-word units from captioned TV series. *The Language Learning Journal*, *51*(4), 401-415. https://doi.org/10.1080/09571736.2023.2211614

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*(1), 31–55. https://doi.org/10.1017/s0272263115000224

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Puimège, E., Montero Perez, M., & Peters, E. (2021). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study. *Second Language Research*, *39*(2), 471-492. https://doi.org/10.1177/02676583211049741

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal, 47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 42*(3), 525–549. https://doi.org/10.1017/S027226311900055X

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal, 47*(4), 479–496. https://dx.doi.org/10.1080/09571736.2019.1616806

Pujadas, G., & Muñoz, C. (2023). Measuring the visual in audio-visual input: The effects of imagery in vocabulary learning through TV viewing. *ITL-International Journal of Applied Linguistics*, *174*(2), 263-290. https://doi.org/10.1075/itl.22019.puj

Rodgers, M. P. (2016). Extensive listening and viewing: The benefits of audiobooks and television. *The European Journal of Applied Linguistics and TEFL, 5*(2), 43–57.

Rodgers, M. P. (2018). Extensive viewing: Extra-Curricular language learning outside the classroom walls. The TESOL *Encyclopedia of English Language Teaching.* Wiley-Blackwell. https://doi.org/10.1002/9781118784235.eelt0932

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* Palgrave Macmillan.

Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs. A corpus-informed study. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 173–209). John Benjamins.

Side, R. (1990). Phrasal verbs: Sorting them out. *ELT Journal, 44*, 144–152. https://doi.org/10.1093/elt/44.2.144

Spring, R. (2018). Teaching phrasal verbs more efficiently: Using corpus studies and cognitive linguistics to create a particle list. *Advances in Language and Literary Studies, 9*(5), 121–135. https://doi.org/10.7575/aiac.alls.v.9n.5p.121

Spring, R. (2019). Using short animations to help teach phrasal verbs from a cognitive typology standpoint: Implications from two years of comparative data. *STEM Journal, 20*(4), 105–122. https://dx.doi.org/10.16875/stem.2019.20.4.105

Teng, F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of. learner-related factors. *System*, *105*, 102736. https://doi.org/10.1016/j.system.2022.102736

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2023). Frequency of exposure influences accentedness and comprehensibility in learners' pronunciation of second language words. *Language Learning*, *73*(1), 84-125. https://doi.org/10.1111/lang.12517

Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609–624. https://doi.org/ 10.1016/j.system.2013.07.012

Vanderplank, R. (2016). 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners?. *Language Teaching, 49*(2), 235–250. https://doi.org/10.1017/S0261444813000207

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). Routledge.

Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667–686. https://doi.org/ 10.1177/1362168814559800

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics, 168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, *56*(2), 161-180. https://doi.org/10.1017/S0261444822000507

Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *TESOL Quarterly*, *44*(2), 250–273. https://doi.org/10.5054/tq.2010.219945

Zhang, P. (2022). How does repetition affect vocabulary learning through listening to the teacher's explicit instruction? The moderating role of listening proficiency and preexisting vocabulary knowledge. *Language Teaching Research.* https:// dx.doi.org/10.1177/13621688221140521

APPENDIX A

FORM RECOGNITION TEST

Choose the correct phrasal verbs you heard in the series. There is only one correct answer in each row. Alternatively, choose "I don't know" if you are not sure.

1.	a) Pile up	b) pile on	c) pile at	d) I don't know
2.	a) Pull down	b) pull out	c) pull into	d) I don't know

APPENDIX B

MEANING RECALL TEST

Write the meanings (definition, synonym, or L1 translation) of the following phrasal verbs as they were presented/shown in the series. Some of these phrasal verbs have more than one meaning, in that case please write as many meanings as you can recall from the series. Tick (√) I don't know if you are not sure.

Phrasal verb	Meaning	I don't know
Seal off		
Let down		

https://doi.org/10.17323/jle.2023.17207

The Role of Goal Orientations and Communication Strategies in Willingness to Communicate in EMI Classrooms

Mu-Hsuan Chou

Wenzao Ursuline University of Languages, Taiwan

ABSTRACT

Background: In English-medium instruction (EMI) classrooms, goal orientations, strategies, and communication play pivotal roles in facilitating effective learning. Achievement goal orientations (AGOs) guide and control learner competence-relevant behavior in academic performance. Communication strategies (CSs) are communication aids for learners to cope with problems or breakdowns while speaking the target language. Strategic competence is an indispensable affective-cognitive factor that promotes learners' willingness to communicate (WTC) in a target language.

Purpose: This study aims to investigate the role of AGOs and CSs in predicting WTC and the effect of English proficiency on AGOs and CSs in EMI classrooms.

Method: An online questionnaire survey regarding the perception of AGOs, CSs, and WTC was conducted with 595 university students taking one EMI course in social science and humanity domains in Taiwan. The items were on a 6-point Likert scale ranging from 'strongly disagree' to 'strongly agree'. Hierarchical multiple regression was adopted to predict WTC in EMI classrooms. One-way between-group MANOVAs were adopted to examine the individual and joint effect of English proficiency on the AGOs and CSs.

Results: The hierarchical multiple regression model showed that task goal orientations and social affective strategies strongly and positively predicted the university students' WTC in the EMI classroom. Performance-avoidance goal orientations and message reduction and alteration strategies were found to negatively predict WTC in EMI settings. Students' English proficiency neither predicted their WTC nor affected their AGOs in the EMI classroom. High-proficiency students adopted accuracy-oriented, fluency-oriented, and negotiation for meaning while speaking strategies more frequently than low-proficiency students.

Conclusion: It is suggested that a supportive and dynamic classroom environment with higher-order learning tasks involving cooperation, reflection, and objective assessment criteria can be incorporated into EMI programs. Besides, instruction in CSs and the use of multimedia teaching aids can facilitate EFL learners' comprehension of subject-specific materials and encourage them to engage more in EMI classrooms.

KEYWORDS

achievement goal orientations, communication strategies, willingness to communicate, English proficiency, EMI

Citation: Chou M.-H. (2023). The Role of Goal Orientations and Communication Strategies in Willingness to Communicate in EMI Classrooms. *Journal of Language and Education, 9*(3), 24-37. https://doi.org/10.17323/jle.2023.17207

Correspondence:
Mu-Hsuan Chou,
mhchou@gmail.com

Received: May 07, 2023
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

The importance of communicative competence has been highlighted, but neglected, in the English curricula for primary and secondary education in Taiwan for several decades. According to the English curriculum, in addition to fostering linguistic knowledge and skills, English should be learned for the purpose

of communication[1]. However, the overemphasis on English reading and writing skills in the high-stakes University Entrance Examination has resulted in limited use of speaking skills and strategies in secondary school. Students' English speaking ability is frequently challenged when teacher-student interaction and oral presentations involving communication in English and subject-specific knowledge are required for English-medium instruction (EMI) courses in universities (Chou, 2018; Tsou & Kao, 2017). This lack of opportunity to speak English in high school and necessary English proficiency, along with other situational factors, such as subject-specific topics, has resulted in student reticence and passivity in EMI classrooms (Chou, 2018; Sang & Hiver, 2021; Shao & Gao, 2016). Communication strategies (CSs) may help students overcome these challenges by providing them with alternative ways, such as repetition, code-switching, and message replacement, to communicate their meaning in English. As a result, to what extent communication strategies predict students' willingness to communicate in EMI classrooms is worth examination.

One important variable that influences student motivation and strategy use is achievement goal orientations, which include the goals of developing academic ability, demonstrating ability, and avoiding the demonstration of lack of ability (Elliot & McGregor, 2001; Midgley et al., 1998; Wolters, 2004). These goal orientations provide a framework for how individuals perceive their own academic ability and how they interpret and react to specific learning tasks; this results in different patterns of cognition, motivation, and behavior. A number of researchers have found strong and positive associations between achievement goal orientations, motivation, and use of general learning strategies in academic learning (Lee et al., 2010; Liem et al., 2008; Miller et al., 2021). While most of the literature in East Asian countries and regions has emphasized (1) course design and material development to increase student participation, (2) instructional approaches and assessment practices to improve student comprehension, and (3) academic achievement for individual EMI programs (Gundsambuu, 2019; Jiang et al., 2019; Joe & Li, 2013; Macaro, 2020; Tsou & Kao, 2017), research on strategic and motivational properties in EMI classrooms in Taiwan remains underinvestigated. By understanding how different types of achievement goals and CSs influence students' willingness to communicate (WTC) in EMI classrooms, teachers can provide students with communication support and create a classroom environment that helps students succeed in EMI courses. Thus, the present study investigates the extent to which the roles of achievement goal orientations and communication strategies predict English as a Foreign Language (EFL) university students' willingness to communicate in EMI classrooms.

LITERATURE REVIEW

Achievement Goal Orientations

Goal orientations generally refer to learners' perceptions of their engagement in learning tasks. Achievement goal orientations (AGOs) are future-focused cognitive representations that guide and control learner behavior in terms of belief in academic abilities, engaging in tasks, and academic performance (Elliot & McGregor, 2001; Midgley et al., 1998). Midgley et al. (1998) classified achievement goals into three types: (1) task or mastery goals (i.e. to develop ability and skill), (2) performance-approach goals (i.e. to demonstrate ability), and (3) performance-avoidance goals (i.e. to avoid demonstrating a lack of ability). It was discovered that students with strong mastery goal orientations motivated themselves to develop competence, and increase their knowledge and skills in academic contexts.

One of the major research topics on AGOs has been their effects on students' academic engagement and performance (e.g., Miller et al., 2021; Noordzij et al., 2021; Wolters, 2004), self-regulated learning (e.g., Lim & Lim, 2020; Zhu & Mok, 2018), self-efficacy (e.g., Huang, 2016; Turner et al., 2021), and strategy use (e.g., Adesope et al., 2015; Liem et al., 2008; Somuncuoglu & Yildirim, 1999). Students with strong task goal orientations were found to participate in high-order, interactive, and reflective learning activities (Lim & Lim, 2020; Zhu & Mok, 2018). They have also been found to use more metacognitive and cognitive strategies for processing information and problem solving (Adesope et al., 2015; Liem et al., 2008; Somuncuoglu & Yildirim, 1999). Nevertheless, the associations between the other two types of AGO (performance-approach and performance-avoidance goal orientations) and academic performance have proven inconclusive in the literature. Performance-avoidance goal orientations have been found to have a positive relation with a superficial level of learning strategies (i.e., surface learning) (Liem et al., 2008) and have a negative impact on strategy use and learning outcomes (Adesope et al., 2015; Somuncuoglu & Yildirim, 1999; Wolters, 2004) and self-efficacy in speaking (Turner et al., 2021). On the other hand, Noordzij et al. (2021), in their meta-analysis of induced achievement goals, found no difference in academic performance, induced performance-approach or performance-avoidance goal orientations.

It is known that AGOs significantly affect students' motivation, behavior, and performance in various educational settings, but there is very limited study on EMI courses in the literature. To date, much work on AGOs has been carried

[1] Ministry of Education (2018). English curriculum of 12-Year Basic Education [十二年國民基本教育課程綱要—英語文領域]. https://reurl.cc/y6ZxXq

out on students taking different academic subjects, such as mathematics, science, and psychology in first (L1) and second language (L2). It is thus worth investigating (1) the relationships among EFL university students' AGOs, communication strategies, and their WTC and (2) the extent to which students' AGOs predict their WTC in EMI classrooms.

Communication Strategies

Communication strategies (CSs), also known as interaction strategies, have been considered verbal and nonverbal communication aids that may be used by learners to cope with problems or breakdowns while speaking the target language. These strategies enable learners to remain active interlocutors in communication. Over the past few decades, many studies have explored the taxonomies of CSs (Cohen & Henry, 2020; Goh & Burns, 2012; Nakatani, 2006). For example, CSs have been categorized into three general types by Goh and Burns (2012): cognitive (e.g., paraphrasing, approximation, and formulaic expressions), metacognitive (e.g., planning, self-monitoring, and self-evaluation), and interactional strategies (e.g., exemplification, confirmation checks, and clarification requests). In his Oral Communication Strategy Inventory (OCSI), Nakatani (2006) classified CSs into eight types: (1) social affective (e.g., control anxiety, encourage oneself to speak English, and risk making mistakes); (2) fluency-oriented (e.g., pay attention to the rhythm and conversational flow, and take time to express oneself); (3) negotiation for meaning while speaking (e.g., make comprehension checks, give examples, or repeat oneself to ensure the listener understands); (4) accuracy-oriented (e.g., pay attention to grammar and expression); (5) message reduction and alteration (e.g., reduce the message and use simple expressions); (6) nonverbal (e.g., make eye contact or use gestures and facial expressions); (7) message abandonment; and (8) attempting to think in English.

A number of researchers have adopted the OCSI to explore the CSs used by second and foreign language learners and possible factors, such as cultural background, L1, and anxiety, that influenced their strategy use (Huang, 2010; Ting et al., 2017; Zhang & Liu, 2013). These researchers also discovered a positive relationship between students with higher English proficiency and their use of CSs. Another two intriguing topics prevailing throughout CS research are the effects of (1) speaking tasks on strategy use (e.g., Chou, 2021; Barkaoui et al., 2013) and (2) strategy instruction on increasing speaking ability and WTC in L2 and foreign language classrooms (e.g., Goh & Burns, 2012; Milliner & Dimoski, 2022; Mirsane & Khabiri, 2016). Researchers have discovered that strategy instruction has a significant and positive impact on students' overall speaking performance and the use of problem-solving, interaction, and communication strategies (Goh & Burns, 2012; Milliner & Dimoski, 2022). Additionally, teaching CSs to EFL learners was found to enhance students' WTC in class (Mirsane & Khabiri, 2016).

Effective communication is a fundamental aspect of the learning process and classroom dynamics. CSs can be adopted to help students convey their thoughts, ideas, and opinions to teachers and peers. To understand the role of CS in EMI classrooms and its effect on students' WTC subject-specific knowledge, there is a need to examine (1) the extent to which CSs predict WTC in EMI courses and (2) whether university students with different English proficiency levels vary in terms of their use of CSs in EMI courses.

Willingness to Communicate (WTC) in English

The concept of WTC is derived from communication in L1, which focuses on a person's trait-like and state-like dispositions while speaking (MacIntyre et al., 1999). MacIntyre et al. (1998, p.547) defined L2 WTC as "a readiness to enter into discourse at a particular time with a specific person or persons, using an L2", regarding it as a strong predictor of communication behavior in L2. MacIntyre et al. (1998) noted that factors influencing L2 WTC involve individual antecedents (e.g., personality), affective influences (e.g., attitude, social situation, communicative competence, and strategic competence), motivation, and situated antecedents (e.g., group atmosphere, desire to communicate with a specific person or under a specific context). Peng and Woodrow's (2010) study discovered that WTC can be divided into two categories: WTC in meaning-focused delivery of content knowledge and WTC in form-focused delivery of linguistic aspects.

Studies on L2 WTC have been increasing sharply for several decades, with interests in the relationships among WTC, perceived communicative competence, motivation, and affective factors such as foreign language anxiety and enjoyment (Dewaele, 2019; Lan et al., 2023; Peng & Woodrow, 2010; Shirvan et al., 2019). These studies consistently found a strong correlation between L2 WTC and perceived communicative competence, which is also an important predictor of L2 WTC. Learners' L2 WTC was found to be affected by a variety of individual and situational variables, including personality, attitude, strategic competence, instructional approach, classroom settings, and emotions (Dewaele, 2019; Li et al., 2022; Mirsane & Khabiri, 2016; Peng, 2020; Shirvan et al., 2019; Vafadar & Foo, 2020). In particular, instruction in speaking strategies has been found to enhance students' WTC in language classrooms. Moreover, Peng (2020) discovered that teachers' interaction strategies, gestures, and effective use of audio/video stimuli increased Chinese EFL students' WTC in class. In addition, recent studies have found that students' task goal orientations have a positive impact on their WTC in language classrooms (Karbakhsh & Safa, 2020; Turner et al., 2021). Several studies have examined the association between task goal orientations and WTC in second and foreign language learning. However, there has been far less research on the effects of individual variables, such as English proficiency, different types of goal orientation, and speaking strategies, on EFL university students' WTC in EMI contexts.

Considering the important role of students' WTC in facilitating knowledge transfer and content understanding in academic courses, investigating the effects of its antecedents, say, AGOs as a type of motivational factor, and CSs, on students' WTC can help teachers understand students' motivation and strategy use, and potentially lead to better student-teacher interaction in EMI classrooms. Thus, the purpose of this paper is to examine the predictive power of individual variables on EFL university students' WTC content knowledge in EMI classrooms. Three individual variables were used to predict WTC: students' English proficiency, the three types of AGO, and six types of CS. Based on the preceding research purpose, the four research questions are as follows:

(1) How well do the three measures of AGOs and six measures of CSs predict university students' WTC in EMI courses? How much WTC variance can be explained?

(2) Which variable best predicts the WTC?

(3) If the possible effect of students' English proficiency is controlled for, is the set of variables (AGOs and CSs) still able to predict WTC?

(4) Do students' overall AGOs and CSs in EMI courses vary with their language proficiency?

METHOD

Participants

The participants were 595 university students (mean age: 21.53 years old; female: 70.1%) taking one EMI course from one university in southern Taiwan. The academic departments offered the EMI courses which were either required or optional, including (1) business-related, (2) politics-related, (3) education-related, (4) design-related, and (5) language- and literature-related courses. The students in this university needed to take an English proficiency test, named the *College Student English Proficiency Test* (CSEPT), organized by the university every year, because the test score serves as a graduation threshold for English and as a placement for students studying English for Academic Purposes. The CSEPT has been developed and validated by the Language Training and Testing Center in Taiwan, and with the results aligned with the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)[2]. The students provided their CSEPT scores from six weeks prior in the questionnaire and the scores were interpreted as CEFR levels according to the CSEPT official website. Of the 595 participants, 121 (20.3%) were placed at CEFR A2, 321 (53.9%) at CEFR B1, and 153 (25.7%) at CEFR B2 levels.

Instrument and Measures

The current study adopted a questionnaire survey comprising three sections: three types of AGO from Midgley et al. (1998), the Oral Communication Strategy Inventory (OCSI) from Nakatani (2006), and WTC in meaning-focused delivery from Peng and Woodrow (2010) (see Appendix 1). There are 18 items of AGOs with six of each type from Midgley et al. (1998). To assess task goal orientations, six items were modified (e.g., "*An important reason I complete tasks in the EMI class is that I like to learn new things*" and "*An important reason I participate in the EMI class is that I want to get a better understanding of the content*"). In the case of the performance-approach orientation scale, six items were modified (e.g., "*I would feel truly good if I were the only one who could answer the teachers' questions in English in the EMI class*" and "*Speaking English better than other students in the EMI class is important to me*"). Similarly, another six items from the performance-avoidance orientation scale were modified (e.g., "*One reason I would not speak English in the EMI class is to avoid looking stupid*" and "*In the EMI class, I speak English so that my teachers don't think my English proficiency is lower than others*"). These items were on a 6-point Likert scale ranging from 'strongly disagree' to 'strongly agree'. The reliability of the questionnaire (Cronbach's α) was .88 for the task goal orientations, .85 for the performance-approach goal orientations, and .89 for the performance-avoidance goal orientations.

As for CSs, 22 items of six types were modified to fit the EMI context, including (1) social affective strategies (e.g., "*In the EMI class, I speak English so that my teachers don't think my English proficiency is lower than others*" and "*In EMI class, I don't mind taking risks even though I might make mistakes*") with the Cronbach's α of .86, (2) accuracy-oriented strategies (e.g., "*In EMI class, I pay attention to grammar and word order during conversation*" and "*In EMI class, I notice myself using an expression that fits a rule I have learned*") with the Cronbach's α of .76, (3) fluency-oriented strategies (e.g., "*In EMI class, I take my time to express what I want to say*" and "*In EMI class, I pay attention to the conversation flow*") with the Cronbach's α of .87, (4) message reduction and alteration strategies (e.g., "*In EMI class, I simplify the message and use simple expressions*" and "*In EMI class, I replace the original message with another message because I feel incapable of communicating my original intent*") with the Cronbach's α of .82, (5) negotiation for meaning while speaking strategies (e.g., "*In EMI class, I make comprehension checks to ensure the listener understands what I am trying to say*" and "*In EMI class, I give examples if the listener doesn't understand what I am saying*") with the Cronbach's α of .76, and (6) nonverbal strategies (e.g., "*In EMI class, I try to make eye-contact when I am talking*" and "*In EMI class, I use gestures and facial expressions if I can't express myself*") with the Cronbach's α of

[2] Language Training and Testing Center (2023). Interpretation of test scores. https://www.lttc.ntu.edu.tw/tw/CSEPT_Test_results

.80. These items were on a 6-point Likert scale ranging from 'almost never' to 'almost always'.

Finally, with regard to the WTC in meaning-focused delivery, five items on a 6-point Likert scale ranging from 'strongly disagree' to 'strongly agree', were modified (e.g., "*I am willing to discuss content knowledge in English in the EMI class*" and "*I am willing to give an oral presentation to the class in the EMI course*"). The Cronbach's α was .88 for this scale.

Data Collection

A group-administered survey involving the distribution of an online questionnaire (using Google Forms) to individual EMI classes was adopted. For practical purposes, the group-administered survey saves time, encourages a higher response rate, and offers immediate clarification (Denscombe, 2021). The researcher first collected all the EMI courses from the university computer system and randomly selected several of them. Next, the researcher approached the course instructors to obtain their permission. Once the instructor agreed, the researcher went to the classroom and explained the research purpose to the students. To ensure the research was ethical, an informed consent form, including the purpose of the research, anonymity, required completion time, and confidentiality, was provided to the participants. In other words, voluntary participation was guaranteed.

Data Analysis

The template of Google Forms allows the researcher to fix the scale within a certain range, say the six-point Likert scale in the present study, thus avoiding the mistakes resulting from entering data manually. However, to ensure data accuracy, the maximum and minimum values, and frequencies of the variables were examined. In addition, normality and multivariate outliers were checked by inspecting the Normal Probability Plot (P-P) of the regression standardized residual, the scatterplot, and the Mahalanobis distance (Pallant, 2020; Tabachnick & Fidell, 2021).

To predict WTC in EMI classrooms, hierarchical multiple regression was adopted using IBM SPSS. The variables were added in a hierarchical manner based on the theoretical ground (Pallant, 2020). The participants' English proficiency was presented as their CEFR level (i.e. the categorical variable) and entered in Block 1. Studies have shown that learners' AGOs influenced their strategy use (Adesope et al., 2015; Liem et al., 2008; Somuncuoglu & Yildirim, 1999) and academic engagement (Miller et al., 2021; Noordzij et al., 2021; Wolters, 2004), so the three types of AGO were entered in Block 2. Finally, researchers have found that learners' strategic competence affected their WTC (Mirsane & Khabiri, 2016; Peng, 2020; Vafadar & Foo, 2020), so the six types of CS were entered in Block 3. In addition to the regression analysis, one-way between-group MANOVAs were adopted to examine the individual and joint effect of the one independent variable (i.e., English proficiency) on the composite (overall) dependent variables of AGO and CS.

RESULTS

Correlations

The participants' task goal orientations were moderately and positively correlated with performance-approach goal orientations, CSs, and WTC (r = .28 to .69, p < .01; Table 1). Similarly, positive and moderate correlations were found among performance-approach goal orientations, CSs, and WTC in EMI courses (r = .20 to .50, p < .01). The six types of CS were found to be strongly and positively correlated with

Table 1

Pearson Correlations, Means and Standard Deviations of the Variables (N = 595)

	1	2	3	4	5	6	7	8	9	10
1. Task goals	–									
2. Performance-approach goals	.47**	–								
3. Performance-avoidance goals	.10*	.50**	–							
4. Social affective strategies	.67**	.35**	.02	–						
5. Accuracy-oriented strategies	.51**	.37**	.16**	.69**	–					
6. Fluency-oriented strategies	.57**	.35**	.06	.75**	.71**	–				
7. Message reduction & alteration	.28**	.20**	.11*	.50**	.42**	.45**	–			
8. Meaning negotiation strategies	.48**	.28**	.08	.66**	.63**	.58**	.58**	–		
9. Nonverbal strategies	.43**	.25**	.03	.62**	.53**	.56**	.57**	.59**	–	
10. WTC	.69**	.34**	–.11*	.64**	.50**	.55**	.22**	.43**	.41**	–
Mean	18.39	26.36	22.82	16.84	12.72	16.59	11.96	10.29	7.33	12.59
SD	3.28	5.34	5.04	3.95	3.22	4.11	2.10	2.46	1.68	2.85

Note. * p < .05; ** p < .01.

each other (r = .42 to .75, p < .01). The participants' WTC was related to CSs (r = .22 to .64, p < .01). Performance-avoidance goal orientations had weak and positive correlations with social affective and message alteration and reduction strategies (r = .11 to .16, p < .05) but a negative correlation with WTC (r = –.11, p < .05). In the case of three types of AGO, some researchers have found that task goals are not related to performance-avoidance goals (Liem et al., 2008; Pekrun et al., 2014; Sins et al., 2008), while others have discovered positive associations among the three types of AGO (Karlen et al., 2019; Lim & Lim, 2020). In the present study, it was discovered that the three types of AGO were positively correlated with each other.

RQs 1–3: Regression Analysis and Predictive Power of the AGO and CS on WTC

Hierarchical multiple regression was adopted to assess the ability of three measures of AGO (i.e., task goals, performance-approach goals, and performance-avoidance goals) and six measures of CSs (social affective, accuracy-oriented, fluency-oriented, message reduction and alteration, negotiation for meaning while speaking, and nonverbal strategies), controlled by the participants' English proficiency to predict their WTC in EMI courses. Preliminary analyses revealed that the tolerance values were between .27 and .93, and the VIF values were between 1.08 and 3.73, suggesting no serious violation of assumptions of multicollinearity. A straight diagonal line from bottom left to top right in the Normal P-P was shown and the standardized residuals in the scatterplot were between 3.3 and –3.3, suggesting no multivariate outliers and no major deviation from normality (Pallant, 2020; Tabachnick & Fidell, 2021). The Mahalanobis distance was 28.35, which was lower than the critical value of 29.59; and the Cook's distance was .04 (cutoff point: less than 1), again suggesting no violation of multivariate normality.

The participants' English proficiency was entered at Step 1, explaining 0.1% of the variance in their WTC. Surprisingly, the students' English proficiency was not a predictor of their WTC in EMI courses. After the entry of three types of AGO at Step 2, the total variance explained by the model was 48.6% (Table 2). The two measures increased an additional 48.4% of the variance in the WTC after controlling for English proficiency, ΔR^2 = .484, ΔF (3, 590) = 185.32, p < .0005. In the second model, the three types of AGO were all significant predictors. Finally, after entering the CSs at Step 3, the total variance explained by the model as a whole was 55.3%, F (10, 584) = 72.34, p < .0005. The CSs increased an additional 6.7% of the variance in the WTC after controlling for English proficiency and AGOs, ΔR^2 = .067, ΔF (6, 584) = 14.68, p < .0005. In the final model, four variables were significant, with task goal orientations showing the highest beta value (β = .434, p < .0005), followed by social affective strategies (β = .278, p < .0005), message reduction and alteration strategies (β = –.125, p = .001), and then performance-avoidance goal orientations (β = –.069, p = .038).

Table 2

Regression Weights of Independent Variables and Model Summary of the Hierarchical Regression

		Block 1			Block 2			Block 3		
	Variables	**Beta**	***t***	***P***	**Beta**	***t***	***p***	**Beta**	***t***	***p***
Independent variables	D_1 English proficiency (CEFR level)	.038	.937	.349	.021	.716	.474	–.021	–.717	.473
	X_1 Task Goals				.660	19.44	.000	.434	10.723	.000
	X_2 Performance-approach Goals				.087	2.22	.027	.044	1.188	.235
	X_3 Performance-avoidance Goals				–.102	–2.93	.004	–.069	–2.084	.038
	X_4 Social Affective Strategies							.278	5.202	.000
	X_5 Accuracy-oriented Strategies							.053	1.204	.229
	X_6 Fluency-oriented Strategies							.074	1.549	.122
	X_7 Message Reduction & Alteration Strategies							–.125	–3.395	.001
	X_8 Negotiation for Meaning Strategies							.007	.170	.865
	X_9 Nonverbal Strategies							.045	1.132	.258
Model summary	R^2	.001			.486			.553		
	F	.887			139.415			72.336		
	p	.349			.000			.000		
	Δ R^2	.001			.484			.067		
	Δ F	.887			185.321			14.683		
	p of Δ F	.349			.000			.000		

RQ4: One-Way MANOVA Results of English Proficiency, Achievement Goal Orientations, and Communication Strategies in EMI courses

A one-way between-groups MANOVA was conducted to investigate English proficiency differences in AGOs in EMI courses. There was no statistically significant difference among the students of three levels of English proficiency on the combined dependent variables, $F = 1.43$, $p = .198$; Wilks' $\lambda = .99$; partial $\eta^2 = .01$ (Table 3). When the dependent variable results were considered separately for the purpose of reducing Type 1 error, the α level was adjusted by dividing the original α level of .05 by the number of factors in each component (Pallant, 2020). In this case, the cut-off point for the significant level of each dependent variable was .017 (.05/3). Thus, there was no statistically significant difference among the students of three levels of English proficiency on the three individual dependent variables.

A one-way between-groups MANOVA was performed to investigate English proficiency differences in CSs (social affective, accuracy-oriented, fluency-oriented, message reduction and alteration, negotiation for meaning while speaking, and nonverbal strategies) in EMI courses. There was a statistically significant difference among English language proficiency on the combined dependent variables, $F = 3.15$, $p < .0005$; Wilks' $\lambda = .94$; partial $\eta^2 = .03$. Again to reduce Type 1 error, when the results of the six dependent variables were considered separately, three differences reached statistical significance using a Bonferroni adjusted α level of .008 (.05/6): accuracy-oriented, fluency-oriented, and negotiation for meaning while speaking strategies (Table 4).

Post hoc comparisons were measured using the Tukey HSD test. The results indicated that in accuracy-oriented strategies, CEFR A2 students ($M = 11.91$, $SD = 3.57$) tended to emphasize the subject and verb of the sentence, pay attention to grammar and word order during conversation, or try to talk like a native speaker less frequently than the CEFR B1 ($M = 12.83$, $SD = 3.14$) and B2 students ($M = 13.14$, $SD = 3.00$). Next, the mean scores in fluency-oriented strategies showed that the CEFR B2 students ($M = 17.84$, $SD = 3.73$) adopted fluency-oriented strategies (e.g., taking time to express what they want to say or paying attention to the conversation flow) more frequently than the CEFR B1 ($M = 16.46$, $SD = 3.98$) and A2 students ($M = 15.36$, $SD = 4.50$). For negotiation for meaning while speaking strategies, the CEFR B2 students ($M = 10.76$, $SD = 2.27$) made comprehension checks to ensure that the listener understood what they wanted to say or gave examples if the listener did not understand what they were saying more frequently than their CEFR A2 peers ($M = 9.80$, $SD = 2.86$).

DISCUSSION

The Predictive Power of AGOs on WTC in EMI Classrooms

The regression model showed that of the three types of AGO, task goal orientation was a stronger predictor of WTC than performance-avoidance goal orientation in the EMI classroom, while performance-approach goal orientation did not predict WTC at all. The students who participated in the EMI class to obtain a better comprehension of the content and to strengthen their professional knowledge were more willing to participate in the discussion of the content knowledge and give presentations in English. Recent studies have shown that students' task goal orientations positively affect their WTC in English (Chou, 2022; Karbakhsh & Safa, 2020; Turner et al., 2021). The findings in the current study corresponded to those of past studies in that students with stronger task goal orientations were more willing to communicate for the purposes of comprehending and producing subject-specific knowledge in English. Studies on the antecedents of task goals showed that classroom context, the evaluation system, students' psychological needs for academic competence, and satisfaction with learning have positive impacts on their task goal orientations (Ames, 1992;

Table 3

One-way MANOVA Results of English Proficiency and Achievement Goal Orientations

Variables	Group	N	M	SD	*F value*	*p value*	η^2
Task Goal Orientations	CEFR A2	121	18.33	3.34	.08	.920	.00
	CEFR B1	321	18.38	3.24			
	CEFR B2	153	18.48	3.34			
Performance-approach Orientations	CEFR A2	121	26.69	5.28	.86	.423	.00
	CEFR B1	321	26.45	5.28			
	CEFR B2	153	25.90	5.52			
Performance-avoidance Orientations	CEFR A2	121	23.82	5.72	3.66	.026	.01
	CEFR B1	321	22.75	4.81			
	CEFR B2	153	22.18	4.85			

Table 4

One-way MANOVA Results of English Proficiency and Communication Strategies

Variables	Group	N	M	SD	*F value*	*p value*	η^2
1. Social Affective Strategies	CEFR A2	121	15.93	4.43	4.45	.012	.02
	CEFR B1	321	16.95	3.87			
	CEFR B2	153	17.31	3.62			
2. Accuracy-oriented Strategies	CEFR A2	121	11.91	3.57	5.42	.005	.02
	CEFR B1	321	12.83	3.14			
	CEFR B2	153	13.14	3.00			
3. Fluency-oriented Strategies	CEFR A2	121	15.36	4.50	13.16	.000	.04
	CEFR B1	321	16.46	3.98			
	CEFR B2	153	17.84	3.73			
4. Message Reduction and Alteration Strategies	CEFR A2	121	11.62	2.45	3.33	.036	.01
	CEFR B1	321	11.94	2.06			
	CEFR B2	153	12.27	1.86			
5. Negotiation for Meaning while Speaking Strategies	CEFR A2	121	9.80	2.86	5.26	.005	.02
	CEFR B1	321	10.26	2.35			
	CEFR B2	153	10.76	2.27			
6. Nonverbal Strategies	CEFR A2	121	7.01	1.82	4.46	.012	.02
	CEFR B1	321	7.32	1.61			
	CEFR B2	153	7.61	1.67			

Hengsadeekul et al., 2014; Karbakhsh & Safa, 2020). As task goal orientations were the strongest predictor of students' WTC in EMI courses, it is suggested that a supportive and dynamic classroom environment with higher-order learning tasks involving cooperation, reflection, and objective assessment criteria can be incorporated into EMI programs.

It is worth noting that although there is no research evidence to support the direct effects of performance-approach and performance-avoidance goal orientations on WTC in either general English or EMI courses, it was found that the students with stronger performance-avoidance goal orientations tended to have less WTC in the EMI classroom. Researchers have found that language learners' reticence and passivity in language classrooms can often be attributed to their fear of negative evaluation, making mistakes, and embarrassment (Sang & Hiver, 2021; Shao & Gao, 2016). This explained why performance-avoidance goal orientations, which refer to learners' avoidance of demonstrating a lack of ability (Elliot & McGregor, 2001; Midgley et al., 1998), negatively predicted WTC in the EMI classroom in this study.

Relationship between AGOs and CSs in EMI Classrooms

The results from previous research have shown that task goal and performance-approach orientations positively impacted learners' use of metacognitive or cognitive strategy and academic achievement (Liem et al., 2008; Lim & Lim, 2020; Miller et al., 2021; Wolters, 2004; Zhu & Mok, 2018). Macaro (2021) also highlights that students' goal orientations and prior knowledge of both language and content are likely to affect their strategic use in EMI classrooms. In the present study, the correlations showed that students with strong task and performance-approach goals employed CSs more frequently in EMI classrooms. While past studies have investigated the positive effects of AGOs on the use of general learning strategies for various academic subjects (Adesope et al., 2015; Liem et al., 2008; Somuncuoglu & Yildirim, 1999), the present findings agree with these studies in that university students with stronger task and performance-approach goal orientations adopted CSs more frequently in EMI classrooms.

Unlike task and performance-approach goal orientations, researchers have discovered that performance-avoidance goal orientations have a negative impact on learning outcomes and strategy use (Adesope et al., 2015; Somuncuoglu & Yildirim, 1999; Wolters, 2004). Contrary to previous research, the results of the current study revealed a positive relationship between performance-avoidance goal orientations and two types of CS (accuracy-oriented and message reduction and alteration strategies). To avoid looking stupid and like they cannot express their thoughts accurately in English, students tend to pay attention to grammar and word order during conversations, simplify their messages,

and use simple expressions while speaking English. Research into the factors that affect AGOs found that tasks with diversity and variety, well-structured and effective teaching in classroom settings, assessment systems that dissuade comparing oneself to peers but instead promote self-reflection and self-improvement, and learning activities that take into account students' psychological needs for competence and satisfaction all had a positive influence on their AGOs (Ames, 1992; Karbakhsh & Safa, 2020). Therefore, it is suggested to introduce a classroom setting that fosters support and interaction, incorporates diverse learning activities, promotes cooperative learning, encourages self-reflection, and establishes transparent assessment standards in EMI classrooms.

The Predictive Power of CSs on WTC in EMI Classrooms

According to MacIntyre et al. (1998), one's strategic competence is an important antecedent of WTC. Of the six types of CS, only two types significantly predict university students' WTC in the EMI classroom, with social affective strategies (e.g., actively encouraging oneself to express opinions or not mind taking risks and making mistakes) being a stronger predictor than message reduction and alteration strategies. Much work has been done to ascertain the influence of CSs on improving speaking ability and WTC in language classrooms (Goh & Burns, 2012; Mirsane & Khabiri, 2016; Vafadar & Foo, 2020). Tai and Tang (2021) investigated the mediating role of anxiety in the relationship between learning strategies and EMI avoidance in postgraduate business programs in Taiwan and discovered that students' use of high-level learning strategies, such as organization, critical thinking, metacognitive self-regulation, effort regulation, and peer learning, reduced their anxiety and encouraged their participation in EMI classes. Furthermore, in a meta-analysis study on the effectiveness of strategy-based instruction on academic performance, Donker et al. (2014) discovered that planning and task value were the most effective strategies, and instructing metacognitive knowledge enhanced the effectiveness of strategies. The current study enhances the findings of these studies by discovering the strong, positive effect of social affective strategies on assisting students in discussing and presenting and on increasing their WTC subject-specific knowledge in EMI classrooms. To enhance students' WTC in EMI classrooms, researchers have suggested that encouraging strategy use and self-regulation skills (Pun & Jin, 2021; Tai & Tang, 2021) and providing multimedia stimuli (e.g., audio/video/picture) help prompt students' WTC in class and maintain students' learning motivation and interest (Mirsane & Khabiri, 2016; Peng, 2020; Vafadar & Foo, 2020). It is thus recommended that instruction in CSs and the use of multimedia teaching aids facilitate EFL learners' comprehension of subject-specific materials and encourage them to engage more in EMI classrooms.

The Predictive Power of English Proficiency on WTC and the Influence of English Proficiency on AGOs and CSs

In language learning studies, researchers have found that students' English proficiency positively predicts their WTC in FL classrooms (Darasawang & Reinders, 2021; Tan & Phairot, 2018). However, in EMI courses where subject-specific knowledge and English are integrated, data from the current study seem to contradict earlier findings. The results showed that the students' English proficiency neither predicted their WTC nor influenced their AGOs in the EMI classroom. Researchers have found that students frequently report language difficulties in understanding subject-specific words and the content of in-class materials, producing academic essays, and participating in discussion in a number of EMI studies (Evans & Morrison, 2011; Jiang et al., 2019; Joe & Li, 2013; Tsou & Kao, 2017). The participants' frequent use of message reduction and alteration strategies in this study supports the fact that expressing ideas regarding subject-specific knowledge in English is a major obstacle in the EMI classroom. In contrast to the students' English proficiency, their task goal orientations (as a type of motivation) and social affective strategies were both positive and strong predictors of their WTC in EMI classrooms.

Studies investigating the relationship between English proficiency and the use of CSs in general English courses have shown that high-proficiency students adopt CSs more frequently than low-proficiency students (Huang, 2010; Ting et al., 2017; Zhang & Liu, 2013). The present study showed that students' overall CSs varied with their English proficiency in their EMI classrooms, and it also enhances the findings of previous studies by showing that CEFR B2 students adopt accuracy-oriented, fluency-oriented, and negotiation for meaning while speaking strategies more frequently than CEFR A2 students in EMI classrooms.

CONCLUSION

This study examined the role of AGOs and CSs in EFL university students' WTC in EMI classrooms. The hierarchical multiple regression model showed that the two strong, positive predictors of the students' WTC were task goal orientations (i.e., goals to motivate oneself to increase one's knowledge and develop competence in academic settings) and social affective strategies. Performance-avoidance goal orientations and message reduction and alteration strategies, on the other hand, negatively predicted WTC. The participants' general English proficiency neither predicted their WTC nor influenced their AGOs but affected the CSs they adopted in the EMI classroom, especially the accuracy-oriented, fluency-oriented, and negotiation for meaning while speaking strategies.

Even though the body of research has offered insights into EMI classrooms, several limitations need to be considered. First, since the participants in these EMI courses majored in social science, arts, and humanities, the results cannot be generalizable to students in the hard sciences. Further research involving science EMI courses would be of great interest and value. Second, the method of investigation is not without problems, since only students' self-reported data from the questionnaires were collected. Additional research using classroom observation would be of great interest and value in understanding student−student and teacher−student communication in EMI classrooms. Finally, only receptive skills were assessed in the English proficiency test, which may limit our interpretation of its predictive power on students' WTC in EMI classrooms.

FUNDING

This work was supported by the National Science and Technology Council: Grant Number MOST 111-2410-H-160-006.

DECLARATION OF COMPETING INTEREST

None declared.

REFERENCES

Adesope, O. O., Zhou, M., & Nesbit, J. C. (2015). Achievement goal orientations and self-reported study strategies as predictors of online studying activities. *Journal of Educational Computing Research, 53*(3), 436–458. https://doi.org/10.1177/0735633115603989

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261–271. https://doi.org/10.1037/0022-0663.84.3.261

Barkaoui, K., Brooks, L., Swain, M., & Lapkin, S. (2013). Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics, 34*(3), 304–324. https://doi.org/10.1093/applin/ams046

Chou, M.-H. (2018). Speaking anxiety and strategy use in learning English as a Foreign Language in full and partial English-medium instruction contexts. *TESOL Quarterly, 52*(3), 611–633. https://doi.org/10.1002/tesq.455

Chou, M.-H. (2021). An integrated approach to developing and assessing EFL students' speaking ability and strategy use. *Language Education & Assessment, 4*(1), 19–37. https://doi.org/10.29140/lea.v4n1.428

Chou, M.-H. (2022). Investigating the mediating role of foreign language anxiety in students' task goal orientations, perceived communication competence, English ability, and willingness to communicate in EMI classrooms. *English Teaching & Learning, 46*(3), 235–254. https://doi.org/10.1007/s42321-022-00116-7

Cohen, A. D., & Henry, A. (2020). Focus on the language learner: Style, strategies and motivation. In N. Schmitt & M. Rodgers (Eds.), *An introduction to applied linguistics* (3rd ed., pp. 165–189). Routledge.

Darasawang, P., & Reinders, H. (2021). Willingness to communicate and second language proficiency: A correlational study. *Education Sciences, 11*(9), 517. https://doi.org/10.3390/educsci11090517

Denscombe, M. (2021). *The good research guide: For small-scale social research projects* (7th ed.). Open University Press.

Dewaele, J. M. (2019). The effect of classroom emotions, attitudes toward English, and teacher behavior on willingness to communicate among English foreign language learners. *Journal of Language and Social Psychology, 38*(4), 523–535. https://doi.org/10.1177/0261927X19864996

Donker, A. S., De Boer, H., Kostons, D., Van Ewijk, C. D., & van der Werf, M. P. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review, 11*, 1–26. https://doi.org/10.1016/j.edurev.2013.11.002

Elliot, A. J., & McGregor, H. A. (2001). A 2 X 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501–519. https://doi.org/10.1037/0022-3514.80.3.501

Evans, S., & Morrison, B. (2011). The student experience of English-medium higher education in Hong Kong. *Language and Education, 25*(2), 147–162. https://doi.org/10.1080/09500782.2011.553287

Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A holistic approach.* Cambridge University Press.

Gundsambuu, S. (2019). Internationalization and English as a medium of instruction in Mongolian higher education: A new concept. *Journal of Language and Education, 5*(2), 48–66. https://doi.org/10.17323/jle.2019.8481

Hengsadeekul, C., Koul, R., & Kaewkuekool, S. (2014). Motivational orientation and preference for English-medium programs in Thailand. *International Journal of Educational Research, 66*, 35–44. https://doi.org/10.1016/j.ijer.2014.02.001

Huang, C. (2016). Achievement goals and self-efficacy: A meta-analysis. *Educational Research Review, 19*, 119–137. https://doi.org/10.1016/j.edurev.2016.07.002

Huang, C.-P. (2010). Exploring factors affecting the use of oral communication strategies. *LongHua Technology University Journal, 30*(1), 85–104.

Jiang, L., Zhang, L. J., & May, S. (2019). Implementing English-medium instruction (EMI) in China: Teachers' practices and perceptions, and students' learning motivation and needs. *International Journal of Bilingual Education and Bilingualism, 22*(2), 107–119. https://doi.org/10.1080/13670050.2016.1231166

Joe, Y., & Lee, H. K. (2013). Does English-medium instruction benefit students in EFL contexts? A case study of medical students in Korea. *The Asia-Pacific Education Researcher, 22*(2), 201–207. https://doi.org/10.1007/s40299-012-0003-7

Karbakhsh, R., & Safa, M. A. (2020). Basic psychological needs satisfaction, goal orientation, willingness to communicate, self-efficacy, and learning strategy use as predictors of second language achievement: A structural equation modeling approach. *Journal of Psycholinguistic Research, 49*(5), 803–822. https://doi.org/10.1007/s10936-020-09714-7

Karlen, Y., Suter, F., Hirt, C., & Merki, K. M. (2019). The role of implicit theories in students' grit, achievement goals, intrinsic and extrinsic motivation, and achievement in the context of a long-term challenging task. *Learning and Individual Differences, 74*, 101757. https://doi.org/10.1016/j.lindif.2019.101757

Lan, G., Zhao, X., & Gong, M. (2023). Motivational intensity and willingness to communicate in L2 learning: A moderated mediation model of enjoyment, boredom, and shyness. *System, 117*, 103116. https://doi.org/10.1016/j.system.2023.103116

Lee, J. Q., McInerney, D. M., Liem, G. A. D., & Ortiga, Y. P. (2010). The relationship between future goals and achievement goal orientations: An intrinsic–extrinsic motivation perspective. *Contemporary Educational Psychology, 35*(4), 264–279. https://doi.org/10.1016/j.cedpsych.2010.04.004

Li, C., Dewaele, J. M., Pawlak, M., & Kruk, M. (2022). Classroom environment and willingness to communicate in English: The mediating role of emotions experienced by university students in China. *Language Teaching Research*. https://doi.org/10.1177/13621688221111623

Liem, A. D., Lau, S., & Nie, Y. (2008). The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome. *Contemporary Educational Psychology, 33*(4), 486–512. https://doi.org/10.1016/j.cedpsych.2007.08.001

Lim, J. Y., & Lim, K. Y. (2020). Co-regulation in collaborative learning: Grounded in achievement goal theory. *International Journal of Educational Research, 103*, 101621. https://doi.org/10.1016/j.ijer.2020.101621

Macaro, E. (2020). Exploring the role of language in English medium instruction. *International Journal of Bilingual Education and Bilingualism, 23*(3), 263–276. https://doi.org/10.1080/13670050.2019.1620678

Macaro, E. (2021). Learner strategies in an English Medium Instruction context. In B. D. Sabato & B. Hughes (Eds.), *Multilingual perspectives from Europe and beyond on language policy and practice* (pp. 63–82). Routledge.

MacIntyre, P. D., Babin, P. A., & Clément, R. (1999). Willingness to communicate: Antecedents & consequences. *Communication Quarterly, 47*(2), 215–229. https://doi.org/10.1080/01463379909370135

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal, 82*(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

Midgley, C., Kaplan, A., Middleton, M., Maehr, M. L., Urdan, T., Anderman, L. H., Anderman, E., & Roeser, R. (1998). The development and validation of scales assessing student' achievement goal orientations. *Contemporary Educational Psychology, 23*(2), 113–131. https://doi.org/10.1006/ceps.1998.0965

Miller, A. L., Fassett, K. T., & Palmer, D. L. (2021). Achievement goal orientation: A predictor of student engagement in higher education. *Motivation and Emotion, 45*(3), 327–344. https://doi.org/10.1007/s11031-021-09881-7

Milliner, B., & Dimoski, B. (2022). The effects of communication strategy training on speaking task performance. *RELC Journal.* https://doi.org/10.1177/00336882221085781

Mirsane, M., & Khabiri, M. (2016). The effect of teaching communicative strategy on EFL learners' willingness to communicate. *Theory and Practice in Language Studies, 6*(2), 399–407. https://doi.org/10.17507/tpls.0602.24

Noordzij, G., Giel, L., & van Mierlo, H. (2021). A meta-analysis of induced achievement goals: the moderating effects of goal standard and goal framing. *Social Psychology of Education, 24*(1), 195–245. https://doi.org/10.1007/s11218-021-09606-1

Nakatani, Y. (2006). Developing an oral communication strategy inventory. *The Modern Language Journal, 90*(2), 151–168. https://doi.org/10.1111/j.1540-4781.2006.00390.x

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th Ed.). Routledge.

Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction, 29*, 115–124. https://doi.org/10.1016/j.learninstruc.2013.09.002

Peng, J. E. (2020). Teacher interaction strategies and situated willingness to communicate. *ELT Journal, 74*(3), 307–317. https://doi.org/10.1093/elt/ccaa012

Peng, J.-E., & Woodrow, L. (2010). Willingness to communicate in English: A model in the Chinese EFL classroom context. *Language Learning, 60*(4), 834–876. https://doi.org/10.1111/j.1467-9922.2010.00576.x

Pun, J., & Jin, X. (2021). Student challenges and learning strategies at Hong Kong EMI universities. *Plos One, 16*(5), e0251564. https://doi.org/10.1371/journal.pone.0251564

Sang, Y., & Hiver, P. (2021). Using a language socialization framework to explore Chinese students' L2 reticence in English language learning. *Linguistics and Education, 61*, 100904. https://doi.org/10.1016/j.linged.2021.100904

Shao, Q., & Gao, X. A. (2016). Reticence and willingness to communicate (WTC) of East Asian language learners. *System, 63*(4), 115–120. https://doi.org/10.1016/j.system.2016.10.001

Shirvan, M. E., Khajavy, G. H., MacIntyre, P. D., & Taherian, T. (2019). A meta-analysis of L2 willingness to communicate and its three high-evidence correlates. *Journal of Psycholinguistic Research, 48*(6), 1241–1267. https://doi.org/10.1007/s10936-019-09656-9

Somuncuoglu, Y., & Yildirim, A. (1999). Relationship between achievement goal orientations and use of learning strategies. *The Journal of Educational Research, 92*(5), 267–277. https://doi.org/10.1080/00220679909597606

Tabachnick, B. G., & Fidell, L. S. (2021). *Using multivariate statistics* (8th Ed.). Pearson.

Takeuchi, O., Griffiths, C., & Coyle, D. (2007). Applying strategies to contexts: The role of individual, situational, and group differences. In A. D. Cohen & Macaro, E. (Eds.), *Language learner strategies* (pp. 69–92). Oxford University Press.

Tai, T. E., & Tang, C. W. (2021). The role of graduate students' learning strategies in reducing their English medium instruction avoidance: The mediation effect of language anxiety. *Asia Pacific Journal of Education, 41*(2), 368–384. https://doi.org/10.1080/02188791.2020.1777937

Tan, K. E., & Phairot, E. (2018). Willingness to communicate among Thai EFL students: Does English proficiency matter? *Journal of Asia TEFL, 15*(3), 590–602. https://doi.org/10.18823/asiatefl.2018.15.3.2.590

Ting, S.-H., Soekarno, M., & Lee, P.-Y. (2017). Communication strategy use and proficiency level of ESL learners. *Journal of Asia TEFL, 14*, 162–170. https://doi.org/10.18823/asiatefl.2017.14.1.11.162

Tsou, W., & Kao, S.-M. (Eds.) (2017). *English as a Medium of Instruction in higher education: Implementations and classroom practices in Taiwan*. Springer.

Turner, J. E., Li, B., & Wei, M. (2021). Exploring effects of culture on students' achievement motives and goals, self-efficacy, and willingness for public performances: The case of Chinese students' speaking English in class. *Learning and Individual Differences, 85*, 101943. https://doi.org/10.1016/j.lindif.2020.101943

Vafadar, H. & Foo, T. C. V. (2020). The effects of communication strategies instruction on Iranian intermediate EFL learners' willingness to communicate. *Asian EFL Journal, 24*(4), 130–173.

Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology, 96*(2), 236–250. https://doi.org/10.1037/0022-0663.96.2.236

Zhang, W., & Liu, M. (2013). Evaluating the impact of oral test anxiety and speaking strategy use on oral English performance. *Journal of Asia TEFL, 10*(2), 115–148.

Zhu, J., & Mok, M. M. C. (2018). Predicting primary students' self-regulated learning by their prior achievement, interest, personal best goal orientation and teacher feedback. *Educational Psychology, 38*(9), 1106–1128. https://doi.org/10.1080/01443410.2018.1497775

APPENDIX 1

Questionnaire Items

Task Goals

TG01 I like to speak English in the EMI class, even if I make a lot of mistakes.

TG02 An important reason why I do tasks (e.g., discussion, conversation, or presentation) in the EMI class is because I like to learn new things.

TG03 I like the tasks in the EMI class best when they really make me think.

TG04 An important reason why I participate in the EMI class is because I want to get a better understanding of the content.

TG05 I participate in the EMI class because I'm interested in it.

TG06 An important reason I interact with classmates and teachers in the EMI class is because I enjoy it.

Performance-approach Goals

PP07 I would feel really good if I were the only one who could answer the teachers' questions in English in the EMI class.

PP08 It's important to me that the other students in my EMI class think that I am good at speaking English.

PP09 I want to speak better English than other students in my EMI classes.

PP10 I would feel successful in school if I spoke English better than most of the other students in the EMI class.

PP11 I'd like to show my teachers that my English ability is higher than the other students while speaking English in the EMI class.

PP12 Speaking English better than other students in the EMI class is important to me.

Performance-avoidance Goals

PV13 It's very important to me that I don't look stupid while speaking English in the EMI class.

PV14 An important reason I practice speaking English in the EMI class is that I don't embarrass myself in English oral communication.

PV15 In the EMI class, the reason I speak English is so my teachers don't think my English proficiency is lower than others.

PV16 In the EMI class, the reason I speak English is so others won't think my English is poor.

PV17 One reason I would not speak English in the EMI class is to avoid looking stupid.

PV18 One of my main goals of not speaking English is to avoid looking like I can't do my work in the EMI class.

WTC (Content Knowledge)

WTC01 I am willing to discuss the content knowledge in English in the EMI class.

WTC02 I am willing to give an oral presentation to the class with notes in the EMI course.

WTC03 I am willing to translate a spoken utterance from Chinese into English regarding the content knowledge in my group.

WTC04 When there is no need talk to the class, I am willing to do a task in English with my peer at my desk.

WTC05 am willing to give an oral presentation to the class without notes in the EMI course.

Communication Strategies (Social Affective)

CS01 In EMI class, I try to enjoy the discussion.

CS02 In EMI class, I actively encourage myself to express what I want to say.

CS03 In EMI class, I don't mind taking risks even though I might make mistakes.

CS04 In EMI class, I try to relax when I feel anxious.

CS05 In EMI class, I try to use fillers when I cannot think of what to say (e.g., well, oaky, you see, you know).

Communication Strategies (Accuracy-Oriented Strategies)

CS06 In EMI class, I try to talk like a native speaker.

CS07 In EMI class, I try to emphasize the subject and verb of the sentence.

CS08 In EMI class, I pay attention to grammar and word order during conversation.

CS09 In EMI class, I notice myself using an expression which fits a rule that I have learned.

Communication Strategies (Fluency-Oriented Strategies)

CS10 In EMI class, I take my time to express what I want to say.

CS11 In EMI class, I pay attention to my rhythm and intonation.

CS12 In EMI class, I pay attention to my pronunciation.

CS13 In EMI class, I pay attention to the conversation flow.

CS14 In EMI class, I try to speak clearly and loudly to make myself heard.

Communication Strategies (Message Reduction and Alteration Strategies)

CS15 In EMI class, I use words which are familiar to me while speaking English.

CS16 In EMI class, I reduce the message and use simple expressions.

CS17 In EMI class, I replace the original message with another message because of feeling incapable of executing my original intent.

Communication Strategies (Negotiation for Meaning While Speaking Strategies)

CS18 In EMI class, I repeat what I want to say until the listener understands.

CS19 In EMI class, I make comprehension checks to ensure the listener understands what I want to say.

CS20 In EMI class, I give examples if the listener doesn't understand what I am saying.

Communication Strategies (Nonverbal Strategies)

CS21 In EMI class, I try to make eye-contact when I am talking.

CS22 In EMI class, I use gestures and facial expressions if I can't express myself.

https://doi.org/10.17323/jle.2023.12008

Compliment Response Strategies in Institutional Discourse within an Emirati Context: Focus on Power and Gender Differences in University Student-Professor Exchanges in English

Tanju Deveci

Antalya Bilim University, Antalya, Turkiye

ABSTRACT

Background: Context plays a significant role in effective communication. Among various aspects of context, culture is particularly important since it necessitates that language be used effectively so that a specific purpose can be achieved successfully. One key element of such communication is the effective use of speech acts including compliment and compliment responses (CR).

Purpose: This research aimed to identify the CR strategies produced by Emirati users of English in a university setting, as a response to a compliment received from an international professor on their academic performance and the psychological effect such a compliment is likely to have on them. It also investigated the influence of gender on CR strategies.

Method: The data were collected using a discourse completion task. Fifty-eight students (33 male and 25 female) participated in the study. The CR strategies were analyzed using Holmes' (1988) classification scheme.

Results: The results showed that a compliment from a professor, irrespective of his/her gender, would make the students happy, with positive effects on their motivation, self-confidence, and feeling of closeness to the professor. The students also thought a CR was necessary for politeness purposes. The most commonly used CR strategy was that of acceptance. The male and the female students produced similar CR strategies in responding to the professor, irrespective of his/her gender. Yet they were more likely to use micro-level strategies (e.g., appreciation token, comment, and promise) with the male professor. The students also used downgrading and disagreeing but only while responding to the male professor. In their conversation with the female professor, they used the strategies of shifting credit and requesting reassurance.

Conclusion: These results provide evidence for the face-enhancing nature of CR strategies as utilized by Emirati users of English with international faculty in a university setting.

KEYWORDS

compliment, compliment response, Emirati university students, gender, power, speech acts

Citation: Deveci T. (2023). Compliment Response Strategies in Institutional Discourse within an Emirati Context: Focus on Power and Gender Differences in University Student-Professor Exchanges in English. *Journal of Language and Education, 9*(3), 38-51. https://doi.org/10.17323/jle.2023.12008

Correspondence:
Tanju Deveci,
tanjudeveci@yahoo.com

Received: February 02, 2021
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

Communication with other people helps us understand ourselves better, and the more we understand ourselves, the better we can communicate and connect with others (Deveci & Nunn, 2018), which points to the significant role interpersonal communication plays in well-being at individual and societal levels. One of the key determinants of effective communication is the context in which communication takes place. The four dimensions of context are physical, temporal, social-psychological, and cultural (DeVito, 2013). Within the scope of this paper, the latter two are of particular importance. Social-psychological context includes the status relationships between those involved in communication while cultural context determines what we say and how we say it according to cultural norms. Both of these contexts require "the ability to use language effectively in order to achieve a specific purpose and to understand language in context" (Thomas,

1983, p. 92). This includes the effective use of speech acts, the theory of which is based on "the assumption that the minimal unit of human communication [is] the performance of certain kinds of acts, such as making statements, asking questions, giving orders, describing, explaining, apologizing, thanking, etc." (Searle, et al., 1980, p. vii). This paper is related to one such act, namely complimenting, and focuses more specifically on *compliment responses* (CRs), the performance of which is heavily influenced by a variety of factors including the interlocutors' gender as well as cultural backgrounds.

Holmes (1988) emphasizes the importance of studying the contexts in which compliments and CRs occur. What is central to the current research is the social-psychological context (DeVito, 2013), which studies CRs in a particular situation involving a higher-lower-status relationship in an institutional setting (a university). A related term, 'social distance,' "determines the way in which interlocutors converse precisely because it is an important determinant of the degree of comfort or politeness/deference in a verbal exchange. This, in turn, determines the constraints felt and the liberties taken in speech exchanges» (Boxer, 1993, p. 103).

Previous research has revealed that social distance in different contexts, including student-professor, can be perceived differently by interlocutors from distinct cultural backgrounds, which impacts strategy choices (Maeshiba et al., 1996). The utilization of a variety of speech acts in student-professor contexts has been studied. Among these were complaint (Önalan & Çakır, 2018), request (Deveci & Hmida, 2017; Yang, 2009), apology (Bataineh, 2005), and refusal (Deveci & Midraj, 2021). However, there is a dearth of studies on students' CR strategies when receiving a compliment from their professors, a situation which is heavily influenced by the role social power plays in determining how interlocutors use the language.

It also appears that in this particular setting (i.e., university professor-student exchanges) how a professor's gender impacts CRs has not been investigated, especially within an Emirati context. Neither has the students' gender. Gender is an important aspect of culture, and how it plays out in daily conversation is of great significance. Tannen (1994) approached gender differences in language use from the cultural difference perspective and noted that men differ from women in that they communicate factual information, while women tend to pay closer attention to building and maintaining relationships. However, it appears that the impact of gender on compliments and CRs has been generally investigated in studies situated in western cultures, with a lack of reference to the Middle Eastern cultures, particularly in the Emirati context, where gender plays a substantial role in daily communication, whether in non-academic or academic settings.

There indeed is empirical evidence that differences exist between male and female Emirati speakers' overall communication styles and linguistic tendencies. About the former, Ahmad (2014) points to several differences. In comparison to Emirati women, Emirati men are more direct and to the point and less hesitant to share opinions. They also take the initiative to speak more often. Emirati men are also noted to use abusive terms and harsh language during conversations with friends and colleagues whereas Emirati women refrain from such language. Ahmad (2014) also notes that Emirati women are more open to conversations with other nationalities than they are to conversations with Emirati men. In a study on expressive writing produced by male and female students in her Emirati context, Almazroui (2009) found that the female students expressed their personal experiences and feeling more often than the male students while the male students exhibited domination and overstatements of self. In a more recent study on expressive writing in the form of reflective essays, Deveci and Ayish (2021) found that male students tended to use adjectives more often than their female counterparts did. However, the female students used attitudinal adjectives with higher frequencies. Together, these studies point to potential differences in the language used by Emirati males and females when producing a CR.

Because 88.52% of the UAE residents are expatriates from a variety of countries[1] (GMI, 2019), the English language is used extensively in daily life. It is also used as the medium of instruction in all UAE universities. Although the profile of faculty may vary to a great extent, students are expected to interact with their professors in socially and culturally appropriate ways. The same is true for faculty as well. Effective instruction includes positive feedback, which helps build strong relationships with students (Kington et al., 2014), thus increasing motivation for learning. Positive feedback can take the form of a compliment *or* be perceived as one by students. And students' perceptions of compliment matter, especially when it comes from a professor of the opposite gender because it is likely to determine how they respond or if they respond at all. It is also possible that female and male students behave differently in such a situation. Therefore, it is important to identify student tendencies in responding to male and female professors' compliments. Toward this end, this research aims to answer the following questions:

(1)

a) What are the psychological effects of a compliment paid to Emirati university students by a professor on their assignment?
b) How do the female and the male students' responses compare?
c) Do their responses differ according to the professor's gender?

[1] GMI. (2019, January 30). United Arab Emirates population statistics. https://www.globalmediainsight.com/blog/uae-population-statistics/

(2)

a) Do the students think a compliment from a professor on their performance would require a response? Why (not)?
b) How do the female and the male students' responses compare?
c) Do their responses differ according to the professor's gender?

(3)

a) What CR strategies do the students use in English?
b) How do the female and the male students' CRs compare?
c) Do their CRs differ according to the professor's gender?

LITERATURE REVIEW

Politeness

The notion of 'face,' borrowed from Goffman, is central to Brown and Levinson's (1987) model of politeness. They elucidate that face is related to embarrassment or humiliation. Face can be 'threatened,' 'lost,' 'maintained,' 'saved,' or 'enhanced.' Considering the mutual vulnerability of face, interlocutors are encouraged to cooperate in maintaining face. Brown and Levinson (1987, p. 312) also note that "while the content of face will differ in different cultures ... the mutual knowledge of members' public self-image or face, and the social necessity to orient oneself to it in interaction, are universal."

There are two categories of face: positive face and negative face. The former is related to people's general desire for recognition, appreciation, and approval. The latter is related to their desire to avoid imposition and to be free in their choices and actions. The notions of positive face and negative face are linked to the notions of positive politeness and negative politeness. While positive politeness is "oriented toward the positive face of the hearer," negative politeness is "oriented toward partially satisfying [the hearer's] negative face" (Brown & Levinson, 1987, p. 70). For communication to be effective, both aspects ought to be respected and maintained.

Brown and Levinson (1987, p. 312) point to the fact that some acts are intrinsically face-threatening; that is, they "run contrary to the face wants of the addressee and/or of the speaker." Three factors determine the extent to which an act is face-threatening and therefore the kinds of politeness strategies an individual would employ. The first one is 'power,' which is the perceived power dynamic between individuals (e.g., a teacher vs. a student). The second one is 'distance,' which is related to social distance between individuals. Compare, for example, a close friend to an acquaintance in the neighborhood. The last one is 'rank,' which refers to the relative sensitivity of a topic in a particular culture (e.g., a person's age).

Brown and Levinson's model of politeness has received some criticism. Al-hindawi and Alhkazaali (2016) identify the main reasons for this. This first one is related to the model being static. That is, a rule-like system of strategies and their linguistic realizations fails to explain human interaction in diverse and dynamic contexts. In support of this, Werkhofer (1992) notes that a variety of social factors including distance and power have a determining role in interactive negotiations. Another criticism rests upon the argument that their theory of politeness is Western-centric with a bias towards individualistic cultures; therefore, it does not explain variations in other cultures (Fukada & Asota, 2004).

Despite the criticism it has drawn, Brown and Levinson's model of politeness has attracted much interest from researchers in diverse cultures. Research results pointed to the various conceptualizations of the notion 'face' in different contexts. We now have a greater understanding of politeness strategies as utilized by people from different cultures as well as their manifestations during exchanges between people from diverse cultural backgrounds. Accordingly, in this study, I, too, use Brown and Levinson's (1987) conceptualization of face as a foundation for understanding how politeness manifests itself in the context of Emirati speakers of English responding to compliments when interacting with male and female international professors. As a face-saving act, compliments can best be conceptualized using Brown and Levinson's model of politeness; they maintain and improve the addressee's face and maintain harmonious relations (Leech, 1983). They, therefore, have been labeled as «social lubricants» (Holmes, 1988, p. 486). Compliment responses, in the same way, save and enhance the complimentee's and complimenter's faces, thus improving social interaction. The impacts of compliments and compliment responses on politeness are further discussed below.

Speech Acts

The foundations of the speech act theory laid by Austin's (1962) seminal work proposed the term 'performative utterances.' While 'performative' is derived from the verb 'perform,' 'utterances' are taken as a functional unit of communication (Richards, Platt, & Weber, 1985). Searle advanced the notion of performative utterances as the basis for the speech act theory. Searle (1965, p. 2) defined a speech act as "the production of the sentence token under certain conditions ...and the minimal unit of linguistic communication." Examples of speech acts include thanking, complimenting, responding to compliments, complaining, apologizing, refusing, and requesting. Searle (1969, p. 18) also noted that

"every meaningful sentence in virtue of its meaning can be used to perform a particular speech act (or range of speech acts)." This points to the significance of context for the speech act theory. Searle (2002) stated that a particular utterance would have a particular meaning under certain conditions. He also emphasized the distinction between 'speaker meaning' and 'sentence meaning' in any given context. These notions can be clarified and illustrated via the speech acts of compliment and CRs.

The Compliment Speech Act

Among the widely cited definitions of the compliment speech act is Holmes' (1988, p. 446): an utterance that "explicitly or implicitly attributes credit to someone other than the speaker, usually the person addressed, for some 'good' (possession, characteristic, skill, etc.) which is positively valued by the speaker and the hearer." Four main functions of the compliment speech act have been identified (Hatch, 1992, pp. 138-139): a) to establish rapport and ease the transition from greeting to the first topic of conversation, b) to reinforce and encourage desirable performance, c) to reinforce the speech act of thanking, and d) to soften the speech act of criticism.

A compliment is considered a face-enhancing act (Taavitsainen & Jucker, 2008) since it serves to enhance the face of the hearer. In this way, a compliment helps oil the wheels of social relations (Holmes, 1988). However, a compliment can also be a face-threatening act when it obligates the hearer to "offer the desirable good to the complimenter" (Holmes, 1988, p. 448). Whether it is face-enhancing or face-threatening, a compliment normally produces a CR (Xiang, 2013). Just as a compliment serves to negotiate solidarity, so does a CR (Ziaei, 2012), and they both play a key role in determining the extent to which communication between interlocutors is successful (Al Harbi, 2017).

CRs were first analyzed by Pomerantz (1978) in American English; she identified three broad strategies: acceptance, rejection, and avoidance of self-praise. A later study by Holmes (1988) also identified three main categories at the macro level (i.e, broad categories of compliment responses including ACCEPT, REJECT, and DEFLECT/EVADE.) Each of these in turn are divided into micro-level strategies, that is, "sub-categories with fairly transparent labels" like 'appreciation' and 'agreeing' (Holmes, 1988, p. 461). These are described in Table 1. Note that the micro-level strategy 'comment'[2] and the macro-level strategy 'promise' in the table are additional ones based on the results of the current study.

Previous Studies on CR Strategies around the World

The interest in research on CR strategies has increased significantly over the years. Many researchers have investigated CRs in face-to-face communication and more recently an interest in complimenting in computer-mediated communication is gaining momentum (e.g., Placencia & Eslami, 2020; Sartini, 2019; Dehkordi & Chalak, 2015). The focus of this study is on face-to-face communication and therefore, the studies reviewed below are all related to face-to-face communication.

Researchers investigated CR strategies in different languages and cultures. Guo, Zhou, and Chow (2012) studied CR strategies in Mandarin Chinese. The data collected using a

Table 1

Compliment Response Strategies

Macro- level	Micro-level	Examples
Accept	Appreciation token	"Thanks", "Cheers"
	Agreeing utterance	"I know", "I am glad you think so"
	Downgrading/qualifying utterance	"It's nothing", "It's not bad"
	Return compliment	"I'm sure you will be great", "Yours was good too"
	Comment	"I have worked hard on this work"
Reject	Disagreeing	"I thought I did it badly", "Don't say so"
	Question accuracy	"Is it right?", "Really?"
	Challenge sincerity	"Don't lie", "Don't joke about it
Deflect/Evade	Shift credit	"No worries", "My pleasure"
	Informative content	"It wasn't hard", "It's really cheap"
	Request reassurance	"Really?"
Promise		"I will impress you more in the future"

[2] Comment was also identified by Herbert (1986) to accompany acceptance.

naturalistic observational method showed that young participants tended toward acceptance strategies. Their use of down-graders and disagreements was rather limited. The researchers argued that this was similar to the behaviors of western cultures. They noted that a possible reason was the prominent use of English in the context where the research was conducted. However, in another study investigating Chinese university students' CR use with a classmate (i.e., a relationship based on equality and solidarity), Tang and Zhang (2009) found that the participants were more likely to evade and reject than to accept. This was particularly the case when the compliment topics involved character and possession. Due to its collectivist culture, the authors argue, the students likely felt that it was their duty to help peers so they did not expect praise/a compliment. The authors also note that people in a collectivist culture are expected to be modest and refrain from showing their wealth. Therefore, the participants tended to reject a compliment on their possessions.

In a more recent study, Tang (2020) investigated the impact of *gender* on responding to different types of compliments (e.g., appearance, ability, possessions) in Mandarin Chinese. She found that the participants' perceived gender roles played a role in the ways they produced CRs, which were influenced by "the social expectations on masculinity and femininity in their particular speech community" (Tang, 2002, p. 545). Tang (2021) also studied Taiwanese Mandarin speakers' CR strategies. In doing so, she focused on the participants' gender as well as the gender and relative social power of their addressees. The findings indicated that these variables all had an impact on the participants' CR strategy-use. The types of microstrategies employed differed according to the social expectations of the addressees with different positions and of a different gender. The addressees' psychological perceptions of the compliment played a role, too. Tang (2021) also found that acceptance was a popular strategy among the participants. She believes this is likely because of the emphasis placed upon positive politeness in Taiwan together with «the value of prioritizing the self rather than the traditional Confucian value of humility» (Tang, 2021, p. 99). The latter, however, appear to contradict findings from earlier research by Yu (2004), who found that Mandarin Chinese speakers from Taiwan often opted for nonacceptance in responding to compliments. The author states that this possibly stems from the Chinese culture attaching high value to relative power and modesty in spoken interactions. She also compares this finding to another indicating that American speakers of English tend to favor acceptance as a CR strategy.

Other researchers studied CR strategies in western languages. For example, Dimova (2014) investigated their use in Bulgarian and German. She found that although the speakers of both languages generally accepted compliments, the Bulgarian speakers were more likely to use this strategy than their German counterparts. The appreciation token (i.e., thanking) was the most frequent micro-level strategy utilized by both groups of speakers. The female participants in both groups were more likely to accept compliments. Exclusive to the male participants was the use of humor and explaining as micro-level acceptance strategies. Regarding the impact of interlocutors on CR strategies, Furkó and Dudás (2012) found that more acceptance was produced during female-female interaction in the Hungarian context.

It is also important to note that some research results point to the presence of certain collectivist elements in individualistic cultures (or vice-versa). To exemplify, Gonzalez-Rodrigues (2012) studied Icelanders' use of CR strategies and found that the most frequent strategy was rejection in the form of disagreement. They also used the strategies of evasion and down-graders. The researcher reasons that this was due to Icelanders' tendency for modesty. The results also revealed that the female participants were less likely to accept compliments. It is interesting to note that the participants tended to accept compliments on their personal belongings rather than their physical appearance. This indicated that the topic of compliments had a determining effect on the participants' choice of CR strategies. Recent research in Iran revealed that CR strategies used by university students were shifting from rejection and/or evasion to accepting (Sarkhosh, 2022). Acceptance was also found more common among female students. The author reasons that the former is likely because of young Persians' increased exposure to and interaction with foreign media (e.g., the Internet, satellite TV). The latter, he argues, may be due to psychological reasons making the compliment receiver want her appearance "to be acceptable and worthy of a compliment" (Sarkhosh, 2022, p. 288). Together, these results indicate that it is not always possible to draw a line between CRs used by people of different cultures and of a different gender.

Studies on CR Strategies in the Arab World

Given that users of English are likely to rely on their mother language and transfer aspects from it while complimenting/responding, both at the pragmalinguistic (linguistic resources) and the sociopragmatic (weighing of the situational variable like power and the selection of the available linguistic resources) levels, it is of interest here to report some recent studies and shine a light on certain trends of Arabic speakers. Dendenne (2021) investigated compliments and CR strategies in colloquial Algerian Arabic based on naturally-occurring data. Results showed that compliments were often accompanied by divine invocations for the well-being of the hearer, which has cultural underpinnings. For example, the utterance "*It was a delicious dish*" uttered after/during having a meal, which may be accompanied by another

one like "*May God grant him [the cook] all his wishes*" can be perceived as a "verbal gift" and, to which the hearer may respond using a formulaic phrase. Dendenne (2021) also states that the hearer may respond to such verbal gifts non-verbally or ignores if he/she is very familiar with them.

Another study conducted by El-Dakhs (2017) found that Egyptian speakers of Arabic preferred the agreement strategy, which included explicit appreciation tokens, attribution to complimenters, and offering gifts. They also used some additional responses, examples of which were prayers for complimenters and terms of endearment. The researcher noted that the participants often used CR Arabic phrases without their literal meanings in a humorous manner. These phrases were generally used to express attribution to the complimenter and self as well as the expectation of self.

The acceptance strategy in Arabic was encountered by other researchers too. For example, Ebadi and Salman (2015) found it in CRs used by Iraqi speakers of Arabic. They also noted that the participants often used formulaic sequences (e.g., Shukren [Thanks]). The results did not reveal statistically significant differences between the female and the male students' utilization of CR strategies in Arabic. In one of the earliest studies on Arabic CRs, Nelson, Al-Batal, and Echols (1996) investigated CRs in Syrian Arabic and found that the majority of the participants (67%) opted for acceptance often utilizing formulaic language.

The utilization of CR strategies by Arabic-speaking users of English has been studied by some researchers, too. For instance, Nouichi (2018) studied Algerian university students' CRs in English. The results showed that a significant proportion of CR strategies (86.4%) was in the form of acceptance. The students mainly produced an appreciation token while a small group of students either agreed with the compliment or provided a supplication to God (3.84% and 2% respectively). Rejection was produced by 7.8% of the students and deflection by 3.2%. Similarly, in her master's thesis investigating female Saudi university students' CRs in English, Al Harbi (2017) found that the students showed a greater tendency towards the acceptance strategy. This was mainly due to their cultural norms indicating that acceptance of a compliment was polite and therefore the expected behavior. Similarly, Alsalem (2015) also found that Saudi university students tended to accept compliments, irrespective of gender. It was also found that the students tended to attribute their success to their professors. However, another study conducted in Jordan revealed that female university students used the acceptance strategy more often than their male counterparts (AlRousan & Awal, 2016). In the Iraqi context, female learners of English were found to use more appreciation tokens and questions than male students (Ebadi & Salman, 2015). The researchers explained that the latter might be due to females' tendency towards asking for reassurance and/or repetition of the compliment paid.

Considering the interlocutors' gender, Almallah (2017) investigated the English CR strategies utilized by Palestinian and Jordanian university students. He found that both the female and the male students opted for agreement with compliments paid by both the same and the opposite genders. However, the female students were found to utilize acceptance on appearance and performances more often than their male counterparts, who had a greater tendency towards disagreement with compliments on appearance made by males. Similar results occurred in previous research conducted by Salameh (2001) in the Saudi context; the female speakers had a greater tendency to accept compliments made by females.

To the best of my knowledge, the only research conducted on Emirati language learners' CRs is one by Al Falasi (2007), which focused on female students. The results showed that the students often committed pragmatic transfer. Some CRs (e.g., "*I am ashamed!*") used as a politeness strategy were considered inappropriate in the target language (i.e., English). It was also found that the students' CRs were long, which was attributed to their belief that longer CRs would express more sincerity.

The aforementioned studies indicate a growing interest in understanding the face-enhancing/threatening nature of the compliment speech act and CR strategies in the Arabic language as well as by Arab non-native speakers of English. However, there are still major lacunae in the study of this speech act, in the UAE context in particular. Neither does the current literature elucidate how Emirati university students respond to positive feedback, which may be considered to be a compliment, from faculty with whom they are in constant contact. It is also important to identify the impact of faculty's status and gender on students' utilization of CR strategies. The current research could make a substantial contribution to the literature.

METHOD

Participants

Data were collected from 58 students registered in Introduction to Linguistics at a UAE-based university. The Institutional Research Board (IRB) approval was sought and granted to utilize data from student responses to an instructional activity designed both to introduce students to the concept of speech acts, compliments in particular, and to demonstrate data collection methods via discourse completion tasks (DCT). Twenty-five (43%) were female and 33 (57%) were male. Their ages ranged from 19 to 25 with an average age of 21. The students were proficient speakers of the English language; before their university studies, they had attained sufficient passing scores from international examinations such as iTOEFL (minimum score of 91) and IELTS (minimum score of 6.5). They also enhanced their English

language skills through the various courses they took before they were registered in Introduction to Linguistics. The comparison of the male and the female students' grades from the compulsory first-year English course taken previously, which has a heavy emphasis on writing skills, also revealed that their average grades were in the same band (B).

Data Collection Tool

Data were collected using a discourse completion task (DCT) comprised of three sections. The first section collected data related to demographics including gender and age. The second section described two hypothetical situations[3] as described below. The students were asked to indicate how a compliment paid by an international male and an international female professor would affect them.

> Situation 1: Imagine you are taking X course from a *male* British professor in his late forties. And you are visiting him in his office to talk about a recent assignment you have written. Your professor seems happy with your work. He hands the assignment to you saying, "Well-done! This is a well-written assignment. Not only did you use the language effectively but you also explained the topic well with good examples and evidence."
>
> Situation 2: Imagine you are taking X course from a *female* British professor in her late forties. And you are visiting her in her office to talk about a recent assignment you have written. Your professor seems happy with your work. She hands the assignment to you saying, "Well-done! This is a well-written assignment. Not only did you use the language effectively but you also explained the topic well with good examples and evidence."

The third section asked the students if they thought the compliments in the hypothetical situations would require a response. They were asked to justify their responses. Following this, they were asked to write the exact words they would use if they chose to respond to their professor.

Although a DCT fails to collect naturally occurring data, it is not always possible or practical for researchers to collect such data (Hartford & Bardovi-Harlig, 1992). In the context of the current study, for instance, the local culture does not allow recording people's voices, particularly those of females, without their permission. Even if permission were granted, the authenticity of such data would still be questionable. There is also evidence from previous research that a DCT and naturally occurring data may reveal similar trends concerning directness and lexical modification in a request speech act (Economidou-Kogetsidis, 2013), which might be the case for CRs on which the current study focused.

Data Analysis

Data related to demographics and data for the first and the second research questions were analyzed using descriptive statistics such as frequencies and averages. A t-test was used for statistical comparisons between these data sets. A p-value of less than .05 was considered significant.

The CR strategies produced by the students in response to the third research question were analyzed using Holmes' (1988) scheme. Data analysis, however, revealed one additional macro-level strategy (i.e., promise) and one micro-level strategy (i.e., comment). Hence, data specific to these were categorized accordingly. To ensure the correct categorization of the strategies, inter-coder reliability was sought. To this end, an independent researcher participated in the analysis. The inter-coder reliability was assessed using the following formula by Miles & Huberman (1994): Reliability=Number of agreements/Number of agreements

Table 2

Effect of the Professor's Compliment on Students

		Female students (n=25)	Male students (n=33)	t	p*	All students (n=58)	t	p*
Compliment by a female professor	Min	4	3			3		
	Max	5	5			5		
	X	4.8	4.6	1.4841	.0717	4.7		
	SD	.37	.6			.52	.5073	.3064
Compliment by a male professor	Min	4	3			3		
	Max	5	5			5		
	X	4.7	4.6	1.0117	.1579	4.6		
	SD	.43	.65			.57		

Note. p<.05.

[3] A two-week interval was given between the data collection phases for the first and the second situations. This was so that the students would be prevented from being influenced by their responses to the first situation.

+ disagreements. The reliability result was 85%. Divergences were discussed, and agreements were reached.

No statistical test was used in analyzing the data for the third research question mainly because of the limited number of participants and too few data available preventing a reliable comparison made between the data sets related to some macro and micro strategies in particular. The chi-square test, for instance, fails to yield (reliable) results if cell frequencies drop below one. Instead, therefore, the results were reported considering the frequencies and the total number of instances. Also important to note is that no statistical test may be required in the absence of a hypothesis (Swinscow, 2009), as was the case in this study.

RESULTS

The first research question aimed to identify the effects of a compliment the students were paid by a professor on an assignment. To this end, they were first asked to indicate the extent to which they thought the compliment would make them happy (1=not at all, 5=extremely). Their responses are given in Table 2.

As shown in Table 2, the students' responses ranged from 3 (moderately happy) to 5 (very happy) with an average of 4.7 (SD=.52) and 4.6 (SD=.57) for a compliment by a female and a male professor respectively. There was no statistical difference between their responses for the two professors (t=.5073, p=.3064 >.05). The t-test conducted according to the students' gender did not reveal differences either.

The students were also asked to justify their reasons. Their responses included the compliment increasing their motivation (f=30), enhancing their self-confidence (f=23), acknowledgment of their efforts (f=20), and making them feel closer to the professor (f=5). One student said, *«It is normal if I produced good work.»*

The second research question asked if the students thought the compliment from their professor would require a response. Table 3 summarizes their responses.

As can be seen in the table, all the students speaking to a female professor said they would say something in return for the compliment they were paid. Only one male student added that he would smile, which is a non-verbal response. Similarly, all the students said they would say something in return for the compliment paid by a male professor. When asked to justify their responses, most of them (f=54) said that it would be rude to behave otherwise. Some said thanking the professor would acknowledge their feedback. The male student that indicated he would smile said he was a shy person and being praised would make him blush.

The third research question was related to the students' strategy use in their CRs. The results are given in Table 4.

CRs to a Female Professor

According to Table 4, a total of 118 micro-level CRs were produced in the DCTs for the female professor. Of this number, appreciation token and commenting accounted for the majority (42.4% and 40.7% respectively). The former mainly included a thank-you note to the professor (e.g., "*Thank you very much*," [4] "*I am grateful for your nice words*."). The latter mainly concerned the amount of time spent doing the assignment (e.g., "*I have worked hard on this work*").

There were some other less frequent micro-level strategies. Five students (4.2%) expressed their agreement with the professor. For example, one student said, "*I thought it would make you happy*." Another one simply said, "*Yes*." Five other students (4.2%) produced a promise committing the students to maintain the quality of their work (e.g., "*I will impress you more in the future*."). On the other hand, four students (3.4%) responded by paying the professor a compliment in return. One said, "*You are the best professor*.» Another remarked, «*You are good at your job*." There were also three instances of evasion in the form of shifting the credit toward the teacher (2.5%) (e.g., "*Thanks to you,*" "*I thought it will be hard but you make it easy to us"*) and three instances of requesting assurance (2.5%) (e.g., "*Did it make you happy?,"* "*I hope you are serious*.").

Table 3

Student Reactions to Compliment

	Reactions	Female students (n=25)	Male students (n=33)	All students (n=58)
Compliment by a female professor	Yes, it would.	25	33	58
	No, it would not.	-	-	-
Compliment by a male professor	Yes, it would.	25	33	58
	No, it would not.	-	-	-

[4] Student responses are reported verbatim – without editing.

Table 4
*Compliment Response Strategies**

Macro-level	Micro-level	CRs to a Female Professor				CRs to a Male Professor			
		Female ss (n=25)	Male ss (n=33)	*Total*		Female ss (n=25)	Male ss (n=33)	*Total*	
		f	*f*	*f*	*%*	*f*	*f*	*f*	*%*
Accept	Appreciation token	27	23	*50*	*42.4*	31	34	*65*	*43.9*
	Agreeing utterance	2	3	*5*	*4.2*	5	3	*8*	*5.4*
	Downgrading/qualifying utterance	0	0	*0*	*0*	1	0	*1*	*0.7*
	Return compliment	4	0	*4*	*3.4*	1	4	*5*	*3.4*
	Comment	25	23	*48*	*40.7*	30	28	*58*	*39.2*
Reject	Disagreeing	0	0	*0*	*0*	1	0	*1*	*0.7*
	Question accuracy	*0*	0	*0*	*0*	0	0	*0*	*0*
	Challenge sincerity	*0*	0	*0*	*0*	0	0	*0*	*0*
Evade	Shift credit	1	2	*3*	*2.5*	0	0	*0*	*0*
	Informative content	0	0	*0*	*0*	0	0	*0*	*0*
	Request reassurance	0	3	*3*	*2.5*	1	0	*1*	*0.7*
Promise		4	1	*5*	*4.2*	3	6	*9*	*6.1*
Total		*63*	*55*	*118*	*100*	*73*	*75*	*148*	*100*

Note. Frequencies and percentages were calculated from the total number of utterances as the students often employed more than one strategy.

The data were also compared considering the gender variable. The female students were found to utilize more micro-level CRs than the male students (f = 63 vs. f = 55). It is also important to note the occurrence of return compliment in the female data set (f = 4). Similarly, a promise was detected four times in the female data set and only once in the male data set.

CRs to a Male Professor

Table 4 shows that there was a total of 148 occurrences of micro-level CRs. The most frequently occurring micro-level strategy was that of appreciation token (43.9%). The most frequent sentence/phrase used was "*Thank you*." Another one was "*I appreciate your kind words*." Comment was the second most frequent micro-level strategy (39.2%). Many focused on the hard work the students engaged in. Note these sample utterances: "*I have truly tried to make sure the assignment checked all the requirements*" and "*I tried to do my best in this assignment*."

Less infrequent, other micro-level strategies were adopted: promise (6.1%) (e.g., "*I will please you always*" and "*I promise I will try and keep improving upon my language skills*"), agreeing utterance (5.4%) (e.g., "*I am glad this good work fulfilled your expectations*"), return compliment (3.4%) (e.g., "*You are a great teacher*" and "*I really think you should be proud of yourself*"), downgrading (.7%) ("*How small it is!*"), disagreeing (.7%) ("*I wish it was really that good*."), and requesting assurance (.7%) ("*Really*?").

The comparison of the data set according to gender showed that the female and the male students produced a similar number of CRs (f=73 and f=75). However, an appreciation token was used slightly more often by the male students than the female students (f= 4 vs. f=31). Similarly, the male students were more likely to produce a promise than the female students (f=6 vs. f=3). They had also more of a tendency to return the compliment (f=4 vs. f=1). On the other hand, a comment was produced slightly more often by the female students than the male students (f=30 vs. f=28). And the female students were slightly more likely to agree with the compliment by the professor (f=5 vs. f=3). Although used only once each, downgrading, disagreeing, and requesting assurance were used only by the female students.

Comparison of CRs according to Professors' Gender

Table 4 shows that the students generally produced similar CRs in their responses to the male and the female professors. Yet it is still important to note that the total number of the micro-level strategies employed with the male professor (f=148) was higher than that of the micro-level strategies

with the female professor (f=118). For instance, the former included 15 more appreciation tokens, ten more comments, and four more promises. Infrequently, downgrading and disagreeing were also used only with the male professor. On the other hand, shifting credit was only used with the female professor. Also, requesting assurance was used slightly more often with the female professor than the male professor.

DISCUSSION

The results of this study showed that both the male and the female students believed a compliment from their professor, whether a male or a female, would make them quite happy. They said this would increase their motivation to engage in further learning. This finding is in line with the findings of previous research. For example, Hancock (2000) also found that university students who received positive verbal feedback, which could be considered a compliment (Pillet-Shore, 2012), from their professors significantly increased the amount of work they did outside the classroom. The students in the current study also maintained that the compliment they were paid would enhance their self-confidence and make them feel closer to the professor. This feeling, earlier research found, would cause more positive and less negative teacher-student interaction (Conroy et al., 2014). Some students in the current study also remarked that the compliment would indicate the professor's acknowledgment of their hard work, which provides support for previous research indicating verbal praise provided college-aged students with information on their competence and the value of their accomplishments, thus enhancing their well-being (Barker, 1992).

These data lend credence to the speech act theory suggesting a compliment enhances the hearer's face (Taavitsainen & Jucker, 2008). The positive effects on the students' self-confidence and feeling of closeness to the professor, who is of a higher position, also support the compliment's role in enhancing social relations (Holmes, 1988). Further, the students' answers to the second research question indicated that the students, irrespective of their gender, found it necessary to produce a CR. This was the case for the exchanges with both the male and female professors. Considering the Middle Eastern context in which the current study was conducted, this is an important finding, supporting the observation that a compliment is often accompanied by a CR (Xiang, 2013). This finding also underscores the nature of a compliment as a face-threatening act if not responded to properly (Ziaei, 2012; Al Harbi, 2017), which was reflected in the students' responses related to a CR being courteous. From the social-psychological perspective (DeVito, 2013), the students' approach can be said to have considered the higher-lower status relationship they often engaged in. However, it is also important to note that the higher status of the complimenter may render CRs even more face-threatening than in normal circumstances.

The third research question investigated CR strategies utilized by the students in responding to compliments from an international male and an international female professor. The results showed that in both scenarios the students of both genders used similar micro-level strategies, an appreciation token and comment being the most common ones. The former may have been prompted by the students' general tendency to be humble, which is one of the values considered to make up the Emirati culture[5]. A simple phrase of 'Thank you' is humble and expresses gratitude[6]. The appreciation token commonly used by the male and the female students in this study, irrespective of the professor's gender, also supports the findings of the earlier research indicating that accepting a compliment was a polite, expected reaction in the Gulf region (Al Harbi, 2017; Alsalem, 2015). However, the observation that Emirati females tend not to accept a compliment from a male unless he is a family member (Al Falasi, 2007) is not supported by the data in the current study. The academic context in which the current study was conducted must have caused this difference. It is a context in which female students are usually in contact with male faculty from different backgrounds. Previous research showed that female students were more likely to contact their professors than male students regarding their coursework (Halawah, 2006), which would generate ample praise for their performance.

Commenting is also an important strategy to note, which was noted by Herbert (1986) too. Its occurrence in the current study can be attributed to the academic context in which the study was set; conversations on student performance have been found to trigger comments in the form of justification for a grade students considered too low (Deveci, 2010, 2015).

It is also interesting to note the significantly higher number of micro-level strategies used when responding to the male professor in comparison to the female professor. This was mainly caused by the male students' more frequent use of an appreciation token and a return of the compliment. Albeit less frequent, the male students also produced twice as many promises as the female students. It is also important to note the female students' tendency to produce a greater number of strategies in their CRs to the female professor. For instance, they produced more promises and appreciation tokens than the male students. This finding is similar to the

[5] Perrett, R. (2018, August 9). Why cultural preservation is on top of the agenda here. Retrieved from https://adwonline.ae/cultural-preservation-top-agenda/

[6] Whitmore, J. (2015, June 10). How to gracefully accept a compliment. https://www.huffpost.com/entry/how-to-gracefully-accept-_b_7042718

findings of Ebadi & Salman (2015), indicating that Iraqi female speakers of English also had a greater tendency to utilize the appreciation strategy than their male counterparts. Taken together, these data seem to indicate that the male Emirati students might have felt more at ease responding to a compliment by a male professor while the female Emirati students felt more comfortable responding to a female professor's compliment. For female Emiratis, this finding may suggest that a CR to the opposite gender is more face-threatening than it is for male Emiratis. This may be attributed to the social forces in the region limiting (and sometimes even preventing) Emirati women's interaction with men outside their families (Dariela et al., 2017), despite the ever-increasing interaction between female tertiary students and male faculty members. But then the female Emirati students' tendency to use a greater number of strategies, acceptance in particular, is line with findings of earlier research focusing on the effects of interlocutors' on CRs (Almallah, 2017; Furkó & Dudás, 2012; Sarkhosh, 2022; Salameh, 2001).

Infrequent as it was, the students' utilization of promise (that of the female students in particular) bears reiterating. Promises act as a positive politeness strategy indicating cooperation on the part of the speaker (Brown & Levinson, 1987). In this way, the students performed an illocutionary act to maintain the professor's positive impression of them and thereby the good relationship that has been established. Thus, a promise in CRs can be said to enhance the complimenter's positive face (Ogiermann, 2009), but it might be misunderstood by an outsider to the Emirati linguaculture. The students' realization of promise may be due to the heavy emphasis placed upon obedience in the Emirati culture; individuals are often raised to show respect to people of higher status, one way of which is to be obedient. During student-teacher interaction, this may manifest itself in the form of a promise to maintain good work, which naturally results in the teacher's positive impression of the student. This approach is common in high power distance cultures (Jandt, 2018) like that of the UAE. Therefore, the likelihood of a student's promise to sustain the quality of work complimented upon may be rather low in many western cultures with a low power orientation.

It is also important to note the presence of reassurance requests by the male students only in their CRs to the female professor. Similarly, a request reassurance was used by a female student responding to the male professor. Ebadi & Salman (2015) also found that female students in their study opted for a question when paid a compliment. Only very few students indeed produced a request in the current study, but I believe this still raises the question of whether a compliment from a professor of the opposite gender is likely to cause Emirati tertiary students to seek confirmation. The possible socio-cultural motivation behind this warrants further investigation.

CONCLUSION

This study sought to identify the effects of a compliment paid in English from international male and female professors on Emirati tertiary students. It was found that the students, irrespective of their genders, thought it would enhance their well-being through its positive impacts on their motivation for studying and self-confidence. The students also felt their interpersonal relationships with the professors would be improved. In addition, the students thought that they would need to be polite and produce a compliment response. The type of compliment responses they produced mainly included acceptance. To this end, they often produced an appreciation token.

Although the female and the male students used similar strategies to a large extent, there were some differences in their strategy use when interacting with the male and the female professors. These differences point to the possibility of the male students feeling more at ease responding to a compliment from a male professor. And the opposite appears to be the case for the female students. Collectively, these findings underscore the importance of Emirati English language speakers' socio-linguistic behavior patterns they appear to adopt despite the universal politeness strategies (Brown & Levinson, 1987). Such orientations ought to be recognized by the international faculty dominating the tertiary education in the region. Only in this way can interpersonal relationships between distinct cultures run smoothly and chances of cross-cultural misunderstandings decrease enormously.

At the very end, it is important to acknowledge that this study has some limitations. One of these is the data collection technique used: a DCT may fail to adequately represent the actual language use in natural settings. However, the context in which this study was conducted does not allow researchers to record conversations; one of the reasons is the fact that female participants – in particular – reject being recorded even when their performance is only audio-taped. Future researchers can aim to collect naturally occurring data where possible. It would also be interesting to do a comparative study between Emirati Arabic and other varieties of English present in the Emirati land (e.g., Australian, British, American). In this context, it is also important to investigate how compliments and CRs are negotiated in such a setting where people from different L1 backgrounds (e.g., Arabic, American/British English, Hindi, Pakistani) communicate using 'English as an international language' – often beyond the native English speaker authority as the sole norm provider (see, e.g., Rose & Galloway, 2019; Seidlhofer, 2011). This would help identify possible pragmatic transfer students likely to make between the two languages. Another limitation is related to the relatively small sample size, which could be increased in future studies. As

well, the limited number of the hypothetical situations used in this study can be increased in future studies. Neither can the results of the study be generalized to the larger UAE context, not even to the immediate university context in which the study was undertaken. Therefore, future research can include a larger sample size, possibly from different university contexts. It would also be appropriate to investigate university professors' thoughts on students' utilization of CRs. This would help identify how Emirati students can be helped to improve their intercultural communication through language-focused instruction where relevant.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

Ahmad, K. Z. (2014). Gender differences and work-related communication in the UAE: A qualitative study. *International Journal of Business and Management, 9*(3), 185-195. https://doi.org/10.5539/ijbm.v9n3p185

Al Falasi, H. (2007). Just say "thank you": A study of compliment responses. *The Linguistics Journal, 2*(1), 28–42.

Al Harbi, R. S. M. (2017*). Responses of female non-native speakers to English compliments: A cross-generational study of Saudi Arabian university students and lecturers* [Unpublished master's thesis]. Auckland University of Technology.

Al-hindawi, F., & Alkhazaali, M. A. R. (2016). A critique of politeness theories. *Theory and Practice in Language Studies, 6*(8), 1537–1545. https://doi.org/10.17507/tpls.0608.03

Almallah, M. R. A. (2017). *Gender differences in using compliment responses strategies: A case study* [Unpublished doctoral dissertation]. Eastern Mediterranean University.

Almazroui, K. M. (2009). I will write you more letters versus write back: Looking at UAE boys' and girls' written discourse. *INTED2009 Proceedings* (pp. 1452-1477). https://library.iated.org/view/ALMAZROUI2009IWI

AlRousan, M. Y., & Awal, N. M. (2016). Compliment responses among male and female Jordanian university students. *GEMA Online Journal of Language Studies, 16*(1), 19–34.

Alsalem, N. K. (2015). *Compliment responses: A comparison of Saudi English learners and native speakers of American English in an academic environment* [Unpublished master's thesis]. Southern Illinois University.

Austin, J. L. (1962). *How to do things with words*. Harvard University Press.

Barker, K. (1992). Workshop summary: The many faces of motivation in the learning process. *Aspects of Educational and Training Technology Series, 25,* 39–41.

Bataineh, R. F. (2005). American university students' apology strategies: An intercultural analysis of the effect of gender. *Journal of Intercultural Communication, 9*, 1–19. http://dx.doi.org/10.36923/jicc.v5i2.405

Boxer, D. (1993). Social distance and speech behavior: The case of indirect complaints. *Journal of Pragmatics, 19*(2), 103–125. https://doi.org/10.1016/0378-2166(93)90084-3

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Conroy, M. A., Sutherland, K. S., Snyder, A., & AlHendawi, M. (2009). Creating a positive classroom atmosphere: Teachers' use of effective praise and feedback. *Beyond Behavior, 18*(2), 18–26.

Dariela, A., Kepharta, C., Nikiforakisa, & N., Zenkerb, C. (2017). *Emirati women do not shy away from competition: Evidence from a patriarchal society in transition (Working paper 0011)*. New York University Abu Dhabi, Division of Social Science Working Paper Series. https://nyuad.nyu.edu/content/dam/nyuad/academics/divisions/social-science/working -papers/2017/0011.pdf

Deveci, T. (2015). The complaint speech act set produced by university students speaking English as a foreign language. *Literacy Information and Computer Education Journal (LICEJ), Special Issue, 4*(1), 2161–2171. https://doi.org/10.20533/licej.2040.2589.2015.0287

Deveci, T. (2010). The use of complaints in the interlanguage of Turkish EFL learners. *Colombian Applied Linguistics Journal, 12(*2), 25–42. https://doi.org/10.14483/22487085.81

Deveci, T., & Ayish, N. (2021). Qualitative adjectives in EFL students' reflective writing essays. *Journal of Language and Education*, *7*(1), 64-77. https://doi.org/10.17323/jle.2021.10979

Deveci, T., & Hmida, I. B. (2017). The request speech act in emails by Arab university students in the UAE. *Journal of Language and Linguistic Studies, 13*(1), 194–214.

Deveci, T., & Midraj, J. (2021). **"**Can we take a picture with you?" Emirati university students' realization of the refusal speech act with tourists. *Russian Journal of Linguistics*, 25(1), 66–88. https://doi.org/10.22363/2687-0088-2021-25-1-68-88

Deveci, T., & Nunn, R. (2018). Intrapersonal communication as a lifelong learning skill in engineering education. *Journal of Higher Education, 8*(1), 68–77. http://dx.doi.org/10.2399/yod.17.009

DeVito, J. A. (2013). *Human Communication: The basic course* (12th ed). Pearson Education.

Dehkordi, Z. G., & Chalak, A. (2015). English compliment response strategies on social networks by Iranian EFL learners. *Journal of Language Teaching and Research, 6*(2), 452-459. https://doi.org/10.17507/jltr.0602.28

Dendenne, B. (2021). Complimenting on-the-go: Features from colloquial Algerian Arabic. *Journal of Pragmatics, 172,* 270-287. https://doi.org/10.1016/j.pragma.2020.11.013

Dimova, D. S. (2014). *Compliment response strategies in Bulgarian and German* [Unpublished doctoral dissertation]. The Philipps University of Marburg.

Economidou-Kogetsidis, M. (2013). Strategies, modifications and perspective in native speakers' requests: A comparison of WDCT and naturally occurring requests. *Journal of Pragmatics, 53,* 21-38. 7. http://dx.doi.org/10.1016/j.pragma.2013.0.014

Ebadi, S., & Salman, A. R. (2015). Using compliment responses in Arabic and English: Focusing on male and female EFL learners in Iraq. *Journal of Applied Linguistics and Language Research, 2*(7), 157–178.

El-Dakhs, D. A. S. (2017). The compliment response strategies of Egyptian Arabic-English bilinguals. *Language and Dialogue, 7(*3), 387-412. https://doi.org/10.1075/ld.7.3.03sal

Fukada, A. & Asato, N. (2004). Universal politeness theory: Application to the use of Japanese honorifics. *Journal of Pragmatics, 36*(11), 1991–2002. https://doi.org/10.1016/j.pragma.2003.11.006

Furkó, B. P., & Dudás, É. (2012). Gender differences in complimenting strategies with special reference to the compliment response patterns of Hungarian undergraduate students. *Journal of Debreceni Egyetemi Kiadó, Argumentum*, *8,* 136-157.

Gonzalez-Rodriguez, M. F. (2012). *Compliment responses in Icelandic* [Unpublished master's thesis]. The University of Amsterdam.

Guo, H., Zhou, Q., & Chow, D. (2012). A variationist study of compliment responses in Chinese. International *Journal of Applied Linguistics, 22*(3), 347–373. https://doi.org/10.1111/j.1473-4192.2012.00315.x

Halawah, I. (2006). The impact of student-faculty informal interpersonal relationships on intellectual and personal development. *College Student Journal, 40*(3), 670-678.

Hancock, D. R. (2000). Impact of verbal praise on college students' time spent on homework. *The Journal of Educational Research, 93*(6), 384-389. https://doi.org/10.1080/00220670009598733

Hartford, B. S., & Bardovi-Harlig, K. (1992). Experimental and observational data in the study of interlanguage pragmatics. In L. F. Bouton, & Y. Kachru (Eds.), *Pragmatics and language learning* (pp. 33–52). University of Illinois.

Hatch, E. (1992). *Discourse and language education*. Cambridge University Press.

Herbert, R. K. (1986). Say —thank you, or something. *American Speech, 61*(1), 76- 88.

Holmes, J. (1988). Paying compliments: A sex preferential politeness strategy. *Journal of Pragmatics, 12*(4), 445–465. https://doi.org/10.1016/0378-2166(88)90005-7

Jandt, F. E. (2018). *An introduction to intercultural communication*. SAGE.

Kington, A., Sammons, P., Regan, E., Brown, E., Ko, J., & Buckler, S. (2014). *Effective classroom practice*. Open University Press.

Leech, G. (1983). *Principles of pragmatics*. Longman.

Maeshiba, B., Yoshinaga, N., & Kasper, G. (1996). Transfer and proficiency in interlanguage apologizing. In S. M. Gass & J. Neu (Eds.). *Speech acts across cultures: Challenges to communication in a second language* (pp. 155–190). Mouton de Gruyter.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage Publications.

Nelson, G., Al Batal, M., & Echols, E. (1996). Arabic and English responses: Potential for pragmatic failure. *Applied Linguistics, 17*, 411–432. https://doi.org/10.1093/applin/17.4.411

Nouichi, F. (2018). *Comparative pragmatic study of Eastern Algerian Arabic and English speech act of thanking and greeting and responding to them: The case of third-year students of English at the Universities of Mila, Constantine, Guelma, Annaba, and Jijel* [Unpublished doctoral dissertation]. University of Fréres Mentouri.

Ogiermann, E. (2009). *On apologizing in negative and positive politeness cultures*. John Benjamins Publishing Company.

Önalan, O., & Çakır, A. (2018). A comparative study on speech acts: Formal complaints by native speakers and Turkish learners of English. *Eurasian Journal of Applied Linguistics, 4*(2), 239–259. 7. http://dx.doi.org/10.32601/ejal.464128

Pillet-Shor, D. (2012). The problems with praise in parent-teacher interaction. *Communication Monograms, 79*(2), 181–204. https://dx.doi.org/10.1080/03637751.2012.672998

Placencia, M. E., & Eslami, Z. (2020). *Complimenting behaviour and (self-)praise across social media: New contexts and new insights*. John Benjamins Publishing Company.

Pomerantz, A. (1978). Compliment responses: Notes on the co-operation of multiple constraints. In J. Schenkein (Ed.). *Studies in the organization of conversational interaction* (pp. 79–112). Academic Press.

Richards, J. C., Plat, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. Longman.

Rohm, F. W. (2010). American and Arab cultural lenses. *Inner Resources for Leaders, 2*, (3). https://www.regent.edu/acad/global/publications/innerresources/vol2iss3 /rohm.pdf

Rose, H., & Galloway, N. (2019). *Global Englishes for language teaching*. Cambridge University Press. https://dx.doi.org/10.1017/9781316678343

Salameh, A. (2001). *Compliment responses in American English, Saudi Arabic and the English of Saudi EFL learners* [Unpublished doctoral dissertation]. University of Leicester.

Sarkhosh, M. (2022). The role of object distance and gender in Persian compliment responses. *Pragmatics and Society, 13*(2), 272-294. https://doi.org/10.1075/ps.19044.sar

Sartini, N. W. (2019). Compliment response strategy of Balinese women on social media: A cyberpragmatic study. *Advances in Social Science, Education and Humanities Research, 338,* 373-376. https://doi.org/10.2991/prasasti-19.2019.64

Searle, J. R. (1965). What is a speech act? In M. Black (Ed.), *Philosophy in America* (pp. 221–239). Allen and Unwin.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Searle, J. R. (2002). Speech acts, mind, and social reality. In G. Grewendorf, & G. Meggle (Eds.), *Speech acts, mind, and social reality: Discussions with John R. Searle* (pp. 3–6). Kluwer Academic Publishers.

Searle, J., Kiefer, F., & Bierwisch, M. (1980). *Speech act theory and pragmatics*. D. Reidel Publishing Company.

Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.

Swinscom, M. J. (2009). *Statistics at square one* (11th ed.). Wiley-Blackwell.

Taavitsainen, I., & Jucker, H. A. (2008). "Methinks you seem more beautiful than ever": Compliments and gender in the history of English. In A. H. Juckerand, I. Taavitsainen (Eds.), *Speech acts in the history of English* (pp. 195–228). John Benjamins Publishing Company.

Tang, C. (2021). Sociopragmatics of compliment-responding behavior in Taiwanese Mandarin. *Poznan Studies in Contemporary Linguistics, 58*(1). 99-12. https://doi.org/10.1515/psicl-2022-0004

Tang, C. (2020). The interplay of cultural expectation, gender role, and communicative behavior: Evidence from compliment-responding behavior. *Pragmatics and Society, 11*(4), 545-569. https://doi.org/10.1075/ps.16041.chi

Tang, C. H., & Zhang, G. C. (2009). A contrastive study of compliment responses among Australian English and Mandarin Chinese speakers. *Journal of Pragmatics, 41*(2), 325–345. https://doi.org/10.1016/j.pragma.2008.05.019

Tannen, D. (1994). *Gender and discourse*. Oxford University Press.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics, 4*(2), 91–112. https://doi.org/10.1093/applin/4.2.91

Xiang, Z. J. (2013). *A study of gender differences in compliments and responses of Chinese international students* [Unpublished master's thesis]. University of Malaya.

Werkhofer, K. (1992). Traditional and modern review: The social constitution and the power of politeness. In R. J. Watts, S. Ide, & K. Ehlich (Eds.), *Politeness in language: Studies in its history, theory and practice* (pp. 155–199). Mouton de Gruyter.

Yang, L. (2009). The speech act of request: *A comparative study of Chinese and American graduate students at an American university* [Unpublished master's thesis]. Bowling Green State University.

Yu, M.C. (2004). Interlinguistic variation and similarity in second language speech act behavior. *The Modern Language Journal, 88*(1), 102–119. https://doi.org/10.1111/j.0026-7902.2004.00220.x

Ziaei, N. (2012). Translation on the basis of frequency: Compliment and compliment response. *Translation Journal, 16*(3), 4–23.

https://doi.org/10.17323/jle.2023.17569

Building Scientific Knowledge in English: Integrating Content, Cognition and Communication in Secondary School CLIL Biology

Pilar Gerns [1,2]

[1] University of Navarra, Pamplona, Spain
[2] UNIR, Rioja, Spain

ABSTRACT

Background: The focus of this paper is on Dalton-Puffer's construct of the Cognitive Discourse Function (CDF) (2013), which offers CLIL teachers a practical framework through which they can more easily understand the complex idea of integrating the content, cognition, and language required for their subject. These functions have mainly been addressed from classroom observations or task prompts, and little is known about their teachability and effectiveness on students' content knowledge.

Purpose: This paper explores whether the CDF of 'comparing' (a subcategory of 'classify') can be taught to Spanish seventh-grade CLIL biology students (N = 37) and examines the effect of teaching it explicitly on their written performance.

Method: An operational framework was developed to define this CDF and an exploratory study was performed in which students were asked to hand in written comparisons. Quantitative and qualitative pre-and post-tests were applied.

Results: Significant results were obtained for the experimental groups, which improved in both content and language learning, scoring higher on inclusion of content points, justification of their scientific claims, concept formation and use of lexico-grammatical forms.

Conclusion: These findings add to our understanding of the importance of integrating cognition and language in teaching and learning natural sciences, within which CDFs can be a useful starting point.

KEYWORDS

Content and Language Integrated Learning (CLIL), Cognitive Discourse Function (CDF), secondary school science education, knowledge construction

Citation: Gerns P. (2023). 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

Correspondence:
Pilar Gerns,
mgerns@unav.es

Received: July 03, 2023
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

Across Europe and many other parts of the world where English is not the vehicular language, school subjects are now being taught in English. In European countries, this approach is generally understood within the paradigm of Content and Language Integrated Learning (CLIL): when suitable didactic techniques are used, students can learn scientific content at the same time as they improve their English language skills and begin to develop scientific literacy in what is effectively the world language of science. However, the actual process of integrating content and language in CLIL classrooms still remains a challenge (Breeze et al., 2014; Llinares, 2015; Villabona & Cenoz, 2021), not least because CLIL teachers seem to continue giving their classes relying on their previous, traditional didactics, where content and language are taught separately. This suggests that the idea of integrated learning has not yet fully reached the classroom. There is a need to bring existing research into CLIL classroom practice (Llinares et. al., 2012, 2015; Nikula et al., 2016) and provide CLIL teachers with concrete, pedagogical tools they can use to foster integrated learning, that is, co-development of the cognitive, content-oriented demands of the subject

and the language that students need to express their emergent knowledge effectively (Nikula, 2017; Morton, 2020).

Various European research initiatives address the question of integrating content and language learning in classroom practice. These include the Council of Europe subject language descriptors (Vollmer, 2011), the genre-based approach (Llinares et al., 2012; Lorenzo, 2017) and the Graz-group *Pluriliteracy Teaching for Learning* (PTL) model (Meyer et al., 2015; Coyle & Meyer, 2021; Coyle et al., 2023). One promising departure in this area is the construct of the Cognitive Discourse Function (CDF), proposed by Dalton-Puffer (2013), which offers CLIL teachers a practical conceptual framework they can use in their CLIL planning, teaching, and assessment (Morton, 2020).

The novelty of this approach lies in its interdisciplinary nature, which offers content and language teachers a shared pedagogical tool that unites contributions from both disciplines, namely the notions of thinking skills in science and other subject education and the notion of discourse function in applied linguistics (Dalton-Puffer, 2013; Bauer-Marschallinger, 2022). Dalton-Puffer essentially identifies a 'zone of convergence' between content and language learning objectives, recognising that learning in school tends to occur through a limited set of cognitive operations (such as 'explain' or 'define'), which are expressed using a finite set of verbs. The cognitive demands of content courses can thus be linked to specific linguistic patterns, which allow to them be identified and taught (Dalton-Puffer, 2013). The CDF can thus be considered an 'entry point' for students that introduces them to subject-specific ways of thinking and communicating, which are an integral component of subject literacies (Veel, 1997; Rose & Martin, 2012; Zwiers, 2014; Polias, 2015).

In one landmark paper, Dalton-Puffer (2013) presents an operational framework of the most important CDFs *in* and *for* school learning, bringing together 50 different functions from different curricula and learning frameworks, which are then systematized according to their communicative intentions yielding seven main types, as shown in Table 1. These are: categorise, define, describe, evaluate, explain, explore, and report.

The idea is basically that when students are asked to perform a CDF-task, such as to 'explain' the process of photosynthesis, they have to process the subject contents cognitively and verbalize this process through some corresponding linguistic forms. If properly scaffolded, the effort to express the newly learned knowledge in English will foster students' subject learning, leading them over time to more intertwined, deeper learning (Vollmer, 2011; Meyer et al., 2015, 2018; Morton, 2020).

Research on CDFs so far has mainly been conducted from descriptive classroom observations or by eliciting students' CDF use through concrete tasks, analysing whether and how CLIL users acquire these discursive forms without explicitly teaching them (see Lose, 2007; Breeze & Dafouz, 2017; Lorenzo, 2017; Dalton-Puffer et al., 2018; Evnitskaya & Dalton-Puffer, 2020; Doiz & Lasagabaster, 2021; Llinares & Nashaat-Sobhy, 2021; Salvador-Garcia & Chiva-Bartoll, 2022; Whittaker & McCabe, 2023; Llinares & Nikula, 2023). The results of these studies show that CDFs do occur in the course of ordinary CLIL classes, regardless of the school subjects and students' age. However, they also pinpoint the fact that almost no explicit CDF teaching and learning takes place and that, in general, little attention is paid to the functional side of subject-specific language in CLIL contexts. The data present low incidences of explicit metatalk and few output opportunities for students, who thus fail to develop a pro-

Table 1

Cognitive Discourse Functions

Type	Members	Communicative Intention
categorise	categorise, classify, compare, contrast, exemplify, match, structure, subsume	I tell you how we can cut up the world according to certain ideas.
define	define, identify, characterise	I tell you about the extension of this object of specialised knowledge.
describe	describe, label, identify, name, specify	I tell you details of what I can see (also metaphorically).
evaluate	evaluate, argue, judge, take a stance, critique, comment, reflect, judge	I tell you what my position is vis a vis X.
explain	explain, reason, express cause/effect	I tell you about the causes or motives of X.
explore	explore, hypothesise, predict, speculate, guess, estimate, simulate	I tell you something that is potential (i.e., non-factual).
report	report, inform, summarise, recount, narrate, present, relate	I tell you something external to our immediate context on which I have a legitimate knowledge claim.

Note. From *"Cognitive Discourse Functions in Austrian CLIL Lessons,"* by C. Dalton-Puffer, S. Bauer-Marschallinger, K. Brückl, V. Hofmann, J. Hopf, L. Kröss, & L. Lechner, 2018, European Journal of Applied Linguistics, 6(1), pp. 5–29 (https://doi.org/10.1515/eujal-2017-0028). Copyright 2018 by the De Gruer.

ductive command of CDFs. As natural acquisition of these academic functions seems not to occur automatically when teaching CLIL, research has stressed the need for more explicit teaching and practice of CDF forms (Dalton-Puffer, 2007; 2017; Meyer et al., 2018; Morton, 2020; Nashaat- Sobhy, 2020; Coyle et al., 2023).

Some first steps have already been taken in this direction by working on the conceptual and linguistic framework of the different CDFs using predominantly a Systemic Functional Linguistics (SFL) approach to make them operable. Such is the case for 'categorize' (Evnitskaya & Dalton-Puffer, 2020), 'define' (Nashaat- Sobhy, 2020; Llinares & Nashaat- Sobhy, 2021), 'describe' (Dalton-Puffer, 2004), 'evaluate' (Llinares & Dalton-Puffer, 2015; Whittaker & McCabe, 2023; Hasenberger, ongoing), 'explain' (Dalton-Puffer, 2004; Lose, 2007; Connolly, 2019), and 'predict' (a subcategory of 'explore') (Dalton-Puffer, 2007), though only a few have been validated in terms of usefulness by developing explicit teaching programmes. Work in this direction includes a study by Breeze and Gerns (2019) in a secondary-level CLIL history class, where some general CDF-related features were explicitly taught for 'describing' and 'explaining', four doctoral theses (Connolly, 2019; Gerns, 2021; Bauer-Marschallinger, 2022; Hasenberger, ongoing) and Nashaat-Sobhy's (2020) validated framework for working with the CDF 'defining' in the university setting. Classroom materials are now being developed to teach, scaffold and assess these CDFs in the CLIL classroom[1] (see Coetzee-Lachmann, 2019; DeBoer & Leontjev, 2020; Del Pozo & Llinares, 2021; Coyle et al., 2023). However, more research is clearly needed to clarify the cognitive and linguistic parameters of the various CDFs across school subjects and test their practical applicability, in particular, in secondary L2 education, where the students' content, cognitive and language skills are undergoing intense development.

This study aims to contribute to this ongoing research by first defining the CDF of 'comparing' (a subcategory of 'categorise') in the context of science, where it constitutes a fundamental method and offers a good starting point to introduce students to scientific working culture and its specific modes of thinking and communicating, which are a key component of scientific literacy (Lemke, 1990; Roberts & Bybee, 2014). By enabling students to articulate their own scientific knowledge in a much clearer, more organised and professional way, teachers thus empower them to develop the necessary cognitive and linguistic skills to take part in society's wider scientific debates in and beyond school (Mercer, 1995; Norris & Phillips, 2009; Oliveira & Weinburgh, 2017; Coyle & Meyer, 2021).

Second, this paper aims to examine the effect of explicitly teaching the CDF of 'comparing' on students' subject-based written performance in a secondary school (7 grade) CLIL science class. To this end, some quantitative and qualitative pre-and post-test results will be shown from an empirical study in which the CDF of 'comparing' was explicitly taught to a group of Spanish CLIL biology students.

Defining the CDF of 'Comparing' in CLIL

To determine whether explicit teaching of the CDF of comparing promotes integrated learning between contents and language, an interdisciplinary approach was adopted, which integrates contributions made by the natural, educational, and linguistic sciences. This is briefly presented below.

'Comparing' in Natural Sciences

The act of comparing constitutes one of our first and most natural forms of human thought. From childhood onwards, we interact with the world by making comparisons, grouping things into pairs, and examining their shared and unshared features, that is, their similarities and differences. These comparison-based discoveries are nothing other than our epistemic way of making the world understandable and operational.

However, comparing does not only present a natural human way of thinking – it is also a fundamental method of scientific inquiry, integral to our modern sciences (Darian, 2003; Martinez, 2018). The idea of studying the living word through a comparative approach dates back to the Ancient Greeks, when Aristotle was one of the first to analyse the physiology of dissected animals and plants by identifying patterns of similarity and difference (Carpi & Egger, 2008). This method has developed since and constitutes an essential, distinctive methodology for biological science, being used to classify plants and animals into taxonomies, or to investigate comparative anatomy in the study of evolution, heredity, and adaptation to environmental pressures (Flannery, 2010). It is thus a core cognitive skill in biology, and an essential prerequisite for disciplinary literacy. In fact, comparing is frequently used in biology classes, for example, when viruses are taught together with bacteria, or the concept of warm-blooded with that of cold-blooded. Teaching students how to compare can be a good starting point to introduce them to comparative methods and help them think like scientists.

'Comparing' in Education

In education and cognitive psychology, comparing is understood to be an important component in terms of lower-or-

[1] LongAd-CLIL project (2019-2021). A longitudinal corpus-based analysis of advancedness in CLIL. https://uam-CLIL.org/longad-CLIL/

der thinking skills (Bloom, 1956). It is defined as one of the most basic analytical lower-order thinking skills, grouped together with others (such as 'matching', 'classifying', 'error analysis', 'generalizing' and 'specifying'), which are essential to carry out more advanced and complex operations (such as 'define', 'predict', or 'classify'). Furthermore, different taxonomies (see Anderson & Krathwohl, 2001; Marzano, 2001; Biggs & Tang, 2011) situate comparing at the crossroads between the cognitive dimensions of 'understand' and 'analyse' and the knowledge dimension of concepts. The reason is that comparing consists of identifying commonalities between at least two items, which often requires moving from a surface, experiential-driven understanding towards a more abstract and deconstructive one.

Studies have shown that teaching how to compare properly in content areas seems to have positive effects on students' learning (Hammann & Stevens, 2003; Silver, 2007; Goldstone, 2010; MacArthur & Philippakos, 2010; Clark et al., 2020). For example, it helps to develop: (1) a structured and orderly way of thinking, (2) abstract and relational ways of learning, (3) detail-focused, and longer-lasting comprehension, since two things taught together seem to prove more memorable than one, and ultimately (4) advanced literacy skills.

'Comparing' in Linguistics

According to the Oxford English Dictionary (OED) (1989), the verb 'compare' is a late Middle English word, derived from the old French word 'comparer' and the Latin 'compara-re' ('com'- with + 'par'- equal; 'compar'-like), meaning literate to 'pair' or 'match together'. The transitive verb 'to compare' and its corresponding noun form 'comparison' can be defined in either one of the following ways:

(1) 'to speak of or represent as similar'; 'to make like'; 'to liken'.

(2) 'to place together so as to point out the similarities and differences of (two or more things)'.

Based on these definitions, 'comparing' seems to be characterized by 'matching' or 'bringing' at least two relatable items together, that is, 'likening' them, and analysing them on some shared and unshared features, i.e., on their similarities and differences.

Academic discourse functions have been studied in the context of L2 teaching, in particular, by researchers from English for Specific Purposes (ESP) and Halliday's Systemic Functional Linguistics (SFL) tradition (Halliday & Matthiessen, 2004), which address the importance of such functions in academic writing (Polias, 2015; Smith, 2019). 'Compare' is considered one of the most common discourse functions in expository text, alongside others such as sequencing, classification, explanation, cause-effect, and problem-solution (De La Paz & McCutchen, 2010). This can also be regarded as an aspect of Cognitive Academic Language Proficiency (CALP) (Cummins, 2008), which is crucial for students' academic development (Llinares et al., 2012; Morton, 2020).

It is clear that cognitive functions (such as 'comparing') are tied to specific lexico-grammatical structures (such as 'is like', 'different to' or 'unlike') which allow teaching and learning through them. Considerable research has been conducted to clarify the analytical patterns of these discourse functions (see Cheong, 1978; Widdowson, 1979; Darian, 2003; Dixon; 2005; Huddleston, 2017; Evnitskaya & Dalton-Puffer, 2020).

Five-Point Framework for 'Comparing'

Based on the previous literature, I developed an operational framework that brings these contributions together. As a result, the act of comparing can be defined through at least five elements, shown in Table 2, each of which can be triggered by a corresponding question. This framework relies, in particular, on the work of Raphael and Kirschner (1985),

Table 2

Five-Point Framework for the CDF of 'Comparing'

Components	Questions
Items being compared: "topic" (X) + "target" (Y)	What two items are compared? What is the "topic" and what is the "target"?
Points of comparison: "criteria"	What (points) can these items be compared on? Can you name the criteria explicitly? Can you use the same criteria for both groups?
Types of comparisons: general, specific comparison; similarities, differences.	What type of comparison will be performed? Are the items being compared 1. in a general/ specific way or 2. in terms of similarities and differences. Are these points reasoned?
Language *of* and *for* comparing: "comparitor"	How do we express a comparative relationship? What lexico-grammatical forms are used?
Comparative structure: parallel organization	How do we structure the comparative points? In a block-method, or a point-by-point one?

Widdowson (1979), Cheong (1978), Darian (2003), Dixon (2005), Polias (2015) and Huddleston (2017), as they present some concrete analytical descriptions of how to make a scientific comparison.

The first element addresses the question of what two items will be compared, referred to as 'topic' and 'target' (Cheong, 1978). The second element consists of establishing the criteria on which these items can be compared on, which are variously known as the 'parameter' or 'basis' of a comparison (Widdowson, 1979; Dixon, 2005). These points are usually abstract concepts. Then, the third element refers to the conceptual categories a comparison can adopt, which include Halliday and Hasan's (1976) general and specific comparison and a comparison in terms of similarities, differences and matters of degree (Cheong, 1978; Darian, 2003; Halliday & Matthiessen, 2014; Polias, 2015). Fourth is the language *of* and *for* comparing, which are the specific lexico-grammatical forms and structures (such as 'similar to', 'different') through which comparative relationships can be expressed, known as the comparitor. Significant research has been done in ESP and SFL to clarify these forms. Finally, a comparison is also characterised by its parallel structure, according to which items can be compared and contrasted in a systematic way. See Gerns (2023) for further analysis of these aspects.

Exploratory Classroom Study

Rationale and Research Questions

Given the importance of validating explicit CDF-based teaching tools and verifying their effect not only on students' language but also on their content and cognitive learning (Badertscher & Bieri, 2009; Bauer-Marschallinger, 2022), the present paper aims to validate the CDF framework for comparing and establish whether explicitly teaching it in a CLIL science context can have a positive effect on students' subject knowledge. To do this, a classroom-based study was designed and executed with secondary CLIL biology students (7th grade, age 12-13), who were previously divided into a study and a control group. They were asked to hand in three written comparisons, which served as pre-and post-tests. The idea was to compare the results obtained before and after the instruction to examine whether the study group students displayed their content knowledge significantly better after the instruction phase than the control group.

The following research questions were addressed:

Can explicit CDF-based teaching (CDF 'compare') have a significant effect on students' written subject performance, considering:

(1) the number of content points (inclusion of differences and similarities)?

(2) the number of reasoned content points (using linguistic devices, i.e., explicatory or defining forms, to support scientific claims)?

(3) the number of content points related to comparative concepts (criteria) (looking at nominalisations, abstract nouns, topic sentences and superordinated concepts)?

(4) the number of lexico-grammatical comparative forms (comparitors)?

These questions align with the framework's core points for comparing (points 2 to 4). The study thus investigates how students co-develop the cognitive, content-oriented and linguistic skills needed to express their emergent knowledge more effectively.

METHOD

Context and Participants

An exploratory study was conducted with 37 lower-secondary (7grade) CLIL biology students (aged 12-13 years), their CLIL biology teacher and their EFL teacher in a charter-like, middle-class school in the north of Spain (Navarra) for three months (January-March 2019). It was decided to work with lower-secondary students, who had just moved from primary to secondary education as this is a decisive stage in students' formal education. The participating school offered biology in three seventh-grade classes with a CLIL modality of 50%, teaching half of its students (approximately 15 per class) in English, who were all included in the study, and half in their L1 (Spanish). These small groups allowed a more personalized teaching approach.

It is a small sample, and cannot be claimed to be representative, but it does offer initial insights into how a group of secondary students learned their science content through a CDF approach. The participants' mother tongue was Spanish (L1), and English was their foreign language (L2), except for three bilingual students (2 Spanish/Chinese and 1 Spanish/Polish). The CLIL biology and EFL teacher ranked the students' English level at a low B1 according to the CEFR.

Design and Method

A class unit was designed in collaboration with the CLIL biology and EFL teacher on the curricular topic of biodiversity, in which the CDF of comparing was integrated into the ordinary science lesson plan, that is, into the class tasks, oral interactions, and materials, providing the students with some CDF-rich lessons. The idea was to integrate the CDF of 'comparing' in a contextualized way so as not to interrupt the regular course of the class or make it look like a subject-ex-

ternal element (see Breeze & Gerns, 2019; Nashaat-Sobhy, 2020; Bauer-Marschallinger, 2022).

We used a pre-and post-test design to measure the possible effects of instruction. Of the three available groups taught by the same CLIL biology and EFL teacher, two were chosen as study groups (S1, S2), and one as control group (CG). The study groups received explicit CDF-based input and output opportunities (see Appendices A-H), while the control group received some lessons only focusing on content aspects without explicit indications or support dedicated to the CDF of 'comparing'; however, as the teacher included comparative expressions in their classroom talk, and as the written texts (see Appendix E) had some comparative lexico-grammatical and structural features, their classes could be regarded as representing an implicit CDF approach.

Intervention

The intervention included several classes during the first two months of the project (January/February 2019), where the teacher progressively explained the five points of the comparing framework. The first step involved raising students' awareness of the role that comparing plays in science (biology) education and jointly defining what makes a good scientific comparison (see Appendix A). Second, work was done on the first three points of the framework (*items being compared*, *comparative concepts* (criteria) and *types of comparison*) practising it with different scientific topics (see Appendix B). In a third step, the language *of* and *for* comparing (*comparitors*) was introduced, encouraging students to formulate some comparative, content-based statements; here language support was aligned with students' scientific knowledge (see Appendices C and D). Last, some structural devices were introduced, such as parallel and paragraph organisation, and the previous points were revised by deconstructing jointly a written comparison (consult text and tasks in Appendices E-G). As in other studies, a step-wise, joint, in-class learning approach was chosen (see Lose, 2007; Bauer-Marschallinger, 2022; Nashaat-Sobhy, 2020), based on the learning cycle idea from the genre tradition (Rose & Martin, 2012), Cummins's four learning quadrants (2008) from the ESL tradition and Vygotsky's (1934) 'collaborative learning' approach, which recommends moving from more supportive to more autonomous learning. Thus, students' receptive skills were first activated, followed by their deconstructive and productive ones. The idea was to draw students' attention to the different dimensions that converge when making a comparison, namely the conceptual, cognitive, and linguistic aspects.

Data Collection

To elicit comparisons, we asked the students to hand in three written comparisons: one before and two after the instruction classes, allowing us to trace better students' learning progress and detect any improvement over time. These served as pre- and post-tests (see Table 3). For the pre-test, the students had to compare between herbivores and carnivores, without receiving any kind of guidance or support. Then, during the instruction, the students were asked to rewrite their first written comparison in order to put what they had learned so far during these lessons into practice (see Appendix H). In the post-instruction phase, the students were asked to hand in two more comparisons to examine their ability to process and present new biological content in a comparative way. This time, they had to compare vertebrates with invertebrates and mammals with reptiles. All three tasks followed a similar design based on the previous analysis of 'comparing' and were reviewed by the participating teachers regarding clarity, content authenticity, curricular demands, and level of difficulty.

Data Analysis

The results were measured in quantitative terms by first counting the number of forms used by the individual students in their writings, and then by calculating average for the groups (S1, S2, CG) in each of the three task collection phases (pre-and post-tests). A paired t-test was run with a p-value at 0.05 (Rasinger, 2008), which indicates whether the students' progress was statistically significant. As the sample is small, caution is recommended. The idea is not to present large-scale representative data but to offer first insights into students' CDF content performance and draw some general conclusions on the effects of explicit CDF-teaching.

RESULTS

Table 3

Study Design

CDF-based tasks	Purpose	Schedule
1. compare herbivores and carnivores.	Pre-test	January 2019
(rewrite first comparison)	Instruction	January/February 2019
2. compare vertebrates and invertebrates.	Post-test (1)	Beginning of March 2019
3. compare reptiles and mammals.	Post-test (2)	End of March 2019

The quantitative results are presented in the following, and a qualitative pre- and post-test example is provided to illustrate students' content performance better.

Inclusion of Content Points

Starting with the results concerning the inclusion of content points (differences and similarities), the students could present up to 7 points per task. The researcher counted the number of task-relevant points presented by each student, identifying 564 points of differences and 155 of similarities, and calculated the average number per group and phase. The points of difference and similarity were analysed separately.

On the one hand, when considering differences, as shown in Figure 1 and Table 4, all three groups learned to include more points from their pre- to their respective post-tests. In the pre-tests the students included an average of 4 points, while in their respective post-tests they included an average of 6 points. Based on the paired t-test results (see Table 4), the gains were highly significant for all groups ($p < 0.05$) except the control group, who improved significantly only when post-test 2 was considered.

On the other hand, regarding students' pre- and post-test inclusion of similarity content points (see Figure 2 and Table 5), at the beginning, the students hardly presented any similarities, whereas after the instructional phase most study-group students (S1, S2) included them, especially in their second post-tests, where they doubled their use (including 2 to 3 forms). This can be considered a significant, delayed improvement ($p<0.05$) (see Table 5). In contrast, the control group's use of similarities declined.

Figure 1

Mean Number of Content Points (Differences) Used per Student

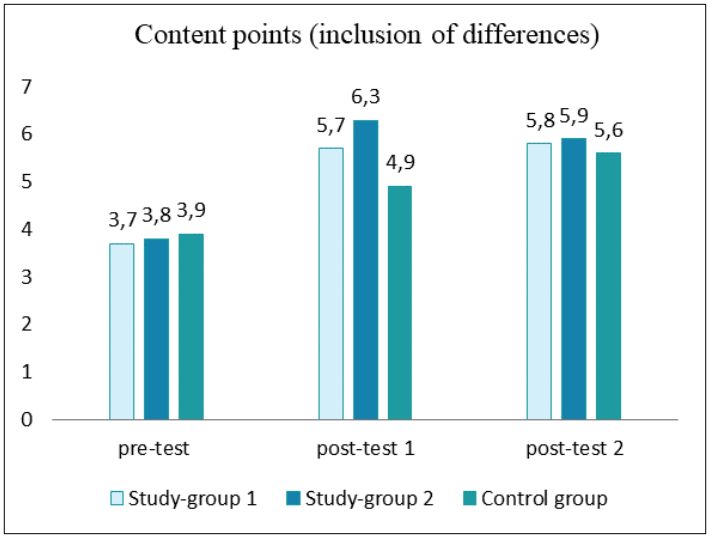


Table 4

Paired T-Test: Mean Number of Content Points (Differences) Used between Pre-Test and Post-Tests

Study Variable	Groups	Pre-Test		Post-Test 1		Post-Test 2		P-Value
		Mean	(Sd)	Mean	(Sd)	Mean	(Sd)	
Mean number of content points (differences)	S1	3.69	(1.58)	5.69	(1.35)			< 0.0007 **
	S2	3.83	(2.21)	6.25	(2.22)			0.0184 *
	CG	3.89	(1.27)	4.89	(2.67)			0.4070
	S1	3.69	(1.58)			5.75	(1.61)	0.0010**
	S2	3.83	(2.21)			5.92	(2.15)	0.0319*
	CG	3.89	(1.27)			5.56	(1.24)	0.0167*

Note. p<*=significance at 0.05; p<**=significant at 0.001

Figure 2

Mean Number of Content Points (Similarities) Used per Student

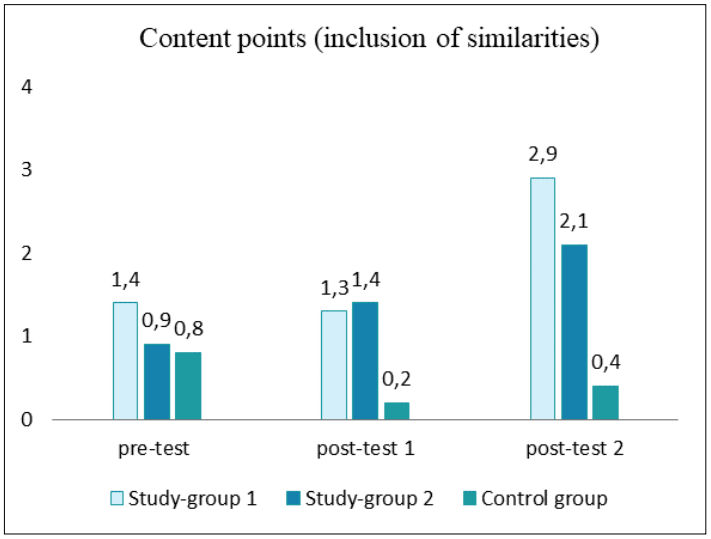


Table 5

Paired T-Test: Mean Number Content Points (Similarities) Used between Pre-Test and Post-Tests

Study Variable	Groups	Pre-Test		Post-Test 1		Post-Test 2		P-Value
		Mean	(Sd)	Mean	(Sd)	Mean	(Sd)	
Mean number of content points (similarities)	S1	1.38	(0.89)	1.31	(1.20)			0.8172
	S2	0.92	(1.31)	1.42	(1.62)			0.2562
	CG	0.78	(0.67)	0.22	(0.44)			0.0509
	S1	1.38	(0.89)			2.94	(1.44)	0.0024 **
	S2	0.92	(1.31)			2.08	(1.38)	0.0413*
	CG	0.78	(0.67)			0.44	(0.73)	0.3466

The following pre and post-test extract (Figure 3, examples 1 and 2) from one study-group student offers a qualitative insight into how most study-group students learned to present the content points in a more integrated, complete, and subject-purposeful way.

As can be seen, in the pre-test (example 1) only one similarity point (*both have a skull*) is included, which offers little relevant information and is presented separately from the other content points. In the post-test (example 2), the point of similarity offers more task-relevant information and is more integrated within the text. The importance of identifying similarities is recognised as a way of establishing class-memberships, seen by the assertion that herbivores and carnivores share the feature of having a backbone to classify them as vertebrates.

Furthermore, most pre-tests presented the differences in a vague and incomplete way, that is, specifying information about one animal group, while omitting this information for the other. Here, the student explains canines in carnivores but fails to provide the corresponding information about herbivores, indicating, for example, that they have a horny pad instead (*herbivores don't have canines and the carnivores have great canines*). Something similar can be seen later in example 5 (*The carnivores have canins and the herbivores not. The herbivores had large incissors to shop grass and canines no*). However, in the post-test most students, like the present one, learned to specify the content points more and analyse the information equally in both groups.

Elaboration of Content Points

Finally, we examined whether the students simply listed or elaborated the content points by providing additional information. The number of content points that were further reasoned, through the use of different linguistic resources (such as synonyms, explanatory/exemplifying forms, conditionals, action verbs, consecutive, relative clauses, etc.), was

Figure 3

Pre- and Post-Test Example: Presenting Content Points (Differences and Similarities)

(1) **PRE-TEST**	(2) **POST-TEST**
Different The herbivores eat plants and the carnivores eat meat (difference) also the herbivores don't have canines and the carnivores have great canines. (difference). Herbivores have broad incisors and the carnivores have pointed incisors. (difference). Common They both have a skull (similarity). (8AM)	Now, I am going to say the differences and the things in common of reptiles and mammals. The first thing in common they have is that both have a backbone; so they are vertebrates (similarity). Now, some differences, reptiles are cold-blooded (ectothermic), on the contrary mammals are warm-blooded (endothermic). (...) (difference) (8AM)

Figure 4

Mean Percentages of Elaborated/Reasoned Content Points per Student

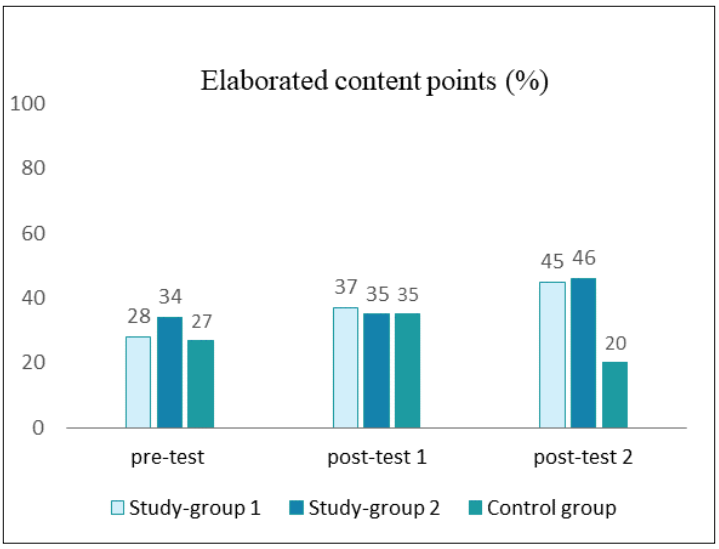


Table 6

Paired T-Test: Mean Value of Elaborated Content Points Between Pre-Test and Post-Tests

Study variable	Groups	Pre-Test		Post-Test 1		Post-Test 2		p-value
		Mean	(SD)	Mean	(SD)	Mean	(SD)	
Mean value of content points related to comparative concepts	S1	1.38	(1.54)	2.56	(2.53)			0.1308
	S2	1.58	(1.56)	2.67	(2.61)			0.2231
	CG	1.22	(1.30)	1.78	(1.56)			0.3251
	S1	1.38	(1.54)			3.50	(2.68)	0.0199*
	S2	1.58	(1.56)			4.17	(2.17)	0.0040**
	CG	1.22	(1.30)			1.69	(0.56)	0.8771

identified and as before, means were calculated in relation to the total number of content points.

Figure 4 and Table 6 show the results of this analysis. When contrasting the pre-test results of all three groups with their respective two post-tests, we see that all three groups improved slightly in developing their content points more. Initially, in the pre-test, the students developed an average of 26-34% of their content points. Then, in the first post-test, these data improved slightly for all, and finally, in the second post-test, the study group learners continued improving significantly on their initial results (see Table 6), whereas the control group decreased the amount of support provided for their content points.

In qualitative terms, the pre-and post-test extracts in examples 3 and 4 (Figure 5) show how especially the study group students learned to elaborate their content points more after the instruction phase.

As the pre-test (example 3) shows, most students presented content points without developing them further or connecting them with theoretical issues. By contrast, in the post-test (example 4), most study-group students improved considerably, clarifying their content points by presenting definitory and explicatory forms. Here, the student explains two biological concepts (cold-blooded and warm-blooded animals) by defining them, using several linguistic resources, such as explanatory and exemplifying forms ('this means'; 'for example'), conditional-if, a consecutive form ('as a result of') and an action verb ('to keep'). The underlying relationship is shown between the different contents (the blood groups, heat regulation mechanism, the environment and eating habits) and there is an attempt to derive a general scientific principle that reptiles are found in warmer climates due to their thermoregulation. This reveals a more complex and interrelated understanding of the contents and of typical scientific ways of proceeding (Schalk et al., 2016).

Figure 5

Pre- and Post-Test Example: Reasoning/Elaborating Students' Comparative Claims

(3) **PRE-TEST**	(4) **POST-TEST**
(…) Herbivores need to eat a lot but carnivores don't eat everyday. Herbivores stomach is large and has many chambers, but carnivores stomach is simple. Herbivores intestine is big and very long but carnivores intestine is short and small (…) (2CF)	(…) The main difference between the two groups is that reptiles are cold-blooded and mammals are warm-blooded. This means that reptiles' body temperature depends on their environment. If it is cold, their bodies are cold too. However, mammals can regulate their heat by themselves. For example, they get cooler through sweat. As a result of this difference, reptiles normally live in warm places, but mammals can live in any environment. Reptiles need less food or energy than mammals to keep their bodies warn. (…) (2CF)

Figure 6

Mean Percentages of Content Points Related to Comparative Concepts per Student

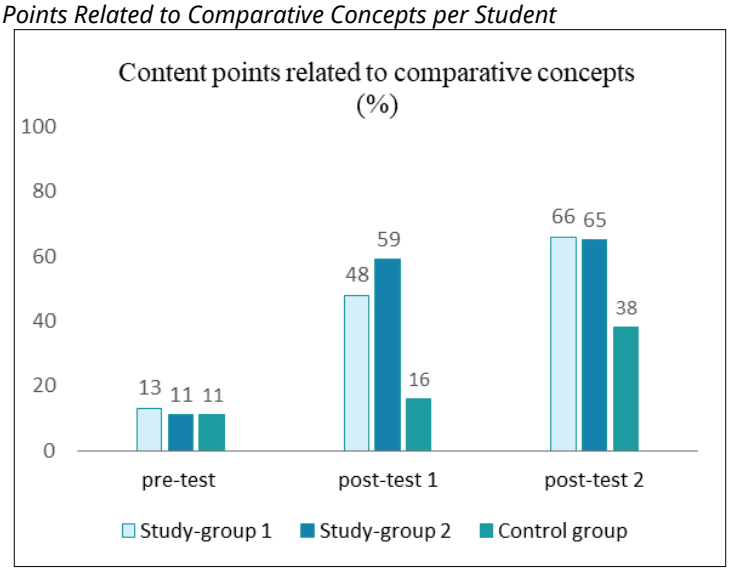


Table 7

Paired T-Test: Mean Value of Content Points Related to Comparative Concepts Between Pre-Test and Post-Tests

Study Variable	Groups	Pre-Test		Post-Test 1		Post-Test 2		P-Value
		Mean	(Sd)	Mean	(Sd)	Mean	(Sd)	
Mean value of elaborated content points	S1	0.50	(0.82)	2.75	(2.44)			0.0051 *
	S2	0.42	(0.90)	3.67	(2.84)			0.0059 *
	CG	0.44	(0.88)	0.67	(0.87)			0.4468
	S1	0.50	(0.82)			3.81	(2.76)	0.0001***
	S2	0.42	(0.90)			3.83	(2.72)	0.0009**
	CG	0.44	(0.88)			1.89	(1.76)	0.0563

Figure 7

Pre- and Post-Test Example: Using Comparative Concepts (Criteria)

(5) **PRE-TEST**	(6) **POST-TEST**
The herbivore animals eat plants and the carnivores eat meat, fish and eggs for example. The carnivores have canins and the herbivores not. The herbivores had large incissors to chop grass and carnivores no. (11AF) **(implicit criteria)**	First I'm going to compare the **legs (explicit criterion)** that they (reptiles and mammals) have. They can be legless or tetrapodal. Legless mean that they haven't got legs and tetrapodal means that they have 4 legs. Reptiles can be legless or tetrapodal. However mammals are tetrapodal, they have 4 legs. Second I'm going to explain how is their **habitat (explicit criterion)**. Both can live in a terrestrial and freshwater ecosystem. In marine ecosystem only can live reptiles. (11AF)

Content Points related to Comparative Concepts (Criteria)

Third, students' writings were examined on their ability to form comparative concepts (criteria) and base their search for similarities and differences on them. To measure this abstract ability, the number of content points that were explicitly related to a comparative concept using nominalisations, abstract nouns or main idea topic sentences was counted per student and phase, and averages were calculated in relation to the total number of content points (previously measured).

As Figure 6 shows, all post-test results from the three groups indicate a significant improvement (especially the second post-test). In the pre-test, the students only subordinated and introduced a few content points (about 12%), while in the two post-tests, this changed significantly, as the two study groups subordinated 35-54% more forms, and the control group learned to subordinate between 5 to 27% more forms. According to the paired t-test results (see Table 7), this improvement was highly significant for the study-group students (S1, S2) in their two respective post-tests ($p<0.05$), while for the control group, the improvement failed to reach significance.

From a qualitative perspective, the following pre- and post-test extracts from a study group student (Figure 7, examples 5 and 6), show the improvement process many study-group students went through, as they learned to relate diffuse content points under some corresponding superordinated concepts (criteria).

In the pre-test (example 5) two animal groups (herbivores, carnivores) are compared on their diet and teeth, but at no point are these two categories explicitly referred to, whereas in the post-test (example 6) the student explicitly names the comparative concepts in a topic sentence ('First I'm going to compare the legs'; 'Second I'm going to explain how is their habitat') and structures his/her comparative analysis on it.

These two examples show that explicit mention of the comparative concept (criterion) in a topic sentence is not an indispensable element when making a comparison, as both the pre-test (example 5) and post-test (example 6) are two valid comparative statements. However, it does show that

Figure 8

Mean Number of Lexico-Grammatical Comparative Forms Used per Student

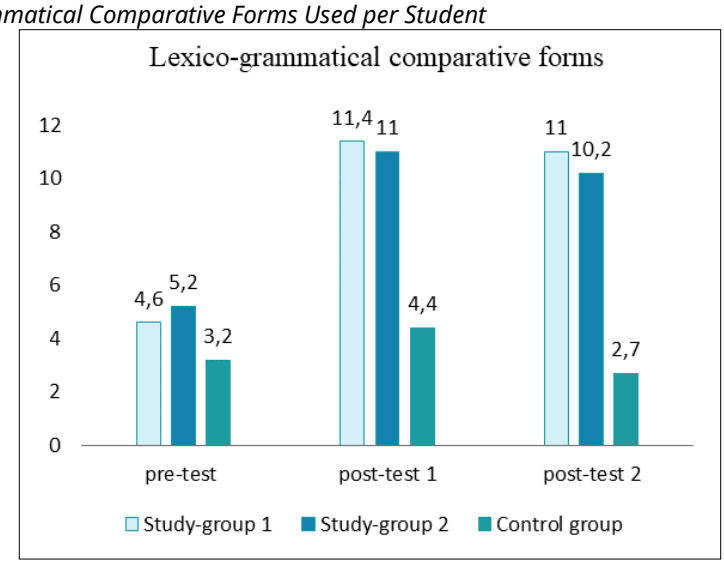


Table 8

Paired T-Test: Mean Number of Lexico-Grammatical Comparative Forms Used between Pre-Test and Post-Tests

Study Variable	Groups	Pre-Test		Post-Test 1		Post-Test 2		P-Value
		Mean	(Sd)	Mean	(Sd)	Mean	(Sd)	
Mean number of lexico-grammatical comparative forms	S1	4.94	(3.21)	11.38	(5.82)			0.0006 **
	S2	4.92	(3.68)	11	(7.38)			0.0145*
	CG	3.56	(2.01)	4.44	(2.79)			0.2495
	S1	4.94	(3.21)			10.81	(5.56)	0.0005***
	S2	4.92	(3.68)			11	(8.05)	0.00981*
	CG	3.56	(2.01)			2.78	(2.17)	0.4097

the comparison must have a common comparative point, because otherwise, it would not be a comparison but a description.

Use of Comparative Language (Comparitor)

In the second step, students' writings were analysed in terms of lexico-grammatical comparative forms, the 'comparitors' (Darian, 2003). There are three main types: contrast forms (such as 'is different', 'unlike'), comparison forms (like 'similar', 'equal to'), and grammatical comparative forms (such as 'less/more than'). In the analysis, all linguistic attempts that somehow indicated a comparative relation were counted, including forms with linguistic errors, as long as their communicative intention was understandable. Average numbers per student, group and phase were calculated. A total of 848 linguistic comparative forms were identified in the whole sample, among which 458 were contrasts, 216 were comparisons, and 174 were grammatical comparative constructions. These numbers can be explained by the fact that most of the content points presented were differences and there were few similarities.

As Figure 8 shows, the study group students learned to include considerably more lexico-grammatical forms after the instruction classes, progressing from using initially an average of 4 to 5 forms per student to including around 11 in their post-tests, which is a significant improvement ($p<0.05$) (see Table 8). On the other hand, the control group students also experienced a slight increase between their first and second writing (passing from 3 to 4 forms), but it quickly declined in their second post-test, where they returned to their initial starting point. It should be noted that from the beginning all groups already used some kind of linguistic comparative forms, which may suggest a certain prior familiarization with these forms acquired in their EFL classes.

The following pre- and post-tests (Figure 9, examples 7 and 8), illustrate how most study group students moved from an initial low and basic use of lexico-grammatical expressions (such as 'common'; 'both', 'different', 'but') towards a more

Figure 9

Pre- and Post-Test Example: Using Lexico-Grammatical Comparative Forms (Comparitor)

(7) **PRE-TEST**	(8) **POST-TEST**
Common: • Carnivore animals and herbivore animals, have a skull that protect their brain. • **Both** have a jaw that used to eat. **Different**: • Carnivores have: Incisors, Canines, Premolars and molars **that** used to eat. • Herbivores have: at the beginning they not have teeth, the have large chewing molars and premolars. (10CM)	**I'm gone a talk about** the physical of **both** groups. **Both** have a skin, reptiles' skin is covered by scales, **while** mammals' skin is covered by hair **that is** beneficial **for** heat themselves, **for advertising. Both** kinds of animals are tetrapodal **that means that** they have four legs **but** not always, **some** mammals **like** dolphins have flippers and **some** reptiles **like** snakes have not any legs, **so** they are legless. **So** leaving behind the physical **now can I talk about** their reproduction. **I'm sure** about you know the meaning of oviparous and viviparous. **Well in one hand** mammals are viviparous **that** they born from their mammas body and they are nursed with milk from mother glands, **but in the other hand** reptiles are viviparous, they lay eggs in the land. (10CM)

varied and combined use of advanced forms (such as 'while', 'whereas', 'on the one hand… on the other'), making their comparative claims more explicit.

The students had no great difficulties in picking up and using these forms; they probably knew them from their EFL class and they seemed to understand their functional purpose as they used them in the right place (to state a contrast and/or similarity). However, they had difficulties using them correctly, that is without making lexico-grammatical or orthographic mistakes. For example, they confused comparatives with superlatives, forgot about the suffix 'than' when using comparative grammatical forms or used incorrect word order or prepositions, which suggests inadequate grammatical, lexical and orthographic knowledge or perhaps lack of attention due to cognitive overload.

Moreover, as shown in example 8, the students were able to combine these forms with other linguistic devices, which had not been explicitly taught, which were probably transferred from their L1 and/or EFL classes. These include, consequence forms ('so'), the English possessive 's', explicative and defining forms (such as 'that means that', 'that is', 'like'), quantifiers ('some', percentages), and they developed some awareness to the need to guide readers and adopt a writerly stance, using future tenses, direct pronouns ('I', 'you'), discourse markers ('well', 'so'), temporal features ('now') and referring to a shared background knowledge ('I am sure about you know the meaning of …'). Their post-tests thus show greater literacy development by decentring (take distance) from the text and establishing a dialogue with the reader.

DISCUSSION

This paper has examined the teachability and effect of explicitly teaching CDFs in CLIL science.

As for the teachability of CDFs, the CLIL science and the English teacher had no difficulty in co-designing and delivering science lessons that integrated the CDF of *comparing*. They quickly understood that language not only refers to English (as L2) but to the language demands specific to science, which goes beyond learning a set of technical terms, but includes using abstract nouns and topic sentences to form concepts, supporting scientific claims through linguistic devices and using discursive structures and specific lexical-grammatical forms to express comparative relationships (Lemke, 1990; Mortimer & Scott, 2003; Evnitskaya & Morton; 2011; Rose & Martin, 2012; Nikula, 2017). A functional rather than formal approach was adopted to the study of language and content, where clarity (communicative intention) was favoured over form. That is why little attention was paid to grammatical and orthographic errors, so as not to overload students with corrections and avoid turning the class into a language class.

The intervention showed that teaching students how to compare from a scientific perspective and in a progressive, explicit and interactive way was extremely useful. A gradual learning path was chosen, which started by first co-defining the importance of *comparing* for learning natural sciences, then moved on to practise some of its constituent components, to review lexico-grammatical forms for expressing the emerging scientific knowledge better and finally to ad-

dress textual organization in paragraphs and parallel structures. The language and cognitive support were only introduced in so far as they aided subject learning.

As for the effect of explicitly teaching CDFs on students' subject performance, significant results were obtained for the experimental groups, which improved in both content and language learning, when looking at the inclusion of content points, justification of scientific claims, concept formation and use of lexico-grammatical forms.

Based on the results, the study-group students moved from an initial, weak understanding of *comparing* towards a more proficient one as students learned to work harder on content and language. They included a greater number not only of differences but also of similarities and provided more specific, and task-relevant information. They also explained and linked their initially diffuse content points more, relying on linguistic devices (such as linking words, defining and explicatory forms) and clarifying the relationship between content concepts (i.e., blood type, environment and temperature), showing a clearer subject understanding. Moreover, the students also learned to structure their claims and make them more explicit as they related the contents to the superordinate concepts, introduced them with topic sentences and signalled comparative relationships using appropriate lexico-grammatical forms, indicating a more advanced linguistic command and proficiency in abstract reasoning. By and large, the explicit instruction helped students reach better scientific achievement levels as they developed more scientific concepts, practised scientific thinking (i.e., using similarities to establish class membership) and produced more sophisticated writing about science. The students understood that *comparing* in science is an inherent part of studying science, requiring certain cognitive and linguistic forms. The process of writing in science class, which is not frequently used, undoubtedly supported the students' learning process, helping them to plan, organise and visualise their scientific understanding in a linear and logical way (Schleppegrell & Colombi, 2005; Oliveira & Weinburgh, 2017).

Most of the pre-test results coincide with the findings made by other studies, which observed students' natural - that is, not-explicitly-taught - CDF performance in different predominantly secondary-level but also tertiary European CLIL contexts, (see Lose, 2007; Coetzee-Lachmann, 2009; Breeze & Dafouz, 2017; Lorenzo, 2017; Dalton-Puffer et al., 2018; Evnitskaya & Dalton-Puffer, 2020; Whittaker & McCabe, 2023). They show that students have difficulties performing CDFs when left on their own, often presenting them in an incomplete way, without including many concepts or explaining them in depth. Such was the case in the pre-test phase, where students failed to present the contents accurately and completely, doing so in a somewhat superficial way, lacking solid argumentation and supporting information. In addition, the students seem to prefer to focus on concrete, tangible content elements rather than relating them to abstract concepts (i.e., comparative criteria), a phenomenon that Evnitskaya and Dalton-Puffer (2020) have analysed in terms of horizontal and vertical content performance. As Nashaat-Sobhy (2020) explains, vertical performance of this kind is particularly challenging.

Overuse of horizontal content performance may be linked to weak conceptual and procedural knowledge, and an inadequate understanding of how knowledge should be displayed using CDFs (Breeze & Dafouz, 2017; Evnitskaya & Dalton-Puffer, 2020). Clear proof of this is that the students at first only superficially included similarity points, to comply with the task instruction, and failed to name the comparative concepts (criteria). It seems that if students are left on their own without explicit support, they have considerably more difficulty presenting the content points in a subject-purposeful way.

The post-tests are consistent with recent interventional studies (Connolly, 2019; Breeze & Gerns, 2019; Nashaat-Sobhy, 2020; Bauer-Marschallinger, 2022; Hasenberger, ongoing), showing how even general, non-explicitly focused CDF support can be very helpful in equipping students to display their content knowledge better, as was the case with the control group who received implicit CDF-based classes. When contrasting the study and control group's CDF-performance, interestingly, both groups did improve in terms of including differences and building comparative concepts (*criteria*). However, the control group did not progress in adding similarities, using lexico-grammatical forms or justifying their statements on their own. An implicit CDF-teaching approach seems to help students pick up some CDF-related features, probably the more familiar ones. In this sense, Breeze and Gerns (2019) reported improvements in a group of Spanish secondary CLIL history students who worked on academic writing skills involving the CDFs of 'describing' and 'explaining', leading to a higher number of content points and improved structure.

In more explicit studies, where the CDF of 'defining' was taught in Spanish tertiary education (Nashaat-Shobhy, 2020), the CDF of 'explaining' in German bilingual secondary chemistry classes (Connolly, 2019) and all seven CDFs were taught in Austrian CLIL secondary history and science classrooms (Bauer-Marschallinger, 2022; Hasenberger, ongoing), a reciprocal effect could be observed on students' content and language learning. It seems that teaching with CDFs helps students retain their subject knowledge better (Nashaat-Sobhy, 2020), increases their use of CDF-related expressions and linking devices (Bauer-Marschallinger, 2022; Hasenberger, ongoing) and fosters their self-confidence (Connolly, 2019). The present post-test results confirm that an explicit teaching approach has a stronger effect on stu-

dents' integrated content and language learning, making them focus more on linguistic and structural features (such as comparitors, linking devices/sequential connectors, explanatory and defining forms, topic sentences, paragraph, and parallel organisation among others), which they would most probably not pay attention to if not prompted.

The results also show how teaching science through a CDF approach may require more time and practice since some points of the framework developed over a longer period (see delayed improvement in presenting similarities in the second post-test).

CONCLUSION

The CDFs construct offers content teachers a practical framework to unlock the importance of language and cognition in the teaching and learning of natural sciences, two dimensions that often receive little attention. This integration can be promoted in a natural way, since the content teacher can point out certain linguistic or cognitive aspects from the scientific perspective without resorting to more complex language models from the EFL class. One of the advantages of the CDFs is that they are basic knowledge-building blocks common to all school subjects (Morton, 2020) and thus inherent to science education.

Furthermore, the study demonstrates the integrated learning potential of CDFs, which can equip students to deal better with content, cognitive and language aspects in a CLIL science class. The CDFs help students advance in their conceptual knowledge and its corresponding language, something students are expected to develop during their years of schooling, which they will not learn in their language class or outside school. These functions, therefore, offer teachers a good starting point to introduce students to subject-specific ways of making meaning and communicating it appropriately, enabling them to participate in scientific discourse in and beyond school. It would thus be desirable that content teachers include more of these cognitive functions in their class input and output to help students in their knowledge-building process. For this, a gradual and collaborative learning path is recommended, making the conceptual, cognitive and linguistic demands visible to students, as most require support to notice and use these tools.

Some of the limits of the present study were the small sample size and the lack of variety between the different written tasks, which follow a similar task design based on a comparative analysis. It would have also been interesting to analyse teacher's and students' classroom interaction and have taught the CDF of *comparing* in another science class (e.g., chemistry).

Further research is therefore needed to operationalize the different CDFs across school subjects and levels, and to provide practical pedagogical tools to foster students' scientific knowledge in different CLIL contexts.

ETHICS STATEMENT

This study formed part of a research project, which was accepted by the Research Ethics Committee (CEI) of the University of Navarra (approval number 2019.164).

ACKNOWLEDGEMENTS

The author is grateful to the teachers and students who participated in the study and also wishes to thank the Fundación Ciudadanía y Valores (Funciva) for their generous support.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Badertscher, H., & Bieri, T. (2009). *Wissenserwerb im content and language integrated learning* [Knowledge acquisition in content and language integrated learning]. Haupt.

Bauer-Marschallinger, S. (2022). *CLIL with a capital I* [Unpublished Doctoral dissertation]. University of Vienna.

Biggs, J. B., & Tang, C. S. (2011). *Teaching for quality learning at university*. Open University Press.

Bloom, B. S. (1956). T*axonomy of educational objectives, Handbook I: Cognitive domain*. Longman.

Breeze, R., & Dafouz, E. (2017). Constructing complex cognitive discourse functions in higher education. *System, 70*, 81–91. https://doi.org/10.1016/j.system.2017.09.024

Breeze, R., & Gerns Jiménez-Villarejo, P. (2019). Building literacies in secondary school history: The specific contribution of academic writing support. *EuroAmerican Journal of Applied Linguistics and Languages, 6*(1), 21–36. https://doi.org/10.21283/2376905X.10.149

Breeze, R., Llamas Saíz, C., Martínez Pasamar, C., & Tabernero Sala, C. (2014). *Integration of Theory and Practice in CLIL*. Rodopi/Brill.

Carpi, A., & Egger, A. E. (2011). Comparison in scientific research. In A. Carpi & A. E. Egger (Eds.), *Process of science* (pp. 129-142). Visionlearning.

Cheong, L. K. (1978). S*yntax of scientific English*. Singapore University Press.

Clark, S.K. Judd, E., Smith, L.K., & Ahlstrom, E. (2020). Examining the effects of integrated science and literacy instruction to teach second graders to write compare and contrast informational texts. *Early Childhood Education Journal, 107*(3), 567-579. https://doi.org/10.1007/s10643-020-01106-9

Coetzee-Lachmann, D. (2009). *Assessment of subject-specific task performance of bilingual geography learners* [Unpublished doctoral dissertation]. University of Osnabrück.

Connolly, T. (2019). *Die Förderung vertiefter Lernprozesse durch Sachfachliteralität* [The promotion of in-depth learning processes through subject literacy] [Doctoral dissertation]. Johannes Gutenberg University Mainz.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In N. H. Hornberger (Eds.), *Encyclopedia of Language and Education* (pp. 487–499). Springer. https://doi.org/10.1007/978-0-387-30424-3_36

Coyle, D., & Meyer, O. (2021). *Beyond CLIL: Pluriliteracies teaching for deeper learning*. Cambridge University Press.

Coyle, D., Meyer, O., & Staschen-Dielmann, S. (2023). *A deeper learning companion for CLIL: Putting pluriliteracies into practice*. Cambridge University Press.

Dalton-Puffer, C. (2004). Academic language functions in content and language integrated classrooms. *Vienna English Working Papers, 13* (1), 23–48.

Dalton-Puffer, C. (2007). Academic language functions in a CLIL environment. In D. Marsh & D. Wolff (Ed.), *Diverse contexts – converging goals* (pp. 201–210). Peter Lang.

Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualizing 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., & 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, K., Hofmann, V., Hopf, J., Kröss, L., & Lechner, L. (2018). Cognitive discourse functions in Austrian CLIL lessons. *European Journal of Applied Linguistics, 6*(1), 5–29. https://doi.org/10.1515/eujal-2017-0028

Darian, S. (2003). *Understanding the language of science*. University of Texas Press.

DeBoer, M., & Leontjev, D. (Eds.). (2020). *Assessment and learning in content and language integrated learning (CLIL) classrooms*. Springer. https://doi.org/10.1007/978-3-030-54128-6

De la Paz, S., & Mc Cutchen, D. (2010). Learning to write. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 32-54). Routledge.

Del Pozo, E., & Llinares, A. (2021). Assessing students'learning of history content in Spanish CLIL Programmes. In *International Perspectives on CLIL* (pp. 43-61). Palgrave Macmillan Cham.

Dixon, R.M. (2005). Comparative constructions in English. *Studia Anglica Posnaniensia, 41*, 5-27.

Doiz, A., & Lasagabaster, D. (2021). An analysis of *the use of cognitive discourse functions in English-medium history teaching at university.* English for Specific Purposes, 62, 58–69. https://doi.org/10.1016/j.ESP.2020.12.002

Evnitskaya, N., & Dalton-Puffer, C. (2020). Cognitive discourse functions in CLIL classrooms. *International Journal of Bilingual Education and Bilingualism*, 1–20. https://doi.org/10.1080/13670050.2020.1804824

Evnitskaya, N., & Morton, T. (2011). Knowledge construction, meaning-making and interaction in CLIL science classroom communities of practice. *Language Education*, 25, 109–127. https://doi.org/10.1080/09500782.2010.547199

Flannery, M. C. (2010). Biology today: Compare, compare. *The American Biology Teacher, 72*(6), 378-381. http://dx.doi.org/10.1525/abt.2010.72.6.13

Gerns, P. (2021). *Cognitive discourse functions in content and language integrated learning: An empirical study in secondary school science* [Unpublished doctorl dissertation]. University of Navarra.

Gerns, P. (2023). Qualitative insights and a first evaluation tool for teaching with cognitive discourse function: "comparing" in the CLIL science classroom. *Porta Linguarum An International Journal of Foreign Language Teaching and Learning*, (40), 161–179. https://doi.org/10.30827/portalin.vi40.26619

Goldstone, R. L., Day, S., & Son, J.Y. (2010). *Comparison*. Springer.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (2004). *Halliday's introduction to functional grammar.* Routledge.

Hammann L.A, & Stevens R. J. (2003). Instructional approaches to improving students' writing of compare-contrast essay. *Journal of Literacy Research, 35*(2), 731-756. https://doi.org/10.1207/s15548430jlr3502_3

Hasenberger, T. (n.d.). *English for the natural sciences: Developing and implementing a curriculum for a new subject at upper-secondary schools* [Unpublished doctoral dissertation]. University of Vienna.

Huddleston, R. (2017). Comparative constructions. In R. Huddleston & G. K. Pullum (Ed.), *The Cambridge Grammar of the English Language* (pp. 1097–1170). Cambridge University Press.

Lemke, J. L. (1990). *Talking science*. Ablex Pub. Corp.

Llinares, A. (2015). Integration in CLIL: A proposal to inform research and successful pedagogy. *Language Culture and Curriculum*, 28(1), 58-73. https://doi.org/10.1080/07908318.2014.1000925

Llinares, A., & Dalton-Puffer, C. (2015). The role of different tasks in CLIL students' use of evaluative language. *System, 54*, 60-79. https://doi.org/10.1016/j.system.2015.05.001

Llinares, A., Morton, T., & Whittaker, R. (2012). *The roles of language in CLIL*. Cambridge University Press.

Llinares, A., & Nashaat-Sobhy, N. (2021). Defining science in primary school CLIL contexts. *Language Teaching for Young Learners, 3*(2), 337–362. https://doi.org/10.1075/LTYL.20010.LLI

Llinares, A., & Nikula, T. (2023). CLIL students' production of cognitive discourse functions: Comparing Finnish and Spanish contexts. *Language and Education*. https://doi.org/10.1080/09500782.2023.2211049

Lorenzo, F. (2017). Historical literacy in bilingual settings. *Linguistics and Education, 3*7, 32–41. https://doi.org/10.1016/j.linged.2016.11.002

Lose, J. (2007). The language of scientific discourse. In D. Caspari, W. Halle, A. Wegner, & W. Zydatiß (Ed.), *Bilingualer unterricht macht schule* [Bilingual education makes school] (pp. 97-107). Peter Lang.

MacArthur, C. A., & Philippakos, Z. (2010). Instruction in a strategy for compare–contrast writing. *Exceptional Children, 76*(4), 438–456. https://doi.org/10.1177/001440291007600404

Martinez, P. (2018). The comparative method in biology and the essentialist trap. *Frontiers in Ecology and Evolution, 6*, 130. https://doi.org/10.3389/fevo.2018.00130

Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives. Experts in assessment*. Corwin Press.

Mercer, N. (1995). *The guided construction of knowledge*. Multilingual Matters.

Meyer, O., Coyle, D., Halbach, A., Schuck & K., Ting, T. (2015). A pluriliteracies approach to content and language integrated learning – mapping learner progressions in knowledge construction and meaning-making. *Language, Culture and Curriculum, 28*(1), 41–57. https://doi.org/10.1080/07908318.2014.1000924.

Meyer, O., Coyle, D., & Schuck, K. (2018). Learnscaping: Creating next-gen learning environments for pluriliteracies growth. In J. Buendgens-Kosten & D. Elsner (Eds.), *Multilingual computer assisted language learning* (pp. 18–38). Multilingual Matters.

Mortimer, E.F., & Scott. P.H. (2003). *Meaning making in secondary science classrooms*. Open University Press.

Morton, T. (2020). Cognitive discourse functions. *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural, 3*(1), 7-17. https://doi.org/10.5565/rev/CLIL.33

Nashaat-Sobhy, N. (2020). Operationalizing 'defining' from a cognitive discourse perspective for learners' use. In S. M. Anwaruddin (Eds.), *Knowledge mobilization in TESOL* (pp. 94–112). Brill Sense. https://doi.org/10.1163/9789004392472_007

Nikula, T., Dafouz, E., Moore, P., & Smit, U. (2016). *Conceptualising integration in CLIL and multilingual educatio*n. *Educational Linguistics* (vol. 32). Springer.

Nikula, T. (2017). 'What's the Moment Thingy?'. In J. Langman, & H. Hansen-Thomas (Ed.), *Discourse Analytic Perspectives on STEM Education* (pp. 11-29). Springer International Publishing. https://doi.org/10.1007/978-3-319-55116-6_2

Norris, S., & Phillips, L. (2009). Scientific literacy. In D. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy*. Cambridge University Press.

Oliveira, A. W., & Weinburgh, M. H. (2017). *Science teacher preparation in content-based second language acquisition*. Springer.

Polias, J. (2015). *Apprenticing students into science*. Lexis Education.

Raphael, T. E., & Kirschner, B. M. (1985). *The effects of instruction in compare/contrast text structure on sixth-grade students' reading comprehension and writing products. Research Series* (vol. 161). Michigan State University.

Rasinger, S. M. (2008). *Quantitative research in linguistics: An introduction*. Continuum.

Roberts, D.A., & Bybee, R. W. (2014). Scientific literacy, science literacy, and science education. In N. Lederman & S. Abell (Eds.), *Handbook of research on science education*. Routledge.

Rose, D., & Martin, J. R. (2012). *Learning to write, reading to learn: Genre, knowledge, and pedagogy in the Sydney school*. Equinox Publishing.

Salvador-Garcia, C., & Chiva-Bartoll, O. (2022) The nature of integration in physical education through CLIL. *Journal of Language, Identity & Education*. https://doi.org/10.1080/15348458.2022.2085107

Schalk, L., Saalbach, H., & Stern, E. (2016). Approaches to foster transfer of formal principles. *PloS One, 11*(2), e0148787. https://doi.org/10.1371/journal.pone.0148787

Schleppegrell, M.J., & Colombi, M.C. (2005). *Developing advanced literacy in first and second languages*. Routledge.

Silver, H. F. (2007). *Compare & contrast: Teaching comparative thinking to strengthen student learning*. Association for Supervision & Curriculum Development.

Smith, S. (2019). *Academic writing genres*. Evident Press.

Vanderbeke, M., & Wilden, E. (2017). Sachfachliche Diskursfähigkeit durch fremdsprachliche affordances in bilingualen Schülerlaborprojekten [Subject-related discourse skills through foreign language affordances in bilingual school laboratory projects]. *Zeitschrift für Fremdsprachenforschung, 2*(1), 3-27.

Veel, R. (1997). Learning how to mean – scientifically speaking. In F. Christie & J. R. Martin (Eds.), Genre and Institutions (pp. 161-195). Continuum.

Villabona, N., & Cenoz, J. (2021). The integration of content and language in CLIL. *Language, Culture and Curriculum, 35*, 36 - 50. https://doi.org/10.1080/07908318.2021.1910703

Vollmer, J. H. (2010). *Items for a description of linguistic competence in the language of schooling*. Language Policy Division.

Whittaker, R., & A. McCabe. (2023). Expressing evaluation across disciplines in primary and secondary CLIL writing. *International Journal of Bilingual Education and Bilingualism, 26*(3), 345–362. https://doi.org/10.1080/13670050.2020.1798869.

Widdowson, H. G. (1979). *Reading and thinking in English*. Oxford University Press.

Zwiers, J. (2014). *Building academic language: Meeting common core standards across disciplines*. Jossey Bass Teacher.

APPENDIX A

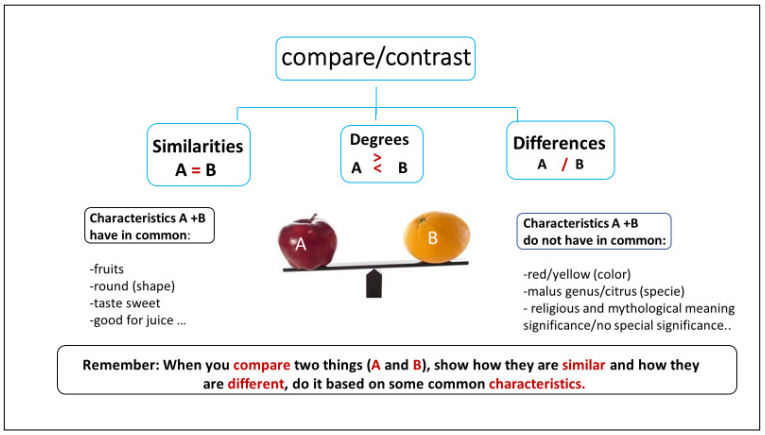
'Comparing' in science

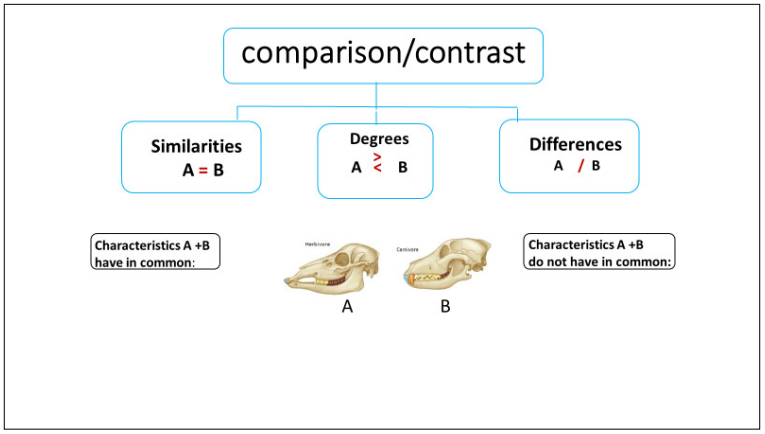
Why do we compare in biology?

What is important for writing a good comparison in science?

APPENDIX B

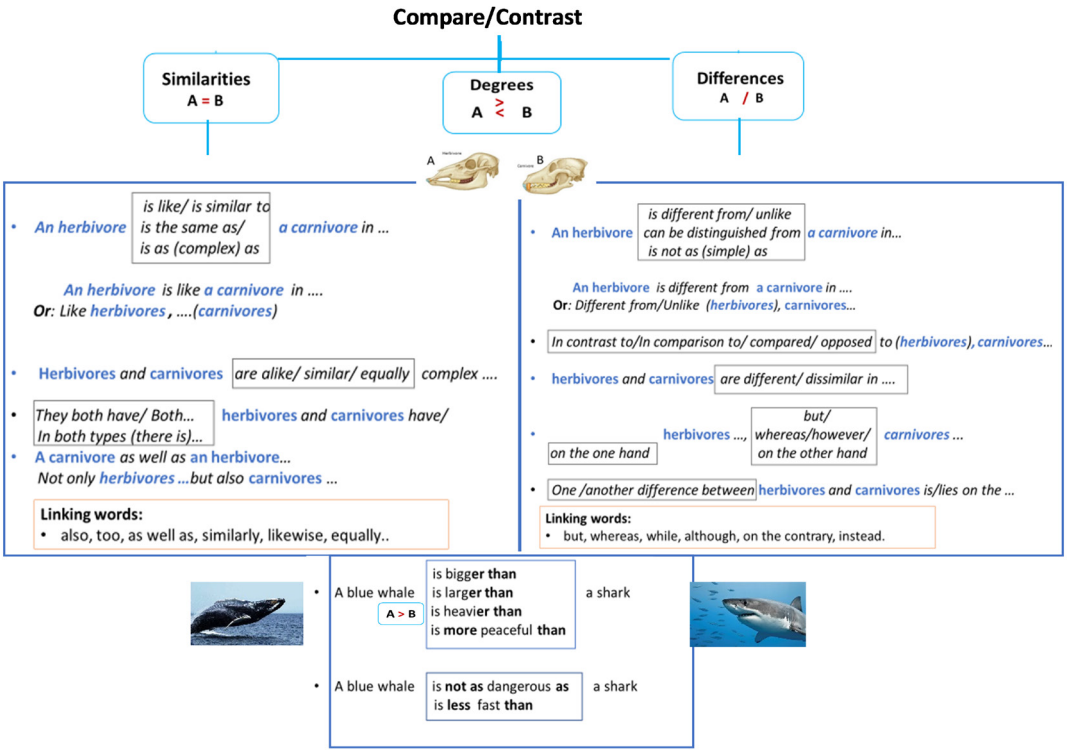
Components of the CDF '*comparing*'





APPENDIX C.

The language *of* and *for* comparing (lexico-grammatical forms; *comparitor*)



APPENDIX D

Individual written tasks

Please form sentences that indicate how herbivores and carnivores are different. To do this, use the contrast forms of the box. Do it as the example indicates.

Example:
***1**. Herbivores are plant eaters, **while** carnivores are meat-eaters.*
***2**. Herbivores have a tooth set that **is very different from** carnivores in…*

3.
4.
5.
6.

Expressions for contrast

- An herbivore | is different from/ unlike / can be distinguished from / is not as (simple) as | a carnivore in…

An herbivore is different from a carnivore in ….
Or: Different from/Unlike (herbivores), carnivores…

- In contrast to/In comparison to/ compared/ opposed to (herbivores), carnivores…
- herbivores and carnivores are different/ dissimilar in ….
- on the one hand herbivores …, but/ whereas/however/ on the other hand carnivores …
- One /another difference between herbivores and carnivores is/lies on the …

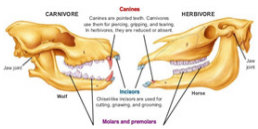
Linking words:
- but, whereas, while, although, on the contrary, instead.

APPENDIX E

Reading and writing tasks

1. Read the following text with your neighbour. Identify comparative **SIGNAL WORDS**. Circle the ones indicating a **similarity in green** and the ones indicating a **difference in red**. Fill these expressions in the table below.

Signal words	
similarity	difference



2. In the text, go paragraph by paragraph and identify the **main idea** on which the two animals (herbivores and carnivores) are compared. In the first paragraph, for example, carnivores and herbivores are compared on their diet. Now try to figure out what these main comparative points are for the other paragraphs.

3. Find out what the text says about the concrete **differences** and **similarity**. What do the two groups have or do not have in common? Include this information in a Venn diagram or a table. Good luck! 😉

Differences and similarities between herbivores and carnivores

In the animal kingdom, two major kinds of animals can be distinguished. These include herbivores and carnivores that will be compared.

Type of food/ Diet

1.Herbivores, as well as carnivores, are similar in being **heterotrophs**. This means both animal groups **depend on others** for food. They require another living thing to live. But, having a closer look at **the concrete type of food** they eat, they differ. Herbivores refer to those animals that **consume only plants**. This includes **leaves[1], grains[2]** and **seeds[3]**. In contrast, carnivores are those animals that eat **meat** to survive[4].

2.Determining **whether a** particular **animal is** a herbivore or a carnivore is **not just** limited to merely **observing** what **particular type of food** they eat, but by simply looking at the **skeletal remains[5]**: at their **teeth.** Both groups possess an identifiable tooth set that allows them to eat. On the one hand, herbivores have **broad, flat canines[6]**, sometimes they are even absent, and with **large, dull incisors[8]**. This helps them to properly **grind[9] the fibres** found in plants and **crush seeds and grains**. On the other side, carnivores have teeth proper to a hunter with **large, pointed canines** and short**, sharp incisors.** This allows them to **tear[10] through skin** and muscle and **chew[11] the meat** from their **prey[12].**

3. Another comparable feature is the **nails**. Herbivores have **flattened nails[13]** or blunt **hooves[14].** Carnivores have **sharp claws[15]** instead**.** Again, the difference between the **nails** has to do with **the type of food** they eat. Carnivores have to **hunt and kill** their **prey** to eat their **flesh**, and these **claws help** them leave their **prey helpless.**

4. In line with this, there is also a difference in the **size[17]** of the **mouth opening** with the **size of** the **skull[18].** As herbivores only **eat plants**, their **mouth opening** is relatively **small** compared with their **head size**. In contrast, carnivores have **larger** mouth openings. This is because carnivores do **not just use** their **mouths to eat but** utilize their sharp teeth **to attack and kill** their prey.

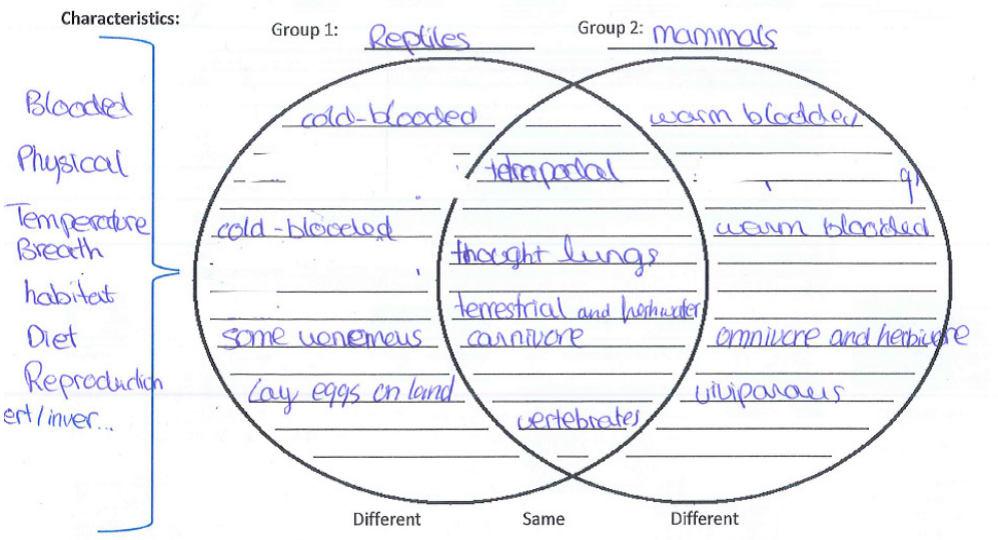
5. A last point to be compared is the **digestive system** of both groups. As digestion starts in the mouth, by breaking down the food into smaller pieces, in both cases, their **teeth are indispensable** for this process. Nevertheless, digestion is not as complex in carnivores as in herbivores. Carnivores have a very **simple, single-chambered stomach** because the **meat** is relatively **easily digested**. Hence[19], their **small intestine** is quite **short** too. On the contrary, herbivores that consume high amounts of **plants**, difficult to digest must do **a lot of chewing**. That is why they have a **complex, multi-chambered stomach** and a **long small intestine**.

Concluding, herbivores only **utilize** their **flat**-based, dull **teeth for consuming food** (plant leaves, seeds and grains) which is why they **only have small mouth openings**. While carnivores **use** their **sharp incisors and canines** as a weapon[20], making it necessary for their **mouth openings** to be relatively **larger**.

APPENDIX F

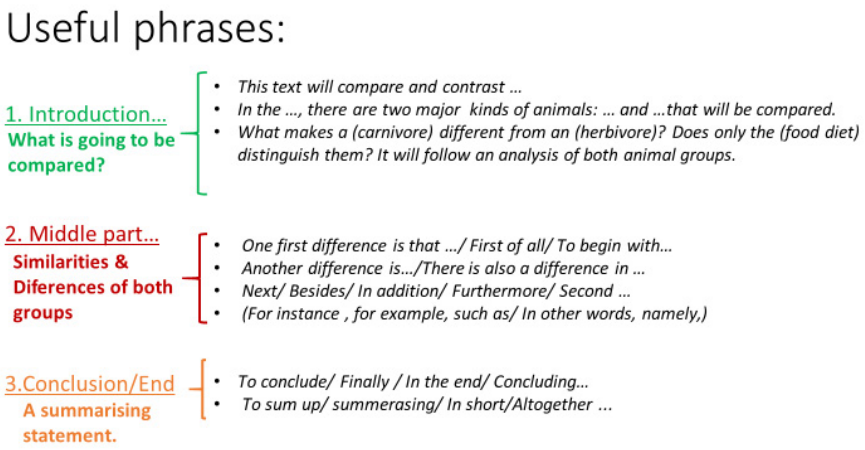
Examples of students' responses

Characteristic:	Group 1: Herbivores	Group 2: Carnivores	Common (=)
Diet/Type of food	***Herbivores:*** *eat plants (leaves, grains, seeds, ect.)*	***Carnivores:*** *eat meat.*	Both are heterotroph
Legs	Herbivores: flattened nails	Carnivores: sharp claws	
Teeth	Herbivores: flat canines dual incisors	Carnivores pointed canines sharp incisors	
Mouth	Herbivores: small mouth opening	Carnivores: larger mouth opening	
Digestion	Herbivores: complex multi-chambered stomach long small intestine	Carnivores: simple single-chambered stomach short small intestine	In both digestion starts in the mouth



APPENDIX G

General structure of a comparison and useful starting sentences



APPENDIX H

Useful framework to redo your first comparison

Remember!! Steps for writing a scientific comparison:

1. **Identify** the groups, that will be compared.
2. **Make a brainstorming** on **characteristics** both groups have in common (**similar**ities) or not (**differen**ces). (Do not list irrelevant information!) (Note them in a Venn diagram or tablet).
3. **Order the points logically.**
4. **Time for writing.**
   Give your text a structure!
   (1. Introduction,**2.** Middle part (Point-by-point Method), **3.** Conclusion).
   Start with an **introduction.**
5. **Link isolated ideas** and **use signal words** to stress **similarities and differences**.
6. Do not forget a summarising **conclusion**! 😉

https://doi.org/10.17323/jle.2023.16907

Stand-Alone Conclusion Section in Open-Access Research Articles: Organizational Structure

Budsaba Kanoksilapatham [1,2]

[1] Silpakorn University, Phetchaburi, Thailand

[2] Academy of Arts, Royal Society of Thailand

ABSTRACT

Background: Research articles (RAs) have been highlighted as one of the most essential channels for academicians to disseminate knowledge. Typically, RAs consist of a nomenclature of the four sections of introduction, methods, results, and discussion, commonly known as IMRD, with each section governed by an organizational structure determined by genre analysis. However, due to the increasing prevalence of the stand-alone conclusion section in recent years and the expansion of open-access journals, understanding how this conclusion section is constructed can benefit scholars in an academic endeavor.

Method: This study aims to identify the organizational structure of the stand-alone conclusion section in open-access journals. An original dataset of 55 open-access journal RAs from four major academic disciplines honored as "Articles of the Year 2021" was analyzed. Only the RAs with stand-alone conclusions within a calculated word range were selected, yielding a final dataset of 25 comparable conclusion sections. Then, anchored on genre analysis, the compiled dataset was analyzed quantitively and qualitatively.

Results: Based on the genre analysis conducted, a set of three moves and their pertaining steps were identified, forming various organizational patterns but one compelling one. Furthermore, the full-fledged two-layer rhetorical structure of the section depicting the frequencies of occurrence of individual moves and steps is quite revealing, highlighting the crucial significance of restating the findings generated from the study being reported. The results demonstrate not only how established this section is but also how it is currently structured, performing its critical function of concluding RAs.

Conclusion: This study has provided insights into integrating analytical elements to successfully stage persuasive arguments in the conclusion section, a skill that needs to be inculcated in novice or early-career researchers and seasoned researchers alike.

KEYWORDS

stand-alone conclusion section, genre analysis, move analysis, open-access journal, research article, structural organization

Citation: Kanoksilapatham B. (2023). Stand-Alone Conclusion Section in Open-Access Research Articles: Organizational Structure. *Journal of Language and Education, 9*(3), 79-89. https://doi.org/10.17323/jle.2023.16907

Correspondence:
Budsaba Kanoksilapatham,
kanoksib@hotmail.com

Received: March 18, 2023
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

Currently, academicians are strongly encouraged to be actively involved in research so that their knowledge and expertise can be put into action, yielding discoveries and moving a field forward. Upon completing a research study, publishing their research discoveries is another significant but challenging step. Research articles (RAs) have been recognized as one of the most important channels to disseminate discoveries or findings. RAs are known to consist of a typical nomenclature of the four major sections of introduction, methods, results, and discussion, commonly known as IMRD. Swales' genre analysis (1990; 2004) has provided a tool for deciphering the structure of each RA. Based on genre analysis, each RA section is regulated by an organizational structure consisting of several "moves" and "steps." A move refers to a portion of a text that performs a principal communicative function. In turn, each move can be substantiated by one or several steps, the sub-units of a move. In fact, the fully developed analyt-

ical framework of genre analysis by Swales in 2004, facilitating the task of RA writing, has expedited the flourishing field to an exponential rate, resulting in a multitude of studies on the RA genre.

Prior to presenting research studies utilizing Swales' genre analysis (1990; 2004), a crucial observation relevant to this study is needed. Based on a common practice adopted by RA authors, RAs can be divided into clearly separated or stand-alone sections. As an illustration, an introduction section performs its function of introducing the topic, whereas the methods section highlights the study's research methodology. These sections are easily recognized by the accompanying and corresponding labels such as "introduction" or "methods." Nevertheless, such definite and distinct boundaries between the RA sections have been compromised. For example, in some RAs, authors can combine the results and discussion sections, yielding the "results and discussion."

To illustrate the constant and growing interest in and the pervasiveness of Swales' genre analysis (1990; 2004) as a tool to demystify the rhetorical structuring of the RA genre, a plethora of studies have concentrated on individual or so-called stand-alone sections of RAs written in English and in various academic disciplines (e.g., applied linguistics discussions by Amirian et al., 2008; dentistry discussions by Basturkmen, 2012; scientific abstracts by Cross & Oppenheim, 2006; engineering introductions by Kanoksilapatham, 2012). Several research studies focus on two stand-alone RA sections (abstracts and introductions by Affandi Adrianto et al., 2021), whereas many studies focus on the combined sections (e.g., results, discussions, and conclusions by Ruiying & Allison, 2003; results and discussions by Suherdi et al., 2020). Few studies analyzed RAs in their entirety (e.g., Kanoksilapatham, 2005 and 2015 in biochemistry and engineering, respectively; Stoller & Robinson, 2013 in chemistry). In addition, a multitude of studies was conducted comparing and contrasting the rhetorical structures of particular sections across disciplines (e.g., conclusions in psychology, Persian literature, and applied linguistics by Adel & Ghorbani Moghadam, 2015; results in sociology and organic chemistry by Bruce, 2009; methods from over 30 disciplines by Cotos et al., 2017; abstracts from three engineering sub-disciplines by Kanoksilapatham, 2013) and across languages (e.g., Iranian and English by Ahmadi, 2022; English and Russian by Elena, 2017 and Zanina, 2017; Chinese and English by Juan & Tao, 2013; Thai and English by Kanoksilapatham, 2007; Malay and English by Loi et al., 2021; English and Spanish by Martin, 2003 and Moreno, 2021; Filipino and Japanese conclusions by Morales, 2012; English and Indonesian abstracts by Pratiwi & Kurniawan, 2021).

The implementation of Swales' analytical framework (1990; 2004) has also been successfully extended to other academic genres, be they written or oral, such as master's theses (Chen & Kuo, 2012), academic presentations (Hu & Liu, 2018; Vera & Williams, 2020), and dissertation acknowledgements (Yang, 2012). Furthermore, genre analysis proved to be effective in discerning the structural organizations of professional genres (e.g., TED Talks by Compagnone, 2015; manuals by Danquah et al., 2020; corporate annual reports by Qian, 2020; story genre by Salmaso, 2017). All in all, genre-based studies, as exemplified above, have demonstrated that Swales' genre analysis has rendered a tremendous contribution pedagogically and scientifically.

As reviewed above, the conclusion section seems to be not as extensively examined as the other RA sections. Ruiying and Allison (2003) are the first scholars who analyzed the combined section of results, discussions, and conclusions using genre analysis. The application of genre analysis on 20 combined sections in applied linguistics from prestigious journals led them to propose a seven-move model. Based on their seven-move model for the combined section, the last three moves were identified to belong to "the allegedly conclusion section."

At this point, an examination of Ruiying and Allison's (2003) model proposed for the combined result-discussion-conclusion section is relevant. Even though the seven-move model was initially established for the combined section of RAs, subsequent scholars adopted the last three-move structure of the model as a taxonomy and an analytical framework to analyze stand-alone conclusions (e.g., Adel & Ghorbani Moghadam, 2015; Morales, 2012; Zamani & Ebadi, 2016). However, it needs to be remarked that these studies do not primarily seek to decipher the rhetorical pattern of the stand-alone conclusion section. Instead, they congruently aimed to pinpoint cultural features manifested in the rhetorical structures. To elaborate, Morales (2012) analyzed 16 RA conclusions authored by Filipino and Japanese writers in English from the field of applied linguistics. Even though the section from the two datasets somehow conformed to Ruiying and Allison's model structure, the cultural influence manifested in the rhetorical organization of this section was elucidated. Moreover, Zamani and Ebadi (2016) investigated 20 conclusions written in English and Persian from the fields of civil engineering and applied linguistics. The detected structural organization was generally compatible with Ruiying and Allison's 2003 model, revealing some discernible variation. However, no statistically significant differences regarding the move structure and frequencies were found.

As far as the conclusion section is concerned, two observations can be made here. First, whereas the structural organizations of the combined section of results, discussions, and conclusions seem to be elucidated by previous studies presented above, no single study concentrating on the stand-alone conclusion section is yet available. Second, several journals are presently enforcing an increasingly widespread guideline that requires RAs to have a stand-alone conclusion. Thus, it is intriguing to examine how the stand-alone conclusion is organized rhetorically. It is recognized that the RA genre has evolved. For example, the rudimentary and

established IMRD nomenclature has been attested. Some examples of the evolutionary features include the combined section of results and discussion, the combined discussion and conclusion section, and the recommended inclusion of the conclusion section by several academic journals. This development path may indicate that academicians are anxious to learn more about the results, and thus the discussion of the results is essentially integrated into the results. The conclusion section is possibly required to accommodate the readers' needs who want to quickly grasp the big picture of the study being reported and the writers' arguments. Although some RA writers merge the conclusion section with the discussion section, many writers still prefer to include this succinct stand-alone section at the end of RAs.

In scrutiny, exponentially developed communication technology has triggered and expedited the evolution of academic writing. According to Mering and Hoeve (2020), due to the free availability of the public internet, open-access journals (as opposed to traditional print-based academic journals) have emerged as a crucial source of full RAs available online. Consequently, scholarly communication and publications can be instantly accessible. It turns out that open-access journals can help alleviate many problems encountered, including the exuberantly high demand for publishing channels and the high cost of printing. Last but not least, to make the ultimate advantage of open-access journals and the final section of the conclusion in which arguments based on the results must be staged persuasively and powerfully, a look at the organizational structure of this RA section in open-access journals is quite illuminating.

With all the above factors combined (i.e., the uncontested conclusion model since 2003, the dynamic nature of scholarly publications, the emergence and gained popularity of open-access journals, and the section position that bears tremendous significance), this study aims to examine the current practices of scholars when crafting RA conclusions by characterizing the structural organization of the section. Specifically, this study seeks to determine the organizational structure commonly followed in the stand-alone conclusions in open-access journals. These insights will provide a better understanding of how the section is currently organized, allowing researchers to successfully communicate and disseminate their discoveries with the audience in a manner that conforms to the current expectations of specific communities.

METHOD

This study used a mixed-method research design. To elaborate, genre analysis is inherently qualitative; however, this study chooses to substantiate the qualitative findings with both quantitative data and additional qualitative evidence based on the conclusions' excerpt instances. The methodological procedures are described in detail in the following sub-sections.

Objectives

Previous genre-based studies, as demonstrated, typically focus on non-open-access and prestigious journals to elucidate best practices in writing RA conclusions. However, given the current phenomenal popularity and proliferation of open-access journals that meet the need for rapid publication while also accommodating the availability and accessibility of published RAs made possible by information technology, RAs published in open-access journals deserve our attention. As a result, this study aims to determine the organizational structure of stand-alone RA conclusions published in open-access journals. Furthermore, comparing the organizational structure of this study with that of Ruiying and Allison (2003) will shed light on the evolution of this particular section, highlighting structural similarities and discernible deviations after two decades of the known taxonomy.

Dataset Compilation

Compiling datasets is critical in genre analysis because it can substantially impact the analysis results. Basic tenets regarding dataset compilation include the representativeness and comparability. Whereas non-open-access journals are generally indexed for prestige (e.g., Web of Science or WoS and SCOPUS), and journal quality is regularly assessed or monitored, such evaluation does not seem to be available for open-access journals. The widespread popularity of open-access journals, on the other hand, piques our interest. One of the commercial publishers of scientific, technical, and medical literature with over 230 peer-reviewed open-access journals and approximately 20,000 articles per year announced on its website in October 2022 that 55 original and review articles were recognized as the "Article of the Year 2021" (https://www.hindawi.com).[1] According to the commercial publisher's website, these articles represent the four major academic disciplines (biological sciences; mathematics, engineering, and computer science; medicine; and physical sciences). They were selected by the corresponding chief editors based on their high academic impact, and they are freely downloadable at https://www.hindawi.com/post/article-year-2021/.[2]

Dataset Preparation

These articles were published in 2021, as indicated in the award title. According to the journal's website, the first glance at these 55 award-winning articles revealed that not

[1] Hindawi. (2023, October 22). Open-access publishing for the scientific community. https://www.hindawi.com

[2] Hindawi. (2023, October 22). Article of the year 2021. https://www.hindawi.com/post/article-year-2021/

all of them are RAs. Two award-winning articles are review articles and thus were excluded from this study dataset, leaving 53 RAs. Further examination of individual RAs revealed that 46 (or 84%) of the 53 RAs contain the conclusion section (combined or stand-alone). This relatively high occurrence suggests that the section has a rather established and preferential status, at least in this dataset. To maintain the dataset's homogeneity and avoid potential confounding results, two RAs with the combined conclusion section were excluded. A closer look at the remaining 44 stand-alone conclusions revealed another issue with comparable word length: some are quite short, with only 57 words, while the others are quite long, with around 600 words. A significant difference in word counts could have a negative impact on the rhetorical orientation and genre analysis outcomes. As a result, the average length of the 44 conclusions was calculated. Based on this average value of 186.36 words, the word range of 100-300 was arbitrarily used as a criterion to finalize the dataset of 25 stand-alone conclusions. All procedures described were to improve the comparability and homogeneity of the final dataset. They are not intended to support or indicate the quality of the conclusions.

These 25 conclusions were coded to facilitate subsequent genre analysis and references. Each code is made up of two information pieces enclosed in parentheses. The first part has three initial letters that correspond to the disciplines to which they belong (i.e., *Bio* for biological sciences; *Mat* for mathematics, engineering, and computer science; *Med* for medicine; and *Phy* for physical sciences). The second part is a number assigned reflecting the sequence in which the selected articles are presented on the website. For example, *(Bio1)* refers to the conclusion of biological sciences and is the discipline's first article. Meanwhile, *(Phy7)* refers to the physical sciences' conclusion, and it is the seventh of the discipline. The RA content remains intact. That is, no modification was made to the conclusion content.

Analytical Framework

As previously stated, Ruiying and Allison (2003) conducted one of the most widely recognized genre-based studies on the combined result-discussion-conclusion section. From an analysis of 20 applied linguistic sections selected from prestigious journals, they concluded that the organizational structure consists of seven major moves. Based on the model, three out of seven moves presumably belong to the "allegedly conclusion section." They are *Move 1: Summarizing the study; Move 2: Evaluating the study; and Move 3: Deductions from the research*. For Move 1, no steps were posited. Move 2 consists of three steps *(Step 1: Indicating significance/advantage; Step 2: Indicating limitations; and Step 3: Evaluating methodology*) and Move 3 consists of two steps (*Step 1: Recommending further study and Step 2: Drawing pedagogic implication*). This study was anchored on Ruiying and Allison's 2003 taxonomy as an analytical framework to analyze the stand-alone conclusion dataset.

Dataset Analysis

Typically, genre analysis units are determined by the communicative functions of the texts rather than syntactic structures (phrases, clauses, sentences, or extended sentences). Individual conclusions were subjected to genre analysis under this tenet. Multiple readings of individual conclusions were performed to obtain a general understanding of the texts and follow the thread of argument presented, as expected in any genre analysis. In the event of doubt or ambiguity, the researcher reread all of the RAs.

Using Ruiying and Allison's 2003 taxonomy, genre analysis was performed three times, each time at one-week intervals, resulting in a clear understanding of the texts and increased confidence in the move/step segmentation. Given that each move can be realized by several steps that share the same communicative function, it is possible that new steps were discovered and thus added to the existing taxonomy. Similarly, if a taxonomy step were not found in the dataset, it would be removed. The removal and addition of moves/steps result in a modified organizational structure that is currently accurate in describing the conclusion texts in the dataset. After the genre analysis was completed, the frequencies of individual moves and steps were recorded.

According to Kanoksilapatham (2012, 2015), the status of individual moves (obligatory, conventional, and optional) can be quite gratifying. This feature aids in determining how stable each move is and how frequently it occurs. To elaborate, "obligatory" refers to a move's very stable status; it appears in every text. Meanwhile, "conventional" and "optional" refer to a move's not-so-stable status. As the labels suggest, "conventional" indicates that the presence of a move is present but not as frequently as "obligatory," and "optional" refers to a relatively unstable status or a move that occurs infrequently. Finally, the organizational structures of individual conclusions were summarized based on the data regarding the presence of moves and steps, their frequencies of occurrence, and their sequence in the conclusions.

RESULTS AND DISCUSSION

The 25 stand-alone conclusion dataset analysis reveals the presence of three moves, as included in the original 2003 taxonomy. Typically, the move and step labels are self-explanatory and transparent, reflecting specific communicative functions, and each move can perform its function or be supported by one or more steps. To facilitate a comparison with Ruiying and Allison's original taxonomy (2003), efforts were made to adhere to the labels used initially, and changes, if necessary, were kept to a minimum. The following sections present a qualitative and quantitative description of each move and step, accompanied by frequencies of occurrence and corresponding excerpts taken directly from the

dataset with the codes attached at the end in parentheses referring to their sources.

Move 1: Summarizing the Study

Move 1 is to summarize the study reported, and it was found in all 25 conclusions (or 100%), thus earning the status of obligatory. Interestingly, this high occurrence rate was congruently reported by Morales (2012) and Zamani and Ebadi (2016), who analyzed the conclusion section. This move establishes the context for an argument to be subsequently developed in Moves 2 and 3. In other words, this move reiterates and reinforces the background information of the study, from which further arguments can be developed in the subsequent moves of the section.

Whereas Ruiying and Allison (2003) did not report any step of this move in their framework, nor did Morales (2012) or Zamani and Ebadi (2016), the current study reveals that this move consists of four steps, reiterating different aspects of the study reported. These details about the study being reported are usually presented in the introduction and methods sections of RAs; thus, to reflect this nature, these steps identified were labeled with the word "restating" (i.e., *Step 1: Restating background, Step 2: Restating objectives, and Step 3: Restating methods, and Step 4: Restating findings)*. Of all the four steps of Move 1, Step 4 (*Restating findings*) was most frequently used (in 21/25 conclusions, or 84%), whereas the other three steps (*Restating background, Restating objectives, and Restating methods*) were used much less frequently (in 40-48% of the conclusions).

Move 1, as found in this dataset, can be represented by using only one single step or a combination of steps to accomplish the move function. To illustrate, the excerpts representing these four steps are presented below, with certain linguistic features or constructions highlighted to reflect the association between language use and communicative functions. It is in the best interests to include corresponding step excerpts from the four disciplines. However, not all conclusions of any disciplines in the dataset use all the steps. As a result, only some steps are accompanied by four excerpts. The only change made to the excerpts is that selected linguistic features, lexical or grammatical, are highlighted.

Move 1 Step 1: Restating Background

- *The chitosan-β-glycerophosphate hydrogel 259* ***is*** *an intelligent drug system* ***widely used*** *to deliver hydrophilic and hydrophobic drugs to different parts of the body. (Bio2)*
- *Since artificial intelligence* ***has greatly changed*** *the data link form, IoT technologies* ***are making significant progress*** *these years, where ontology matching* ***plays an advance role*** *[47–51]. (Math6)*
- *The* ***recent*** *outbreak of COVID-19 is* ***affecting many*** *countries* ***worldwide*** *including Iran as one of the* ***top 10 most*** *affected countries. (Med2)*
- *With the* ***rapid*** *advancement of nuclear power, storage of spent nuclear fuel and transportation safety* ***have become unavoidable issues****. (Phy4)*

Move 1 Step 2: Restating Objectives

- *The role of distinct landscape fragmentation thresholds on the reemergence of YFV in Brazil* ***was tested****. (Bio1)*
- ***This paper proposes*** *a novel asymmetric magnetic pole PMSM for automobiles. (Mat2)*
- ***We provide*** *a proof of conceptual investigation of brain tumour tissues by a multimodal spectroscopy approach. (Med9)*
- ***In order to better understand*** *the conformable fractional derivatives and integrals, … (Phy5)*

Move 1, Step 3: Restating Methods

- ***Experimental conditions comprise*** *a constant-torque experiment and an acceleration-and-deceleration experiment. (Mat3)*
- *Due to a combination of three different spectroscopic methods, an overall spectroscopic-based* ***PCA-DA model was developed*** *to characterize brain tumour tissues with regard to the degree of malignancy. (Med9)*
- *The Al7075/B_4C (50 μm) composites* ***were fabricated*** *by stir casting using a vacuum furnace. The characterization of composites with different weight fractions of reinforcement materials (6, 8, 10, and 12 wt.% of reinforcements)* ***was carried out****. (Phy2)*

Move 1 Step 4: Restating Findings

- *The estimated value of the CG basically* ***converges*** *to the reference value in the simulation and experiment. (Mat3)*
- *… calves from the synbiotic group* ***showed better results*** *in virtually all parameters. (Med7)*
- *Using …,* ***we demonstrated that*** *formalin fixation is a suitable sample preparation method for spectroscopic investigations. (Med9)*
- ***Our simulation work shows that****, in the laser energy and duration range considered here, the laser pulse*

duration only has a small influence on prompt doses and activation. (Phy 7)

In short, Move 1 provides a summary of the study. This move is very stable, being employed in all of the conclusions analyzed. Moreover, this move features four possible steps or strategies to help accomplish the move function. Move 1 Step 4: Restating findings is remarkable as far as its frequency of occurrence is concerned.

Move 2: Evaluating the study

The conclusion section does not only summarize the study as demonstrated in Move 1. Move 2, in turn, substantiates Move 1 by offering evaluative comments related to the reported study. This move was found in 17 (68%) conclusions, earning the conventional status. In congruence with Ruiying and Allison's taxonomy (2003), this move can be represented by a maximum of three steps (*Step 1: Indicating significance or advantage, Step 2: Indicating limitations, and Step 3: Evaluating methodology*). Through this move, the RA writers begin to offer critical comments on their study, contributing to compelling, rational, and valid arguments. The occurrence frequency rates of these steps are relatively much lower than those of Move 1. That is, they were found in six, three, and eight conclusions (or 24, 12, and 32%), respectively. Here are the excerpts corresponding to the three steps of Move 2, with selected linguistic features highlighted.

Move 2 Step 1: Indicating Significance or Advantage

- *These new systems are **more efficient than** chitosan beta-glycerophosphate hydrogels. (Bio2)*
- *This work's dense cross-attention mechanism c**an jointly distribute** the attention weights horizontally and vertically. As a result, it can efficiently mine the most representative features of the hyperspectral data. (Mat5)*
- *… our result **could be useful** in preparing for future outbreaks as well as the current one by considering the results in public health decision-making. (Med2)*

Move 2 Step 2: Indicating Limitations

- ***However, the lack of** nonelectrical test data for transformers and circuit breakers in this study is **not comprehensive enough**. (Mat1)*
- *As the results of our study are very promising but the significance of these results was **not always clear**. (Med7)*

Move 2 Step 3: Evaluating Methodology

- *The **important feature of the proposed real-time estimation method** is that it is suitable for FREWL dynamics and operational characteristics and that it enhances our ability to solve strong nonlinear problems while avoiding the problem of a negative-definite Cholesky factor. (Mat3)*
- *Considering the implementation of high pretest probability, cardiac markers testing, particularly hs-cTnI, **is advised on** admission for selected patients, such as older adults and those with preexisting cardiovascular comorbidities. (Med 1)*
- *For glucose sensing in the food and pharmaceutical sectors, for the long-term usage of nanostructured glucose biosensors, reliability and longevity in extreme environments such as heat/cold, saline, and acid/ primary conditions **need to be taken into account**. (Phy3)*

In short, Move 2 aims to present evaluative accounts associated with the study being reported. In congruence with Ruiying and Yang's 2003 model, three steps were identified. This move and the associated steps had substantially lower frequencies when compared with those of Move 1 and its four steps.

Move 3: Deducting from the Study

Move 3, if used, tends to be the last move concluding the section by offering deductions from the study's findings. The same rate of occurrence as Move 2 was found (in 17 conclusions or 68%), and thus similarly conventional. In line with Ruiying and Allison's 2003 model, this move features two possible steps that help project the study into the future (*Step 1: Recommending further studies and Step 2: Drawing implications)*. These steps have a slightly higher frequency than those of Move 2 but are still lower than those of Move 1. That is, they were found in nine and eight conclusions (36 and 32%, respectively).

Move 3 Step 1: Recommending Further Studies

- *Due to this drug delivery system's desirable features, it is predicted that the drug-containing formulations from this drug system **will enter the global market in the future**. (Bio2)*
- ***In the next work**, we will try to describe the similarity between matching pairs by applying a new method and improve the extraction method to make it more suitable for the characteristics of the sensor ontology. (Math6)*

- ***Future research work*** *will not only include the investigation of other tumour types but also consider the common paraffin embedding procedure. (Med9)*

Move 3 Step 2: Drawing implications

- *Furthermore, such predictions* ***may help other countries*** *that are now battling the outbreak to be more prepared. These measures are also essential to control the epidemic, protect frontline health workers, and reduce the severity of patient outcomes. (Med2)*

- *… erythrocyte-related indicators* ***can provide more clinical information*** *and can be used to monitor the progression of diabetes and its complications. (Med6)*

- *This work* ***can provide a theoretical basis*** *for the optimization design and experimental preparation of novel neutron shielding materials. (Phy4)*

Similar to the 2003 taxonomy, the closing move or Move 3 was represented by two steps. Similar to Move 2 and its pertaining steps, the move and step frequencies are much lower than those of Move 1.

Based on the description of individual moves and steps accompanied by their corresponding excerpts and their frequencies of occurrence presented above, Table 1 presents a conclusion taxonomy delineating the three moves found, the typical sequence of moves and steps, and their frequencies of occurrence which provide a criterion for their status.

To substantiate the qualitative descriptions in the previous section and to justify the above taxonomy, Table 2 illustrates the quantitative descriptions of individual conclusions in the dataset, including a variety of organizational patterns identified and the steps employed in each conclusion as well as their total occurrence frequencies across the 25 conclusions.

As far as the move sequence is concerned, which forms an organizational pattern in the conclusion dataset, Table 2 demonstrates that not all 25 conclusions follow the same pattern. The analysis reveals that four patterns were discovered and are presented here in descending frequency order: 1-2-3, 1-3, 1-2, and 1. Despite a range of patterns, the 1-2-3 pattern was the most common, appearing in 11 conclusions and accounting for 44% of the dataset. The other patterns are more sporadic, appearing in only 6, 5, and 3 conclusions, respectively. Corresponding with the previous descriptions of individual moves and substantiating previous studies in this section, Move 1 is always present. If all three moves are present, the pattern of 1-2-3 is the most common. Finally, it is confirmed that Move 1, the most solid and prevalent move of the conclusion section, can be very powerful, as reflected by its exorbitantly high rate of occurrence.

To focus on the comparison of the organizational structure obtained with that of the original taxonomy (Ruiying & Allision, 2003), a level of congruence can be observed. As detailed in Table 2, Move 2 and Move 3 and their pertaining steps are identical to the original taxonomy, indicating their relative stability over the two decades. It needs to be noted that this study and the previous studies on the conclusion section reviewed earlier (Morales, 2012; Zamani & Ebadi, 2016) were anchored on the same analytical framework of Ruiyang and Allison's 2003 model. Interestingly, the occurrence of the steps associated with Move 1 was not reported. However, the significance of these steps of Move 1 cannot be underestimated either. In this regard, two interpretations of this phenomenon are possible. First, it is plausible

Table 1

Conclusion Taxonomy Depicting Move/Step Frequencies and Move Status (n = 25)

Move/Step	Move Status	Frequency (Percent)
Move 1: Summarizing the study	Obligatory	25 (100%)
Step 1.1: Restating background		12
Step 1.2: Restating objectives		12
Step 1.3: Restating methodology		11
Step 1.4: Restating findings		21
Move 2: Evaluating the study	Conventional	17 (68%)
Step 2.1: Indicating significance		6
Step 2.2: Indicating limitations		3
Step 2.3: Evaluating methodology		8
Move Move 3: Deducting from the study	Conventional	17 (68%)
Step 3.1. Recommending further studies		9
Ste Step 3.2: Drawing implications		8

Table 2

Quantitative Data of Individual Conclusions: Move Sequence and Step Presence (n = 25)

No.	Sequence	Move 1				Move 2			Move 3	
		1.1	1.2	1.3	1.4	2.1.	2.2	2.3	3.1	3.2
1.	12		√		√	√				
2.	123	√	√			√			√	
3.	123	√		√	√		√		√	
4.	1		√	√	√					
5.	123			√	√			√	√	
6.	123		√	√	√			√	√	
7.	12		√		√	√		√		
8.	123	√	√	√	√			√	√	
9.	12		√		√			√		
10.	123	√	√			√				√
11.	13				√					√
12.	13	√								√
13.	123	√			√			√	√	
14.	123	√			√	√				√
15.	123				√		√		√	
16.	13		√		√					√
17.	13		√	√	√				√	
18.	123				√	√			√	
19.	1				√					
20.	12			√	√			√		
21.	12	√		√	√			√		
22.	13	√			√					√
23.	13	√	√	√						√
24.	123			√	√		√			√
25.	1		√	√	√					
		10	12	11	21	6	3	8	9	8

that the previous studies concentrated on cross-cultural comparison analysis of the rhetorical structure of the section across languages. Consequently, adhering to the original 2003 taxonomy would facilitate the comparison of the two cultural contexts. Second, given that the full scheme of the 2003 model was meant for the combined result-discussion-conclusion section, adopting the model to analyze the stand-alone conclusion section might not be precisely justifiable. These interpretations, however, remain speculative and need to be elucidated by subsequent research studies.

At this juncture, it is evident that Move 1 Step 4 (*Restating findings*) seems to be the most vital, with the highest occurrence rate of 84% (in contrast with the other steps, of which their frequencies are quite low or relatively peripheral in the dataset). A pertinent question arises: What are some of the justifications for this phenomenon? To address this question, at the move level, the relatively high occurrence rate of Move 1, compared with those of Moves 2 and 3, suggests that this move is not serendipitous but quite pervasive. At the step level, the high occurrence rate of Move 1 Step 4 (*Restating findings*) may be attributed to its function of highlighting specific findings, acting as a crucial springboard and providing a core element for further arguments to be developed in Move 2 and Move 3. The scrutiny of the dataset reveals that Step 4 of Move 1 co-occurred with at least another step of this move in all (but one) conclusions, implying that the findings must be stated scrupulously to ensure that

they are not only contextualized but also solidly and convincingly presented. In this scenario, the first three steps of Move 1 could thus be employed to set the scene, contextualizing the announcement of findings in Step 4. Overall, the findings suggest that Move 1 is potentially the most central, powerful, and established. Meanwhile, Step 4 of Move 1 is distinctly pivotal, paving the way for subsequent arguments in Move 2 and Move 3.

Move 2 aims to evaluate various facets of the study being presented and Move 3 projects to the future. The existence of these tactics indicates that scholars need to possess critical thinking skills, an essential asset to execute these moves. Since this section comes at the end of the RA, it possibly serves as a closing reminder to the audience about the points the writers wish to restate and to make a strong final positive impression on the audience. This section effectively conveys a sense of completeness by providing a synthesis of the main ideas in Move 2 while highlighting the important takeaways from the study in Move 3, which projects future directions, calls for actions, or offers an overview of potential future research topics.

Another pertinent question is what triggers the need to end a RA with powerful closure. As remarked by Musselin (2018), the academic environment has become increasingly competitive. Therefore, every effort has been simultaneously mobilized to enhance educational institutions' competitiveness. This has led to the emergence of new competition formats, channels, platforms, and a new type of competition. In response to this competitiveness, RA publications have been employed as one of the measures to determine the standing of higher education institutions. Plausibly, quality rivalry in the RA genre has spawned a more elaborate rhetorical structure, encouraging scholars to emphasize Move 1 by including the four steps and further develop arguments analytically and futuristically using Move 2 and Move 3. As such, the conclusion section represents the writers' final opportunity to capitalize on it by making a lasting positive impression on the audience.

Finally, this study has provided a current structural organization of the stand-alone conclusion section in open-access journals from four academic disciplines. Pedagogically, the structural pattern can substantially facilitate the task of RA writing. As illustrated in this study, the conclusion section incorporates moves and steps. The findings imply that without the pattern delineated, it is highly likely that novice scholars will find the RA writing task overwhelming, demanding, and challenging. They will struggle to figure out what content to include in the conclusion section and how to linguistically express it. With the patterns displaying moves and steps and sample excerpts of individual moves and steps, the task can be substantially scaffolded. Furthermore, a greater understanding of the rhetorical structure can contribute to the more appealing nature of the RAs and escalate the community's level of competition. Specifically, with particular attention to the conclusion section, the last opportunity for the RA writers to manifest or express their strife for competitiveness will be successfully exploited.

CONCLUSION

This study aims to accomplish two goals: identifying the organizational structure of 25 stand-alone conclusions from open-access journals using Swales' genre analysis (2004) and determining structural differences in comparison to Ruiying and Allison's 2003 taxonomy. The findings, which are partly consistent with previous research, show that the section comprises three major moves and several steps. The findings, however, shed light on an intriguing feature of Move 1, which was discovered to be more prominent in terms of frequency of occurrence and more extensive in terms of the numerous steps it contains. Based on these findings, a current version of the RA conclusion taxonomy is proposed to reflect the current writing practices of this section.

At this juncture, given the narrow focus on four academic disciplines and award-winning RAs from open-access journals from one commercial publisher, and although the dataset was systematically compiled, the resulting taxonomy cannot claim to be utterly representative of "the conclusion section." The modified taxonomy proposed in this study needs to be tested or verified by additional studies in other disciplines and on a larger dataset to determine its validity. It should be noted that this study did not attempt to link the award-winning status to the quality of the conclusion section. Finally, future research is needed to fine-tune this modified taxonomy.

Notwithstanding these limitations, pedagogical implications could be drawn from the current findings to benefit and empower early-career researchers embarking on their academic journey, particularly aspiring to publish in international and prestigious journals. As elucidated in this study, in addition to the skills in conducting research, the knowledge to use academic language and expressions, and the tactics for powerfully reporting scientific discoveries, the ability to stage persuasive arguments in the conclusion section cannot be overlooked. This particular skill must be systematically cultivated, fostered, and nurtured. These implications are bolstered by the fact that, currently, many prestigious journals strongly recommend the inclusion of this section, be they open-access or non-open-access. Hence, a better understanding of how the conclusion section is constructed and regulated will be pedagogically advantageous, assisting novice scholars to successfully embark on an academic journey and sharpening their awareness of academic conventions or expectations.

ACKNOWLEDGMENTS

The author would like to express sincere appreciation to the Faculty of Management Science, Silpakorn University, which facilitated the conduct and publication of this study.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

Adel, S. M. R., & Ghorbani Moghadam, R. (2015). A comparison of moves in conclusion sections of research articles in psychology, Persian literature and applied linguistics. *Teaching English Language*, *9*(2), 167-191. https://dx.doi.org/10.22132/TEL.2015.53729

Affandi, A. M., Ali, S., Yazid, B., & Nunuk, W. A. (2021). Previous studies have several limitations ….: Indonesian doctoral students', Indonesian academics', and international authors' research gap strategies in ELT research article abstracts and introductions. *Journal of Language and Education*, *7*(2), 25-44. https:// dx.doi.org/10.17323/jle.2021.11735

Ahmadi, L. (2022). Rhetorical structure of applied linguistics research article discussions: A comparative cross-cultural analysis. *Journal of Language and Education*, *8*(3), 11-22. https://dx.doi.org/10.17323/jle.2022.17520

Amirian, Z., Kassaian, Z., & Tavakoli, M. (2008). Genre analysis: An investigation of the discussion sections of applied linguistics research articles. *The Asian ESP Journal*, *4*(1), 39-63.

Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in dentistry and disciplinary variation. *Journal of English for Academic Purposes*, *11*(2), 134-144. https:// dx.doi.org/10.1016/j.jeap.2011.10.004

Bruce, I. (2009). Results sections in sociology and organic chemistry articles: A genre analysis. *English for Specific Purposes*, *28*(2), 105-124. https://dx.doi.org/10.1016/j.esp.2008.12.005

Chen, T. Y. & Kuo, C. H. (2012). A genre-based analysis of the information structure of master's these in applied linguistics. *The Asian ESP Journal, 8*(1), 24-52.

Compagnone, A. (2015). The reconceptualization of academic discourse as a professional practice in the digital age: A critical genre analysis of TED Talks. *HERMES-Journal of Language and Communication in Business*, (54), 49-69. https:// dx.doi.org/10.7146/hjlcb.v27i54.22947

Cotos, E., Huffman, S., & Link, S. (2017). A move/step model for methods sections: Demonstrating rigour and credibility. *English for Specific Purposes*, *46*, 90-106. dx.https://dx.doi.org/10.1016/j.esp.2017.01.001

Cross, C., & Oppenheim, C. (2006). A genre analysis of scientific abstracts. *Journal of Documentation*, *62*(4), 428-446. https://dx.doi.org/10.1108/00220410610700953

Danquah, M. B., Francis, T., Slippe, D. P., & Adentwi-Hayford, D. (2020). A genre analysis of electronic manuals. *AU eJournal of Interdisciplinary Research*, *5*(1), 33-43.

Elena, Z. (2017). Move structure of research article abstracts on management: Contrastive study (the case of English and Russian). *Journal of Language and Education*, *3*(2), 63-72. https://dx.doi.org/10.17323/2411-7390-2017-3-2-63-72

Hu, G., & Liu, Y. (2018). Three-minute thesis presentations as an academic genre: A cross-disciplinary study of genre moves. *Journal of English for Academic Purposes*, *35*, 16-30. https://dx.doi.org/10.1016/j.jeap.2018.06.004

Juan, Z. H. A. O., & Tao, W. U. (2013). A genre analysis of medical abstracts by Chinese and English native speakers. *Journal of Medical Colleges of PLA*, *28*(1), 60-64. https://dx.doi.org/10.1016/S1000-1948(13)60018-0

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

Kanoksilapatham, B. (2007). Writing scientific research articles in Thai and English: Similarities and differences. *Silpakorn University International Journal*, *7*, 172-203.

Kanoksilapatham, B. (2012). Research article structure of research article introductions in three engineering subdisciplines. *IEEE Transactions on Professional Communication*, *55*(4), 294-309. https://dx.doi.org/10.1109/TPC.2012.2223252

Kanoksilapatham, B. (2013). Generic characterisation of civil engineering research article abstracts. *3L: Southeast Asian Journal of English Language Studies*, *19*(3), 1-10.

Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes*, *37*, 74-86. https://dx.doi.org/10.1016/j.esp.2014.06.008

Loi, C. K., Evans, M. S., Lim, J. M. H., & Akkakoson, S. (2016). A comparison between Malay and English research article discussions: A move analysis. *Sage Open, 6*(2), 1-11. https://dx.doi.org/10.1177/2158244016652925

Martín, P. M. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes*, *22*(1), 25-43. https://dx.doi.org/10.1016/S0889-4906(01)00033-3

Mering, M., & Hoeve, C. D. (2020). A brief history to the future of open-access. *Serials Review, 46*(4), 300-304. https://dx.doi.org/10.1080/00987913.2020.1850041

Morales, R. C. (2012). Conclusions in research articles: A Filipino-Japanese contrastive rhetoric study. *Philippines ESL Journal*, *8*, 83-95.

Moreno, A. I. (2021). Selling research in RA discussion sections through English and Spanish: An intercultural rhetoric approach. *English for Specific Purposes*, *63*, 1-17. https://dx.doi.org/10.1016/j.esp.2021.02.002

Musselin, C. (2018). New forms of competition in higher education. *Socio-Economic Review*, *16*(3), 657-683. https://dx.doi.org/ https://doi.org/10.1093/ser/mwy033

Pratiwi, S. N., & Kurniawan, E. (2021). Rhetorical move and genre knowledge development of English and Indonesian abstracts: A comparative analysis. *Studies in English Language and Education*, *8*(3), 885-900.

Qian, Y. (2020). A critical genre analysis of MD&A discourse in corporate annual reports. *Discourse & Communication*, *14*(4), 424-437. https://dx.doi.org/10.1177/1750481320910525

Ruiying, Y., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes*, *22*(4), 365-385. https://dx.doi.org/10.1016/S08894906(02)00026-1

Salmaso, G. (2017). Story genres in SFL: A more flexible taxonomy. Extrapolating a taxonomy of story genres in Spanish to story genres in English. *Journal of Language and Education*, *3*(1), 6-22. https://dx.doi.org/10.17323/2411-7390-2017-3-1-6-22

Stoller, F. L., & Robinson, M. S. (2013). Chemistry journal articles: An interdisciplinary approach to move analysis with pedagogical aims. *English for Specific Purposes*, *32*(1), 45-57. https://dx.doi.org/10.1016/j.esp.2012.09.001

Suherdi, D., Kurniawan, E., & Lubis, A. H. (2020). A genre analysis of research article 'findings and discussion' sections written by Indonesian undergraduate EFL students. *Indonesian Journal of Applied Linguistics*, *10*(1), 59-72.

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

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

Viera, C., & Williams, S. A. (2020). Corpus analysis of engagement discourse strategies in academic presentations. *Research in Corpus Linguistics*. https://dx.doi.org/10.32714/ricl.08.01.07

Yang, W. (2012). A genre analysis of PhD dissertation acknowledgements. *LSP Journal-Language for Special Purposes, Professional Communication, Knowledge Management and Cognition*, *3*(2).

Zamani, G., & Ebadi, S. (2016). Move analysis of the conclusion sections of research papers in Persian and English. *Cypriot Journal of Educational Sciences*, *11*(1), 9-20.

https://doi.org/10.17323/jle.2023.16044

A Corpus-Based Investigation of Phrasal Complexity Features and Rhetorical Functions in Data Commentary

Muhammed Parviz [1], **Ge Lan** [2]

[1] Imam Ali University, Tehran, Iran

[2] City University of Hong Kong, Hong Kong SAR, China

ABSTRACT

Introduction: In academic written texts, linguistic and rhetorical features are often interactively used as a vehicle for writers to construct their texts in order to accomplish specific communicative purposes. However, the effective integration of these resources may pose challenges for developing writers.

Purpose: This study employed a corpus-based genre analysis approach to investigate phrasal complexity features and rhetorical functions in data commentaries written by Iranian undergraduate and graduate students. Through this approach, we aimed to examine a relatively unexplored genre of data commentary in terms of its phrasal complexity features, rhetorical functions, and their relationships. By analyzing these relationships, we sought to provide insights into the writing practices of Iranian undergraduate and graduate students in the context of data commentaries.

Method: This study employed a convenient sampling method to select a total of 76 university students, which included 47 undergraduate students and 29 graduate students. The participants were involved in generating a corpus of 380 data commentaries, which were then thoroughly examined and compared. To identify instances of phrasal complexity features, the researchers utilized the AntConc software tool, applying regular expressions (regex) to extract potential occurrences. Additionally, a Python program was developed and implemented to calculate the frequencies of the identified PCFs. The researchers manually annotated the rhetorical function of the data commentaries to determine their specific usage.

Results: Statistical analysis such as Mann Whitney U test and Spearman correlation test, revealed that graduate students significantly utilized more phrasal complexity features including attributive adjectives, nominalizations, and prepositional phrases (of) compared to undergraduate students. However, a qualitative analysis showed that the use of these linguistic features is influenced by the writing topics. Regarding rhetorical functions, graduate students used more moves and/or steps related to presenting and commenting data, while undergraduate students produced more moves or steps concerning personal asides. Moreover, certain phrasal complexity features and the moves and/or steps were found to be correlated, aligning with recent corpus-based studies.

Conclusion: The study concludes with pedagogical implications.

KEYWORDS

phrasal complexity, rhetorical function, academic writing, data commentary, corpus-based approach

Citation: Parviz M., & Lan G. (2023). A Corpus-based Investigation of Phrasal Complexity Features and Rhetorical Functions in Data Commentary. *Journal of Language and Education, 9*(3), 90-109. https://doi.org/10.17323/jle.2023.16044

Correspondence:
Muhammed Parviz,
mohammad.parviz60@gmail.com

Received: September 27, 2022
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

As an essential medium for developing scientific knowledge, academic writing is seen as an indispensable part of knowledge construction. For this reason, learning academic writing is a demanding job due to the challenges it may pose concerning how to transform knowledge into a disciplinarily appropriate entity for

a particular community. Chief among multifarious linguistic resources closely associated with academic writing, phrasal complexity features (PCFs) are considered an important group of linguistic features for academic discourse (Biber & Gray, 2010). Researchers have recently focused on investigating PCFs in various written genres such as argumentative papers (Lan & Sun., 2019), research articles (Parviz et al., 2020), and MA theses (Parkinson & Musgrave, 2014). Closely intertwined with linguistic features are rhetorical functions. In academic written texts, linguistic and rhetorical features are often interactively used as a vehicle for writers to construct their texts in order to accomplish specific communicative purposes. However, the effective integration of the linguistic and rhetorical resources in academic writing may pose challenges to developing writers such as undergraduate students. Among the existing research on grammatical complexity features (including PCFs), a functional-register approach has been frequently used to interpret the rhetorical functions of selected grammatical features in academic written texts (Biber et al., 2022).

In terms of second language (L2) writing research, most existing research focuses on some common genres such as argumentative papers, theses/dissertations, and exam essays. However, as an important written genre, data commentary (DC) has been under-represented in L2 writing research on PCFs and their rhetorical functions. DC is a multimodal way of presenting data and is described as "the verbal comments on visual materials" (Nordrum & Eriksson, 2015, p. 59). DC is important to student writers, especially L2 students, because 1) it can be found in diverse academic written genres (e.g., case studies, research articles, and business proposals), 2) it also includes major writing techniques such as presenting and interpreting experimental data and statistical evidence in visual elements (e.g., images and diagrams), which student writers need to acquire, and 3) recent studies have demonstrated that visual renderings, articulating, and formulating focal points in a multimodal context such as a data commentary can be an arduous undertaking to be experienced and written by numerous student writers and may present a challenge to both proficient writer students and teachers in science fields (Eriksson & Nordrum, 2018; Jalilifar et al., 2019; Nordrum & Eriksson, 2015; Sancho Guinda, 2012; Wharton, 2012, to name but a few). Therefore, the aims of the present study are (a) to compare PCFs and the rhetorical functions produced by Iranian undergraduate and graduate students, and more importantly (b) to analyze the relationship between the PCFs and the rhetorical functions in data commentary.

LITERATURE REVIEW

The Importance of PCFs in Academic Writing

Phrasal complexity features have recently been included as an important subset of grammatical complexity features in many studies in L2 writing and academic writing (e.g., Staples et al., 2016; Parviz et al., 2020). The increasing studies on the PCFs emanate from two important roles they play in writing research. First, PCFs contribute to the expansive representation of grammatical complexity in writing research. In writing studies, scholars argued the integration of PCFs into the analysis of grammatical complexity. Previously, T-units-based measures[1] have been dominated in applied linguistics to assess grammatical complexity of academic writing (Biber et al., 2011). Then, scholars (e.g., Biber & Gray, 2010; Staples et al., 2016) expressed their concerns over analyzing L2 and academic writing with only T-unit-based measures, and this resulted in a call for integrating phrasal complexity[2] into the analysis of grammatical complexity. As the seminal work on phrasal complexity, Biber et al. (2011) proposed a developmental index of writing complexity features, operationalizing phrasal complexity into a set of phrasal features that functions as noun modifiers (e.g., attributive adjectives, pre-modifying nouns, and prepositional phrase). The integration of these PCFs, which represents phrasal complexity, supplements existing measures/features of complexity at the clausal level to make the representation of grammatical complexity more complete in L2 writing research.

Second, the importance of PCFs in academic writing also arises from their close relationship with writing development. Empirical studies have demonstrated that PCFs are often used to create compressed NPs, which is a grammatical characteristic of indicating the advanced stage of writing development (Biber et al., 2011). Scholars conducted large-scale corpus-based studies to show that in academic written registers, there are co-occurring patterns of PCFs to construct compressed NPs (Biber, 1988). Then, researchers added a functional interpretation to the co-occurrence: the PCFs fulfill an important function, packing intensive information in a compact space (Parkinson & Musgrave, 2014). For example, the NP, "*a solid research plan on the topic of grammatical complexity in second language writing*", only includes PCFs, for instance attributive adjectives (e.g., solid) and prepositional phrases as postmodifiers (e.g., of grammatical complexity). The NP carries rich information to expand the meaning of the head noun (i.e., *plan*) in a compressed structure. Researchers claim that when stu-

[1] A T-unit refers to a main clause and all dependent clauses attached to the main clause, which only capture grammatical complexity based on subordinate clauses.

[2] Phrasal complexity primarily refers to compressed noun phrases (NPs), which is the grammatical structure that head nouns are modified by phrasal features, such as premodifying nouns and prepositional phrases (Biber & Gray, 2010).

dents become academically advanced, they tend to use more PCFs, which indicates their writing development (Staples et al., 2016).

Because of the importance of PCFs, there has been a growing research trend of this grammatical structure in L2 and academic writing from 2011 to present. Based on Biber et al. (2011), many relevant studies were conducted in diverse academic written genres, including but not limited to EAP essays (Parkinson & Musgrave, 2014), research articles (Parviz et al., 2020) argumentative essays (Lan & Sun, 2019), PhD dissertations (Ansarifar et al., 2018), first-year composition (Staples & Reppen, 2016), and diverse genres in British Academic Written English Corpus (Staples et al., 2016).

Rhetorical Functions[3] in Academic Writing

In discourse analysis, rhetorical functions have been considered a key component of academic writing. Over the past three decades, rhetorical functions within the domain of genre analysis have captured the attention of genre analysts and rhetoricians, many of which have been based on Swales' seminal work. Swales (1990) provided a widely used analytical framework to analyze discourse structure of a genre, including rhetorical moves and linguistic features. Swales (1990) also discussed communicative moves, which have generated many empirical studies on varied genres, in which academic written genres are a major part. These academic genres include but not limited to textbooks (e.g., Hyland, 2000), theses and dissertations (e.g., Samraj, 2008), research articles (e.g., Yang & Allison, 2003; Parviz, 2023), among others. These studies have shared profound insights into the schematic patterns, discoursal features, and linguistic-textual mechanisms deployed in diverse academic disciplines. The analyses of rhetorical functions, furthermore, provide extensive implications for the teaching and learning of L2 writing. The successful use of rhetorical functions has been regarded as a characteristic of "good-standard" academic and L2 writing.

Furthermore, grammatical features play an important role in the analyses of rhetorical functions of academic writing. They serve as a medium for achieving specific rhetorical functions, making it critical to build a connection between rhetorical functions and grammatical options to enhance L2 students' learning of academic writing (Charles, 2007). While many studies have attempted to build this connection, only a few have examined the connection between rhetorical functions and NPs and PCFs in academic writing. For example, Jiang and Hyland (2015) conducted an analysis on shell nouns to analyze how this structure is used in research articles to express stances across multiple disciplines. More recently, Jiang and Hyland (2021) compared patterns of metadiscursive nouns, which are often used to express viewpoints on research and to interact with intended audiences as members of disciplinary communities.

In register analysis, researchers have applied a register-functional approach to study both grammatical forms and their rhetorical functions based on large-scale corpora (Biber et al., 2022). For instance, Staples and Reppen (2016) compared grammatical complexity features in two academic genres such as argumentative papers and rhetorical analysis. The rhetorical functions of two PCFs (attributive adjectives and premodifying nouns) were interpreted in their study. They found that more premodifying nouns are used in argumentative papers, "reflecting the greater complexity of topics and relationships being expressed, as well as the need to provide more informational support than in the rhetorical analysis" (Staples & Reppen, 2016, p. 28-29). Staples et al. (2016) also explored grammatical complexity (including PCFs) in diverse written genres across four institutional levels in the university setting. They noted that writers at higher institutional levels used more PCFs to meet "increased information packaging demands as they gain disciplinary content knowledge and are expected to convey this knowledge concisely for their audiences" (p. 77).

Research Gaps and Research Questions

This study was conducted in a relatively new genre of data commentary, which has not been explored substantially in terms of phrasal complexity features, rhetorical functions, and their relationships. In this study, we analyzed the use of phrasal complexity features and the rhetorical functions and their relationships in data commentaries written by Iranian undergraduate and graduate students.

Data commentary is selected as the academic written genre due to two specific reasons. First, it has been considered a challenging genre to not only L2 students but also university students in general (Eriksson & Nordrum, 2018; Nordrum & Erikson, 2015). The primary reason for this may be the multimodal nature of data commentary via which writers need to incorporate both textual and visual information in academic and scientific texts (Swales & Feak, 1997). Second, it could be considered a common genre across multiple written genres in university settings. A number of academic written genres across various disciplines can include data commentary, such as case analyses in economics, scientific reports in chemistry, and research articles in applied linguistics. Generally, data commentary is commonly organized in a general-to-specific pattern, with the overall structure classified into location elements and/or summary statements, as well as highlighting statements, discussions of implications,

[3] In this article, we used the terms "rhetorical functions", "rhetorical moves,» «communicative moves,» and «moves and steps» interchangeably to refer to the various strategic techniques that speakers use to achieve their communicative goals. While each term conveys a slightly different nuance, they all refer to similar concepts.

problems, exceptions, recommendations, or other notable aspects of the data (Swales & Feak, 2012).

Locatives, considered an important element, are typically employed as opening statements in data commentary. Their purpose is to indicate the position of the data, providing a succinct overview of findings through visual information. Locatives direct readers' attention to the visual display of information before proceeding to the main text (Swales & Feak, 2012). Following are examples of locatives (*italicized in examples a & b*) which can elucidate how this overarching element is linguistically employed.

> *Table 5 shows the types of internet misbehavior common among university students.*
>
> *Figure 2 shows a honeycomb solid oxide fuel cell (SOFC) unit with air cooling paths.*
>
> *(Adapted from Swales & Feak, 2012, p. 147)*

Location elements, also known as "endophoric markers", play a crucial role in assisting readers in identifying important information presented in visual materials such as tables, figures, charts, and diagrams (Hyland, 2019). As a type of metadiscourse resource, endophorics can also direct the reader's attention to explanatory and interconnected sections of a text. They are often positioned at the end of a sentence using passive voice, as exemplified below (Hyland, 2019; Swales & Feak, 2012).

> *a. The types of internet misbehavior common among university students are shown in Table 4*
>
> *b. A honeycomb solid oxide fuel cell (SOFC) unit with air cooling paths is shown in Figure 2.*
>
> *(Adapted from Swales & Feak, 2012, p. 147)*

Another integral aspect of a data commentary is the use of highlighting statements. These statements are used to emphasize key points that are supported by the data (Swales & Feak, 2012). Highlighting statements also provide opportunities to demonstrate the ability to recognize "trends or regularities in the data, to separate more important findings from less important ones, and to make claims of appropriate strength" (Swales & Feak, 2012, p. 156). In the process of writing highlighting statements, it is crucial to exercise "caution and critical thinking towards the data" while effectively employing appropriate linguistic devices to express this cautious approach (Swales & Feak, 2012, p. 156). However, the concluding section of a data commentary can be challenging, as it requires presenting significant insights derived from the data, distinguishing experienced writers from novices (Swales & Feak, 2012). To tackle this challenge and demonstrate expertise, Swales and Feak (2012, p. 172) proposed incorporating certain elements in the conclusion of a data commentary such as "explanations and/or implications of the data and explanation of the reasoning process that led to the conclusions". Overall, Swales and Feak (2012, pp. 140-141) listed some of the more common purposes of data commentary as shown in Table 1. They highlighted that while many of the aforementioned objectives are typically addressed in a data commentary, it can be challenging to precisely specify the requirements of this genre. However, despite its importance, this genre has receive limited research attention regarding the analysis of PCFs and rhetorical functions.

Given the widespread presence of data commentary across disciplines, there is a need for further research to explore and examine the use of PCFs within this under-represented genre. By investigating the presence and functions of PCFs within data commentary, scholars can uncover valuable insights into how these linguistic features contribute to the overall rhetorical effectiveness of written works in different academic domains. This research would greatly benefit our understanding of language usage and provide guidance for students and academics in mastering the skills required for effective data commentary writing in their respective fields. In addition, it is important to mention that individual PCFs and their corresponding rhetorical functions have been explored in some studies (e.g., Staples & Reppen, 2016). The analyses in such studies are often based on analysis of specific texts to interpret rhetorical functions of individual PCFs. While acknowledging the importance of these analyses, it is also meaningful to detect the connection between PCFs and rhetorical functions from a quantitative perspective to triangulate the existing findings. In this study, we quantified the relationship between the frequencies of PCFs and the

Table 1

Common Purposes of Data Commentary Proposed by Swales and Feak (2012)

Highlight the results of research
Use the data to support a point or make an argument in your paper
Assess theory, common beliefs, or general practice in light of the given data
Compare and evaluate different sets
Assess the reliability of the data in terms of the methodology that produced it
Discuss the implications of the data
Make recommendations

Note. From *"Academic Writing for Graduate Students: Essential Tasks and Skills,"* by J. M. Swales & C. B. Feak, 2012, University of Michigan Press. Copyright 2012 by University of Michigan Press.

frequencies of rhetorical functions (i.e., moves and steps). Recently, this method has been applied in academic writing research. For instance, Lu et al. (2021b) analyzed the relationship between phraseological features and rhetorical functions in research articles. Consequently, our study would supplement the existing analyses from the qualitative analysis on PCFs and their rhetorical functions, presenting the importance of PCFs in the under-explored genre, data commentary.

Finally, we conducted a comparison between the data commentaries produced by undergraduate and graduate students in order to uncover potential linguistic similarities and differences between the two groups. This comparative analysis aimed to shed light on how these groups employ phrasal complexity features and rhetorical functions in their writing. Understanding this comparison is crucial because all student writers, regardless of their academic level, rely on specific linguistic and rhetorical strategies or writing patterns to achieve their writing goals (Lavelle & Bushrow, 2007). Given that undergraduate students are typically in the nascent stages of their academic careers, and they may encounter data commentaries for the first time, it is essential to understand their linguistic and rhetorical approaches in order to provide the necessary guidance and assistance tailored to their needs. On the other hand, graduate students face the complex and often novel task of academic writing at the graduate level (Lavelle & Bushrow, 2007). With their prior experience gained during their undergraduate studies, these students are expected to possess advanced writing skills. By undertaking this comparative analysis, our goal is to shed light on how undergraduate and graduate students utilize phrasal complexity features and rhetorical functions in their data commentaries. This knowledge may empower us to provide targeted support and appropriate resources to both groups, assisting them in improving their writing skills and excelling in their respective academic pursuits. Additionally, this analysis offers valuable insights into academic writing development, allowing for the assessment of proficiency levels, identification of developmental patterns, examination of genre conventions, and contributing to the advancement of writing pedagogy. Accordingly, the following research questions are formulated:

(1) How do undergraduate and graduate students utilize PCFs in data commentaries they produce? Which specific types of PCFs are more prominently used by each group?

(2) How do undergraduate and graduate students employ rhetorical functions in data commentaries they produce? Which specific types of rhetorical functions are more prominently used by each group?

(3) Is there a correlation between the presence of PCFs and the occurrence of rhetorical functions within both the undergraduate and graduate data commentaries?

METHODS

Participants and Context

The present study was conducted at an Iranian state-run university during the winter semester of 2019-2020, which coincided with a severe outbreak of Covid-19 pandemic. Due to the sudden shift from traditional classrooms to on-line education compelled by the pandemic, the authors of this study were required to adjust all their academic activities and face-to-face classes to on-line mode of delivery. Given the challenging circumstances, participant selection for this study was conducted using a convenient sampling method. The researchers selected participants from three existing intact groups, with the first author of the study serving as the instructor for these groups. It is important to emphasize that random assignment or manipulation of group compositions was not employed during the participant selection process.

In this study, a total of 76 students participated, consisting of 47 undergraduate students and 29 graduate students. Among the graduate students, there were 23 MA students and 6 PhD students. The participants included 51 female students and 25 male students, all of whom were native speakers of Persian majoring in English Language Teaching (ELT). The selected participants were enrolled in different writing courses specifically tailored to their academic level. The courses included "Essay Writing" for undergraduate students, "Academic Writing" for MA students, and "Advanced Academic Writing" for PhD students. These courses, which were offered by the English department of the university, were two-credit online writing courses that spanned over the course of the winter semester, with a total of 16 sessions. These courses were conducted online, with each group having one class per week. Each session lasted approximately 90 minutes.

It is also essential to consider the English language proficiency of the participants and the admission process for Iranian students in undergraduate and graduate studies. Iranian students undergo a rigorous and highly competitive university entrance examination (UEE) to gain admission to BA and MA programs. This national examination evaluates their content knowledge as well as their language skills, including reading, grammar, and vocabulary. Admission to doctoral programs also involves meeting strict academic standards, which include performing well in the UEE and evaluating their research achievements during their master's studies. Therefore, all participants have passed the Iranian nationwide BA, MA, and PhD university entrance examinations, which serve as gatekeeping tests for selection purposes.

Based on their academic background and experiences, it is reasonable to assume that they possess a high command of academic English. Detailed demographic information about the participating students is presented in Table 2.

Furthermore, it should be noted that both undergraduate and graduate students have already completed extensive academic coursework. They have taken various prerequisite writing courses during their undergraduate and graduate studies such as "Introduction to Writing," "Paragraph Development," "Letter Writing," "Essay Writing," and "Advanced Writing." However, the participants' knowledge and experience in writing data commentary may vary, primarily because the genre of data commentary is not commonly taught at the undergraduate level. As a result, undergraduate students (BA) might have prior experience with this type of writing. In contrast, graduate students (MA and PhD), given their advanced academic training, were expected to have a higher level of familiarity and experience in writing data commentary. It is important to acknowledge that the differing levels of experience between undergraduate and graduate students could potentially influence their approach and proficiency in writing data commentary. However, this difference was not considered or analyzed in the current study.

Task

In order to examine phrasal complexity features and rhetorical functions in the data commentaries, a set of 30 suggested topics emphasizing data commentary was initially chosen. These topics were sourced from publicly accessible IELTS materials associated with the IELTS Academic Module writing Task 1 (Cullen et al., 2014), which are seen highly relevant to data commentary writing. The visual nature of this task also aligns with academic writing components, as its discourse modes often mirror those found in authentic university assignments and "reflect some of the features of academic language" (IELTS, 2017)[4]. To ensure impartiality in the writing prompts, we then carefully selected 20 topics that steer clear of sensitive issues related to politics, religion, race, culture, and controversy (samples provided in Appendix A). The visual materials accompanying these prompts were chosen for their familiarity and relevance to daily life, eliminating the need for specialized knowledge.

Procedure

To maintain ethical standards, participants were given a clear explanation of the objectives of the study at the outset. Prior to their involvement, participants were required to provide verbal consent, signifying their willingness to participate and ensuring their rights as subjects were respected. To ensure the research process was feasible and aligned with the participants' busy academic schedules, a pre-selected set of 20 visual materials was utilized. These materials were integrated as in-class activities during eight separate online sessions of the writing course over an eight-week period. This approach aimed to strike a balance between accommodating the participants' demanding academic commitments and maintaining the practicality of the research study. The visual materials included bar/pie charts, diagrams, and line graphs, covering various general topics such as "*adult education, global illiteracy rates, leisure time, food budgets, mobile phones, building construction, cinema attendance, London museums, public transportation, and work performance*". Participants were then required to individually write about their preferred topic in the form of a data commentary, with a minimum length of 150 words and a time limit of 30 minutes. To prevent topic disclosure, participants were not informed beforehand of the specific topic for each session, and it was assumed that the selected visual materials were of general interest and familiar to all participants, aligning closely with their everyday reality (Sancho Guinda, 2012). No participant was allowed to work collectively or in pairs in order to ensure an accurate gauge of how they exploited phrasal complexity features and rhetorical functions in the data commentaries. The commentary writing tasks were presented without any supporting background materials, enabling participants to engage in spontaneous writing without requiring any field-specific knowledge. These data commentary writing tasks were also written without any direct explicit instruction on phrasal complexity features and rhetorical instructions and were gathered as part of the coursework, with no impact on participants' final assess-

Table. 2

Demographics of the Participating Students

Educational Levels	No of the Participating Students	Gender		Total
		Male	Female	
BA	47	12	35	
Master's	23	10	13	76
PhD	6	3	3	

[4] IELTS (2017). IELTS website. Retrieved from: https://www.ielts.org.

ments. After collecting the data commentaries produced by undergraduate and graduate students during the sessions (as shown in Table 3), a corpus was constructed consisting of 380 files. The total number of tokens in the corpus was 72,591 words with an average length of 191.03 tokens. The corpus included two sub-corpora, one for undergraduate students (206 files), and another for graduate students (174 files). It is worth noting that not all graduate students were able to complete all tasks, as some individuals were unable to attend all online sessions of the course.

Linguistic Model of PCFs

The PCFs in the study are derived from the hypothesized index of writing complexity features proposed by Biber et al. (2011). These PCFs are presented in Table 4. The grammatical functions of these PCFs are primarily noun modifiers, except for nominalizations, which can be either head nouns or noun modifiers. The PCFs consist of two pre-noun modifiers (i.e., attributive adjectives and premodifying nouns) and three post-noun modifiers (e.g., prepositional phrases (*of*), prepositional phrases (*others*), and appositive noun phrases). The noun modifiers are at different stages of Biber et al.'s (2011) hypothesized stages of writing development: attributive adjectives and premodifying nouns are at the stages of 2 and 3, respectively, whereas the two types of preposition phrases and appositive noun phrases are at the stages of 4 and 5. Nominalizations are not placed in any stages.

Corpus Processing and PCFs Extraction

The corpus processing involved five steps. First, the original files in the corpus were converted into plain text format using the *AntFileConveter*. Second, the converted files were tagged with part-of-speech (POS) information via the *TagAnt*. The PCFs are based on three core linguistic features: nouns, adjectives, and prepositions. A qualitative check was conducted on precision and recall of the three features to ensure the accuracy of the tags. The rates of precision and recall were all above 90% for the three core linguistic features. Thus, we did not fix the POS tags of the three linguistic features in all files (see Appendix B).

Third, *AntConc* was employed to extract the potential instances of the PCFs based on regular expressions (regex). The regex patterns were built based on the following linguistic sequences:

- "adjective + noun" pattern to identify attributive adjectives
- "noun + noun" pattern to identify premodifying nouns
- "noun + preposition" pattern to identify prepositional phrases as postmodifiers (including both *of*-prepositional phrases, and *other* prepositional phrases)
- "a string + nominal suffixes" pattern to identify nominalizations

Table 3

Description of the Corpus

Corpora	File No.	Total Tokens	Mean Length
Undergraduate (**UG**)	206	42,678	207.17
Graduate(**GR**)	174	29,913	171.91
Total	380	72,591	191.02

Table 4

Linguistic Model of PCFs

Stages	PCFs	Position	Examples
1	Attributive adjectives	Pre-noun	[Fundamental] questions
2	Premodifying nouns	Pre-noun	[Research] methods
N/A	Nominalizations	N/A	[Industrialization]
4	Prepositional phrases (*of*) as post-noun modifiers	Post-noun	The analytical approach [of discourse analysis]
4	Prepositional phrases (*others*) as post-noun modifiers	Post-noun	The methodological diversity [in applied linguistics]
5	Appositive noun phrases	Post-noun	A common method, [corpus-based approach], has been …

– "words in a pair of punctuations" pattern to identify appositive NPs (i.e., parentheses, square bracket, dashes, and commas).

AntConc utilized these regex patterns to extract relevant cases of the PCFs from the corpus. There are two noteworthy points: (1) the list of suffixes to extract nominalizations was based on five high-frequency nominal suffixes in academic registers in Biber et al. (1999)[5], including "-ment", "-ity", "-tion", "-ness", and "-ance", "-ship"; (2) the patterns to extract appositive noun phrases were based on how this phrasal feature was extracted in related previous studies (Lan & Sun, 2019). Then, the extracted potential cases of the PCFs were saved in an Excel file for the manual adjustment.

Fourth, a manual adjustment was conducted. The extracted potential cases in the Excel file were not all accurate cases of the PCFs. In Example1, the second case (e.g., to have their **meal with** their friends) is a "noun-preposition" sequence but the prepositional phrases (i.e., with their friends) are not to modify the head noun (i.e., meal).

To ensure the accuracy of the data, two researchers conducted manual checks on all instances in the Excel file and removed any inaccurate cases of PCFs. During the pilot coding phase, the researchers collaboratively examined 100 instances of each PCF to establish inter-coder reliability, which yielded a high agreement rate of 93.7%. Following this, the two researchers independently examined the remaining instances of PCFs. Given that a larger number of inconsistent instances were found in the prepositional phrases (other) category during the pilot phase, extra attention was given to ensure the accuracy of this specific PCF. In addition, two PCFs (i.e., attributive adjectives and premodifying nouns) might occur more than one time in a case, such as coordinate attributive adjectives (e.g., a valid and reliable method) or multiple noun sequences (e.g., a corpus research method). These cases were marked in the Excel file and were used to adjust frequencies of the PCFs in the next step.

Fifth, a Python program was applied to count the frequencies of the PCFs based on the Excel file with all the adjusted cases of PCFs. The program output is a dataset with filenames and all the frequencies of the PCFs associated with specific files. We then adjusted the frequencies based on the marked cases of coordinate attributive adjectives and multiple noun sequences in the Excel file. By the end of this step, we considered that the frequencies of the PCFs were accurate for statistical analysis.

Coding of Rhetorical Functions

The schematic structures were analyzed via a multi-step process. First, the researchers examined relevant studies in which textual move organization of data commentary were presented (e.g., Eriksson & Nordrum, 2018; Jalilifar et al., 2019; Nordrum & Eriksson, 2015; Sancho Guinda, 2012; Swales & Feak, 2012). Move-step schemas, along with illustrative examples and functional language representative of rhetorical structures of previous research, were then selected, extracted, and saved in a separate single file. This facilitated adoption and adaption of possible functional labels/names for the moves and steps in our data analysis. Finally, following previous recommendations from the leading researchers (e.g., Biber et al., 2007), the researchers rigorously scrutinized the entire datasets to achieve a clear understanding of the texts' structure, patterns, communication functions, and linguistic signals.

Drawing on seminal text analytical methods (Biber et al., 2007; Monero & Swales, 2018), we assigned discourse roles to various (sub) move types based on function-form relations. While *function* is realized by clause or sentence, *form* is realized by lexico-grammatical constituents in text segments (Monero & Swales, 2018). By utilizing this approach, we aimed to gain a comprehensive understanding of how different linguistic elements contribute to the overall structure and organization of the discourse.

The viewpoint of communicative purposes of move as well as linguistic clues were also the central notion for our analysis. By definition, move is "a discoursal or rhetorical unit in a text that performs a coherent and distinctive communication function in written or spoken discourse" (Swales, 2004, pp. 228-229). On the other hand, moves can be the outcome of the hybrid of multiple elements or sub-moves (steps) which are recognized both rhetorically and linguistically. Steps are, therefore, the many text fragments that "together, or in some combination, realize the move" in such a way that "the steps of a move primarily function to achieve

Example 1

Case	Pre-context	Keyword in context	File
1	...decade_NN (_(1980_CD)_) only_RB five_CD	**percent_NN of_IN** people_NNS spent_VVD...	UG1
2	...more_RBR willing_JJ to_TO have_VH their_PP$	**meal_NN with_IN** their_PP$ friends_NNS...	UG1

[5] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. New York: Longman.

the purpose of the move to which it belongs" (Biber et al., 2007, p. 24).

The researchers followed the general steps outlined in Biber et al. (2007) to conduct a pilot analysis in order to refine the coding schema and minimize the potential risk of human errors. To achieve this, two rounds of pilot coding were performed on a subset comprising 25% of the corpus (*n*=95). The first round included 5% of the files (*n*=19), while the second round encompassed 20% of the files (*n*=76). Each of the researchers independently coded the texts, annotating them in the margins. Following each round, the researchers engaged in discussions to address discrepancies and ensure agreement on the rhetorical interpretations of the coded text segments. This process continued until both researchers reached a consensus. To assess inter-coder reliability, Cohen's Kappa coefficient was calculated, resulting in a Kappa coefficient of 0.865, indicating almost perfect agreement between the coders.

Considering our pilot coding, we developed a coding scheme as the benchmark for the coding of moves and steps for the remaining 75% of the corpus (see Table 5). The coding scheme for the data commentary comprised four rhetorical moves and certain rhetorical steps associated with the moves, as illustrated in Table 5. The only exception was that Move 4 (i.e., concluding visual information) has no associated steps identified. The detailed description of the coding scheme of rhetorical functions and examples taken from the datasets are presented in Appendix C.

Statistical Analysis of PCFs and Rhetorical Functions

Statistical analyses were conducted using SPSS. The frequencies of the PCFs and the counts of the rhetorical functions (moves and/or steps) were firstly normalized to 100 words. Then, the normality of the PCFs and rhetorical moves/steps were examined based on the Shapiro-Wilk test (see Appendix D). The test results revealed that the PCFs and moves/steps did not follow a normal distribution, except for attributive adjectives. To answer the first and second research questions, Mann Whitney U tests were employed to explore the differences in phrasal complexity features and rhetorical functions between the undergraduate and graduate data commentaries, respectively. After that, a Spearman correlation test was utilized to explore the relationship between phrasal complexity features and rhetorical moves/steps, generating a correlation matrix to answer the third research question.

RESULTS AND DISCUSSION

RQ1: How do undergraduate and graduate students utilize PCFs in data commentaries they produce? Which specific types of PCFs are more prominently used by each group?

To answer the first research question, the results of the Mann-Whitney U Test showed that four out of six PCFs had significant differences on the normalized frequencies between the undergraduate corpus and the graduate corpus. These four phrasal complexity features were attributive adjectives, nominalizations, noun-noun sequences, and prepositional phrases (of). Table 6 also presents the p values for these four PCFs (e.g., 0.008 for nominalizations, 0.000 for noun-noun sequences). Based on the comparison of the mean ranks of these four PCFs, we categorized the four PCFs into two groups. Group1 consisted of three PCFs (i.e., attributive adjectives, nominalizations, and prepositional phrases (of)) that were found to be significantly more prevalent in the corpus of graduate data commentaries. On the other hand, Group 2 only included one PCF, noun-noun sequences, which were used significantly more in the corpus of undergraduate data commentaries.

In terms of the effect size of the six PCFs, the noun-noun sequences had the largest eta^2 as 0.172. However, these

Table 5

Rhetorical Functions Found in Data Commentary

Move 1: Presenting Visual Information
Step 1: Providing an explanatory note to set the scene
Step 2: Indicating the location of the data
Move 2: Highlighting Visual Information; Comparing and Contrasting Key Points
Step 1: Describing the facts (with/without providing statistical evidence)
Move 3: Commenting on Visual Information
Step 1: Personal asides
Move 4: Concluding Visual Information

Table 6

Mann-Whitney U Test on the PCFs

Linguistic Features	P Values	Effective Size (eta2)	Mean Rank (UG)	Mean Rank (GR)	Description
Attributive adjectives	0.010*	0.017	176.351	205.252	GR > UG
Nominalizations	0.008*	0.007	175.920	205.773	GR > UG
Noun-noun sequences	0.000*	0.172	230.893	139.924	GR < UG
Prepositional phrases (*of*)	0.000*	0.019	162.308	222.066	GR > UG
Prepositional phrases (*other*)	0.869	0.074	188.652	190.514	NA
Appositive NPs	0.094	0.000	184.572	195.401	NA

Note. The test is based on a two-tailed assumption with 0.05 as the alpha level. "*" marks the PCFs with significant difference between the two groups. The sum of the effect size for the PCFs is small (eta2 =0.290).

suggested a small effect size[6]. The remaining PCFs also had a small effect size such as attributive adjectives (eta^2 = 0.017) and nominalizations (eta^2 = 0.007). The total cumulative effect size of the PCFs was also small with the eta^2 = 0.290. This suggested that the six PCFs can only explain 29% of the variance of the data commentary difference between undergraduate and graduate students. Given that writing differences can arise from a number of factors (e.g., linguistic features beyond the PCFs), we concluded that it is not unexpected to get small effects for the PCFs, because only PCFs are not supposed to explain a large portion of the variance of the writing difference in our research context.

The results further suggested that Iranian graduate students produced three types of PCFs (i.e., attributive adjectives, nominalizations, and prepositional phrases (*of*) as post-noun modifiers) much more but noun-noun sequences much less than their undergraduate counterparts. In terms of nominalizations, few studies have included such phrasal complexity features in recent studies on grammatical complexity. The only study is Staples et al. (2016), who presented a similar finding that L1 English students used nominalizations in their writing when they become academically advanced. For attributive adjectives, Staples et al. (2016) found that L1 graduate students used these linguistic features more than their undergraduate counterparts in academic writing, whereas Parkinson and Musgrave (2014) found that EAP students heavily relied on attributive adjectives in their writing. Our finding is similar to Staples et al. (2016).

In the context of data commentary, it was found that Iranian graduate students used more attributive adjectives than undergraduate students. Qualitative analysis indicated that undergraduate students employed a limited range of attributive adjectives in their data commentary prompts (e.g., public transports, selected countries); However, graduate students used a broader usage of attributive adjectives beyond those provided in the prompts and related to the description of data (e.g., "*the highest illiteracy percentage*", "*the different degrees of the effects*"). Thus, it can be concluded that graduate students exhibited a more extensive and proficient use of attributive adjectives in their data commentary.

Concerning noun-noun sequences, Ansarifar et al. (2018) reported that Iranian PhD students used these linguistic features more frequently in their dissertations compared to MA students in their theses. However, our findings reveal a contrasting pattern, indicating that Iranian graduate students employed noun-noun sequences less frequently than their undergraduate counterparts. The higher frequency of noun-nouns sequences in undergraduate data commentaries can be primarily attributed to topic influence, too. This is especially evident in the case for proper nouns such as "*London Museum*" and certain formulaic sequences such as "*mobile phones*" and "*food budget*", because undergraduate students relied on using them in their data commentary. In contrast, graduate students tend to paraphrase these formulaic sequences, aiming to avoid repetitive use of the same lexical chunks. For instance, they rephrased "*food budget*" as "*the budget on their food* " or "*the budget allocated for food* ".

As for prepositional phrases (*of*), Parkinson and Musgrave (2014) found no significant difference between L2 undergraduate students and MA TESOL students. Our finding showed that Iranian undergraduate students used such linguistic features more frequently than Iranian graduate students. This can be explained by the prevalence basic formulaic sequences in undergraduate data commentaries. When a noun is followed by a prepositional phrase (*of*), this construction primarily functions as a post-noun modifier as in "*the result of a survey*" and "*the plan of our research project*" (Lan & Sun, 2019). However, there are situations where L2 students may not be fully aware of their use of this noun modifier when writing. For example, two common and basic formulaic sequences associated with describing quantity,

[6] The interpretation of the effect size is based on Cohen (1988), with eta2 less than 0.300 suggesting a small effect, between 0.300 and 0.500 suggesting a medium effect, and greater than 0.500 suggesting a large effect.

namely, "*a lot of* " and "*the number of* " can be found in undergraduate data commentaries. In contrast, this is not the case in graduate data commentaries. In fact, Iranian graduate students show a tendency to employ a variety of prepositional phrases (*of*), such as "*the cost of education*" and "*the survey of adult education*".

Last but not least, it is important to emphasize that a greater use of a grammatical feature does not always suggest a greater writing development (Staples & Reppen, 2016). Several relevant studies have demonstrated that the repetition of formulaic sequences in writing does not necessarily indicate a greater writing development (Lan & Sun, 2019; Staples & Reppen, 2016). This is particularly the case in our study. Although undergraduate students used more noun-noun sequences and prepositional phrases (*of*), we cannot concluded that their writing development is greater than that of the graduate students. In contrast, graduate students produced a greater diversity of noun-noun sequences and prepositional phrases (*of*), and had less reliance on formulaic sequences when writing about various topics. This suggested a stronger proficiency in utilizing noun modifiers.

RQ 2: How do undergraduate and graduate students employ rhetorical functions in data commentaries they produce? Which specific types of rhetorical functions are more prominently used by each group?

In order to answer the second research question, we conducted a rank-based non-parametric test (Mann-Whitney U) to measure the significance of differences between the variables. The results indicated that graduate students used significantly more M1 (i.e., presenting visual information), M1S2 (i.e., indicating the location of data), and M2 (i.e., highlighting visual information; comparing and contrasting key points) than undergraduate students. Conversely, undergraduate students used significantly more M3 (i.e., commenting on visual information) and M3S1(i.e., personal asides) than graduate students.

Table 7 also presents the *p* values for the five rhetorical moves and steps, for example 0.000 for M1 (i.e., presenting visual information), 0.000 for M1S2 (i.e., indicating the location of data), and 0.000 for M2 (i.e., highlighting visual information). Based on the comparison of the mean ranks of these rhetorical functions, we categorized them into two groups: (a) Iranian graduate students significantly used more M1, M1S2 and M2; (b) undergraduate students significantly used more M3 and M3S. The two groups suggested a preference of rhetorical functions: graduate students included more moves and/or steps related to presenting and interpreting data while undergraduate students included more moves and/or steps related to expressing their personal opinions. However, based on the statistical interpretation of "M4", there was no significant difference observed in the use of M4 between undergraduate and graduate data commentaries. This suggested that both groups tended to employ M4 in a similar manner.

Regarding effect size, we wish to discuss the rhetorical functions included in the analysis, M1, M2, M3, and M4. Among them, M1 had the largest eta^2 as 0.156, followed by M2 eta^2 as 0.089, indicating small effect size for both rhetorical functions. M3 and M4 also had small effect size in our analysis (eta^2 = 0.002 and eta^2 = 0.000), respectively. Taken together, the cumulative effect size of the four rhetorical functions can be considered medium, with the eta^2 = 0.466. This suggested that the four functions can explain around 46% of the variance of the data commentary differences between undergraduate and graduate students. While other factors may also contribute to differences in data commentaries, we concluded that the four rhetorical functions can explain fairly large portion of the variance in our research context.

Table 7

Mann-Whitney U Test on the Rhetorical Functions

Rhetorical Functions	*P* Values	Effective Size (eta^2)	Mean Rank (UG)	Mean Rank (GR)	Description
M1	0.000*	0.156	150.330	236.420	GR > UG
M1S1	0.245	0.004	194.709	183.261	NA
M1S2	0.000*	0.100	158.788	226.282	GR > UG
M2	0.000*	0.089	159.699	225.192	GR > UG
M2S1	0.875	0.000	188.577	190.604	NA
M2S2	0.000*	0.092	176.000	205.670	GR > UG
M3	0.006*	0.020	200.266	176.604	GR < UG
M3S1	0.000*	0.129	216.524	157.133	GR < UG
M4	0.987	0.000	189.431	189.580	NA

Note. *A Bonferroni adjustment was applied to the alpha value (i.e., 0.05), and the adjusted alpha value is 0.006. "*****" marks the moves/steps with significant difference between the two groups. The sum of the effect size for the four rhetorical functions (i.e., M1, M2, M3, M4) is medium (eta^2 = 0.466).*

These findings suggested the difference of using the four rhetorical functions had a medium practical influence, which deserve teaching attention in real classroom instruction.

Moreover, the results demonstrated that both undergraduate students and graduate students incorporated four moves and four steps to construct their data commentary. The occurrences of some of these moves and steps in both datasets may not only corroborate previous studies on data

commentaries (e.g., Eriksson & Nordrum, 2018; Nordrum & Erikson, 2015; Swales & Feak, 2012) but it may also show that both groups conform to recognizable patterns of discourse organization of data commentaries.

Regarding the rhetorical functions, M1, M1S2, and M2 are highly favored by graduate students. M1 specifically serves the purpose of explicitly stating the objectives of visual content and is achieved through two distinct steps. The first step (i.e., M1S1, providing an explanatory note to set the scene) makes a brief note to begin reporting visual information that follows and apprises the readers of data that is going to be discussed, while the second step (M1S2) helps readers notice further salient information presented in non-text materials. As an opening tactic, M1 is commonly employed to commence the data commentary prior to discussing methods of highlighting visual information. This finding may indicate that data commentary typically begins with a general overview to enhance readers' understanding of the text before highlighting key points visually.

Known as an endophoric marker and a text organizer, M1S2 is utilized to demonstrate the position of the non-verbal information and draw readers' attention to important information presented in visual modes (Hyland, 2019; Swales & Feak, 2012). The greater propensity for employing M1S2 may suggest that the references to charts, tables, diagrams, and figures can form a preferred vehicle for drawing attention to a very brief summary of significant information that non-material texts show. The higher incidence of this rhetorical device could be, moreover, attributed to the directive force of endophoric markers in multimodal contexts. This finding echoes Sancho Guinda's (2012) conclusion that multimodal environments can influence endophorics counts, compared to simply verbal texts. On the whole, the high rate of occurrences of M1S2 may show how student writers establish metadiscoursal links between text constituents, argument, and target reader (see Hyland, 2019). Nevertheless, these metadiscursive items (i.e., M1 and M1S2) can function as prefabricate expressions or ready-made language chunks (e.g., *the chart shows that*... and *according to the table*...) that help writers start their writing. This could explain why these devices outnumber other rhetorical choices.

Recognized as the backbone of data commentary writing, M2 is a core rhetorical device via which key points are compared, contrasted, and accentuated normally by relevant evidence, statistics, and examples. The greater propensity for this rhetorical choice among graduate students may reflect their heightened awareness of the nature, characteristics, and rhetorical conventions of the text type, enabling them to highlight and focus on important aspects in the data commentaries they produced. In sum, the increased preference for M1, M1S2, and M2 may not only indicate graduate students' appreciation of the importance of these cardinal rhetorical moves in data commentary but it also demonstrates the extent to which graduate writers are preoccupied with aiding their readers in directing, processing, and unfolding non-text elements. Thus, these findings can support earlier studies that similarly identified the rhetorical functions as the most common and essential moves in data commentaries (e.g., Eriksson & Nordrum, 2018; Nordrum & Erikson, 2015; Swales & Feak, 2012).

M3 (i.e., commenting on visual information) and M3S1 (i.e., personal asides) were highly preferred by undergraduate students, on the other hand. The greater exploitation of M3 may suggest they are far more likely to include subjective explanations, judgments, and comments on their data description. Another reason why this rhetorical option is preferred is that undergraduate students strove to establish their identities as knowledgeable individuals (Sancho Guinda, 2012). However, the presence of M3 is not in agreement with the common purposes and structure of data commentary writing stated by Swales and Feak (2012). Additionally, M3S1 was solely based on undergraduate students' personal judgments or explanations of the data visually presented and was not directly related to the main topic being discussed. M3S1 can also present an opportunity to the writers to address the target readers subjectively and straightforwardly by asides and interrupting the ongoing discussion to make a statement on what has been mentioned (Hyland, 2019). The presence of M3S1 might give rise to some cases of pragmatic infelicity or deviance from academic writing (Sancho Guinda, 2012); yet, it can function as "an important reader-oriented strategy" (Hyland, 2019) that helps writers/speakers accentuate personal viewpoints, determine the nature social relationships, and exhibit relatively innovative linguistic performance. However, not being a move common and cardinal to data commentary, M3S1 can reflect undergraduate students' overt stance-taking through which they express their stance openly via a description and an evaluation of data. Swales and Feak (2012) regard stance as an important element in academic writing since it enables writers to reveal not only what they think but also what they know. In all, this finding can present evidence for previous research in which M3S1 was found in data commentary (Sancho Guinda, 2012).

Finally, the low incidence of M4 by both groups may suggest that students struggle to provide a brief synopsis of the key information visually presented at the conclusion of their data commentaries. One possible explanation for this could be in line with Swales and Feak (2012, p. 172), who consider M4 to be a challenging segment that distinguishes proficient writers from less proficient ones, as it requires the ability to engage in "some original thinking". This further implies that the skill of producing a clear and concise visual data summary demands not only critical thinking skills but also the capacity to identify and prioritize relevant information effectively. Therefore, students who can effectively exploit M4 and present the data coherently and meaningfully are more likely to excel in writing data commentaries compared to those who struggle with this move.

RQ 3: Is there a correlation between the presence of PCFs and the occurrence of rhetorical functions within both the undergraduate and graduate data commentaries?

In order to answer the third research question, we applied a Spearman correlation matrix among the PCFs and the rhetorical moves/steps. This matrix helped us determine the potential relationship between the intended linguistic features and rhetorical functions. As shown in Table 8, the results indicated that there was a correlation between M1 and attributive adjectives (ρ= 0.223) as well as prepositional phrases (*of*) (ρ= 0.212). M1S2 was also correlated with attributive adjectives (ρ= 0.210). M2 was another rhetorical function correlated with attributive adjectives (ρ= 0.216) and prepositional phrases (*of*) (ρ= 0.214). Finally, there was a correlation between M3S1 with prepositional phrases (*of*) (ρ= - 0.285).

The results demonstrated a pattern of *head nouns + prepositional phrases (of)* for M1 (i.e., presenting visual information). Few studies have correlated the PCFs and rhetorical moves or steps, but two recent studies have revealed the relationship between rhetorical functions and phraseological features (i.e., p-frames), which is related to our study. For instance, Lu et al. (2021a) found that the frequent p-frames (*the NOUN of*) largely contributed to the presentation of research information in the introductions of research articles in social sciences. Lu et al. (2021b) suggested that this was particularly the case for research articles in applied linguistics. For example, some frequent p-frames associated with presenting research purposes were *the purpose of* and *the basis of*. In our study, *head nouns + prepositional phrases (of)* fulfilled the function of presenting visual information. This PCF pattern is frequently used to present visual information in data commentary, as in the examples: *the number/amount/quantity/percent/rate(s) of.* Despite the genre differences, our findings can be consistent with Lu et al. (2021a) and Lu et al. (2021b).

Moreover, there was a connection between M1S2 (i.e., indicating the location of the data) and attributive adjectives. Two possible interpretations can be given: First, although locatives generally present a set of data or figures without concentrating solely on any particular item, they can be embedded in a presentational move or step to focus on a comparison of categories, groups, and data or report a relationship between variables (Lim, 2011). This finding may echo Lim's (2011) interpretation that indicating the location of the data is characterized by certain linguistic features such as (comparative) adjectives even though our genre of study (data commentary writing) is different from that of Lim (2011). Our second interpretation is the student writers' data commentary might be influenced by the presence of attributive adjectives in the writing prompts/topics, as discussed already. However, Lim (2011, p. 738) considered the locatives as a *space-saving strategy* to present a relatively abridged report.

The results further revealed the pattern of *attributive adjectives + head nouns*. Few p-frames based on this pattern can be observed in Lu et al. (2021a) and Lu et al. (2021b) for the rhetorical function of presenting research information. This is contradictory to our finding. However, taking a close look at our corpus, we found that when presenting the visual information, it is unavoidable to use some keywords associated with the topics such as *secondary school* and *mobile phone*. As topic would have information on the pattern of

Table 8

Correlation Matrix between the PCFs and the Rhetorical Functions

Rhetorical functions	Attributive adjectives	Nominalization Forms	Noun-noun sequences	Prepositional phrases (*of*)	Prepositional phrases (*other*)	Appositive NPs
M1	**0.223**	0.081	-0.099	**0.212**	0.023	0.017
M1S1	-0.090	0.058	-0.008	-0.107	0.073	0.107
M1S2	**0.210**	0.039	-0.101	0.187	-0.009	-0.056
M2	**0.216**	-0.125	-0.060	**0.214**	-0.022	0.068
M2S1	0.004	-0.045	-0.030	0.021	-0.039	-0.130
M3	-0.092	0.051	0.039	-0.141	-0.052	-0.035
M3S1	-0.138	-0.019	0.003	**-0.285**	-0.039	-0.058

Note. The number in bold only are the correlation coefficients (ρ) with noticeable absolute values.

attributive adjectives + *head nouns* (e.g., Staples & Reppen, 2016), we found that this pattern has been used to present visual information mostly associated with the keywords of the topics. Nevertheless, the pattern of *attributive adjectives* + *head nouns* + *prepositional phrases (of)* can be excluded from the analysis because few cases are found to present visual information in the corpus.

Another connection between rhetorical functions and the intended linguistic resources has to do with M2 (i.e., highlighting visual Information; comparing and contrasting key points) and *attributive adjectives* + *head nouns* pattern. We have two possible interpretations for this pattern and rhetorical connection: a) this pattern suggests that attributive adjectives are used as a communicative vehicle to describe head nouns to meet the rhetorical functions of highlighting and comparing information as in the examples of *the highest percentage/rate/number/rank* vs. *the lowest percentage/percent/rank/bar*; b) as mentioned earlier, the use of attributive adjectives could be associated with the topics, for example *mobile phone* and *developed countries*. It is not easy to avoid using these attributive adjectives when highlight and/or compare the key visual information as in *the percentages for mobile phone users are close to each other*.

Moreover, the results revealed a connection between M2 and prepositional phrases (*of).* This pattern can also align with the p-frame (*the NOUN of*) in Lu et al. (2021a) and Lu et al. (2021b). As Lu et al. (2021a) concluded, the p-frame (*the NOUN of*) contributed to a wide range of rhetorical functions, not only the presentation of research information as discussed above, but also discussing counter claims and/or indicating research gaps in research articles. These functions involved the comparison and contrast between previous literature and recent studies. In our study, the genre is data commentary, however. A reasonable inference is that the pattern – *head nouns* + *prepositional phrases (of)*- can be further used for comparing visual information in the data/table/charts. In contrast to M1, M2 has many cases of *attributive adjectives* + *head nouns* + *prepositional phrases (of)* pattern, for instance *the highest amount of leisure time*.

Finally, the results indicated a negative correlation between M3S1 (e.g., personal asides) and prepositional phrases (*of*). Lu et al (2021a) suggested that some rhetorical functions that are related to *personal asides*, discussed values and limitation of research. This function involved personal comments/ideas on the research. Lu et al. (2021b) mentioned that writers in social sciences tended to include p-frames that express their stance for participating in knowledge construction in particular disciplines. The frequent p-frames with relevant rhetorical functions are *it is important to VERB*, *it has been VERB in*, and *this is the first NOUN*. Based on the examples, we can see that this rhetorical function related to personal stance is expressed mostly by clausal structures instead of nominal structures based on PCFs. Although our genre is different from Lu et al. (2021a), the expression of personal asides can possibly be universal across different written genres. We also found personal asides have rarely been based on nominal constructions with PCFs, but mostly with verbal structures, for example: *in my opinion*, *it is because* … and *I believe we cannot have a conclusion that* …. This could explain why there was a negative correlation between prepositional phrases (*of*) and M3S1.

CONCLUSION

This study uncovered the use of PCFs and rhetorical functions in data commentaries written by L2 student writers in the Iranian university context. As the L2 student writers become more academically advanced, they tend to use more diverse and abundant PCFs (e.g., nominalizations, prepositional phrases as post-noun modifiers) to construct their data commentary description. This suggests that higher academic levels can be associated with advanced linguistic competence in data commentary writing. Moreover, writing proficiency is reflected not only in linguistic abilities but also in discourse dimensions, where advanced L2 student writers demonstrate a better ability to handle different rhetorical functions (e.g., presenting, highlighting, and/or comparing visual information). Additionally, there is a connection between linguistic features and rhetorical functions, indicating the importance of both dimensions in academic writing development. These insights have practical implications for guiding the instruction and assessment of data commentary writing in educational settings. By understanding students' linguistic and rhetorical patterns, educators can target their instruction to specific areas that need improvement. Finally, this study lays the groundwork for future research; further exploration of additional factors influencing the use of PCFs and rhetorical functions, along with investigations into the impact of instructional interventions on students' writing skills, could broaden our understanding of this field.

This study has some limitations. We need to acknowledge that the discussion of the relationship between the PCFs and rhetorical moves/steps is primarily based on recent studies, because a few prior studies have explored the connection between linguistic features and rhetorical functions in relation to phrasal complexity features in academic writing. Therefore, this study can be regarded as an initial and exploratory investigation. To enhance our understanding of this relationship, future studies should consider incorporating additional research to provide further insights. It is also important to acknowledge that the varying levels of experience between undergraduate and graduate students could potentially impact their approach and proficiency in writing data commentary. This difference was not considered and analyzed in the study. Additionally, although we attempted to minimize topic influence by selecting 20 topics, it is essential to recognize that topic influence cannot be completely eliminated. However, using only one topic for research design can help reduce the influence of topic.

Previous studies have suggested various methods for teaching PCFs, including awareness-raising instruction and data-driven learning. In line with these findings, our study proposes an integrated approach that combines grammatical instruction on PCFs with relevant rhetorical moves and steps. This holistic approach allows students to not only understand the grammatical aspects of PCFs but also comprehend how to use them strategically to achieve specific rhetorical purposes. Utilizing corpora is an effective way to facilitate this integration. Corpora can provide valuable resources of authentic writing in specific genres, such as data commentary and research article. By incorporating corpora into the teaching process, students gain access to real-world examples and can analyze how PCFs are appropriately used in their target genres. This enhances their understanding and application of PCFs within the context of academic writing. Moving from the linguistic instruction to the rhetorical instruction can effectively help students use PCFs appropriately in the written genres that they need to work on in their specific disciplines.

The authors would like to thank the students who actively and voluntarily participated in this research project. They also thank the anonymous reviewers.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Muhammed Parviz: Corpus building/data collection, discourse analysis of rhetorical moves/steps, manuscript drafting.

Ge Lan: Corpus tagging, extraction and analysis of linguistic features, statistical analysis, manuscript drafting.

ACKNOWLEDGMENTS

REFERENCES

Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes, 31*, 58-71. https://doi.org/10.1016/j.jeap.2017.12.008

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D., Connor, U., & Upton, T. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. John Benjamins Publishing.

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

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5-35. https://doi.org/10.5054/tq.2011.244483

Biber, D., Gray, B., Staples, S., & Egbert, J. (2022). *The register-functional approach to grammatical complexity: Theoretical foundation, descriptive research findings, application*. Routledge.

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes, 6*, 289-302. https://doi.org/10.1016/j.jeap.2007.09.009

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Publishers.

Cullen, P., French, A., & Jakeman, V. (2014). *The official cambridge guide to IELTS for academic and general training*. Cambridge University Press.

Eriksson, A., & Nordrum, L. (2018). Unpacking challenges of data commentary writing in master's thesis projects: an insider perspective from chemical engineering. *Research in Science & Technological Education, 36* (4), 499-520. https://doi.org/10.1080/02635143.2018.1460339

Hyland, K. (2000). *Disciplinary discourses: Social interaction in academic writing.* Pearson.

Hyland, K. (2019). *Metadiscourse: Exploring interaction in writing*. Continuum.

Jalilifar, A., Parviz, M., & Don, A. (2019). Exploring phrasal complexity features in graduate students' data commentaries and research articles. *Journal of English Language Teaching and Learning*, *24*(11), 118-152. https://doi.org/10.22055/.2019.17801

Jiang, F. K., & Hyland, K. (2015). 'The fact that': Stance nouns in disciplinary writing. *Discourse Studies*, *17*(5), 529–550. https://doi.org/10.1177/1461445615590719

Jiang, F. K., & Hyland, K. (2021). 'The goal of this analysis …': Changing patterns of metadiscursive nouns in disciplinary writing. *Lingua*, *252*, 103017. https://doi.org/10.1016/j.lingua.2020.103017

Lan, G., & Sun, Y. (2019). A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course. *Journal of English for Academic Purposes*, *38*, 14-24. https://doi.org/10.1016/j.jeap.2018.12.001.

Lavelle, E., & Bushrow, K. (2007). Writing approaches of graduate students. *Educational Psychology*, *27*(6), 807-822. https://doi.org/10.1080/01443410701366001.

Lim, J. M. H. (2011). Paving the way for research findings: Writers' rhetorical choices in education and applied linguistics. *Discourse Studies, 13*(6), 725-749. https://doi.org/10.1177/1461445611421364.

Lu, X., Yoon, J., & Kisselev, O. (2021). Matching phrase-frames to rhetorical moves in social science research article introductions. *English for Specific Purposes, 61,* 63-83. https://doi.org/10.1016/j.esp.2020.10.001.

Lu, X., Yoon, J., Kisselev, O., Casal J., Deng, J., & Nie, R. (2021). Rhetorical and phraseological features of research article introductions: Variation among five social science disciplines. *System*, *100*, 102543. https://doi.org/10.1016/j.system.2021.102543.

Moreno, A. & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes, 50*, 40-63. https://doi.org/10.1016/j.esp.2017.11.006.

Nordrum, L., & Eriksson., A. (2015). Using a small, specialized corpus for formative self-assessment practices. In M. Callies & S. Götz (Eds.), *Learner corpora in language testing and assessment. Studies in corpus linguistics* (vol. 70, pp. 59-84). John Benjamins Publishing Company.

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for academic purposes students. *Journal of English for Academic Purposes, 14*, 48-59. https://doi.org/10.1016/j.jeap.2013.12.001.

Parviz, M. (2023). A comparative corpus-based investigation of results sections of research articles in Applied Linguistics and Physics. *ICAME Journal*, *47*(1). https://doi.org/10.2478/icame-2023-0005.

Parviz, M., Jalilifar, A., & Don, A. (2020). Phrasal discourse style in cross-disciplinary writing: A comparison of phrasal complexity features in the results sections of research articles. *Círculo de Lingüística Aplicada a la Comunicación*, *83*, 191-204. https://doi.org/10.5209/clac.70573.

Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes, 7*, 55-67. https://doi.org/10.1016/j.jeap.2008.02.005.

Sancho Guinda, C. (2012). Proximal positioning in students' graph commentaries. In K. Hyland & C. Sancho Guinda (Eds.), Stance and Voice in Written Academic Genres (pp.166-183). Palgrave Macmillan.

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, *33*(2), 149-183. https://doi.org/10.1177/074108831663152.

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexicogrammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing, 32*, 17-35. https://doi.org/10.1016/j.jslw.2016.02.002.

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.

Swales, J. M. & Feak, C.B. (1997). From information transfer to data commentary. In Miller, T. (Ed.), *Functional approaches to written text: Classroom applications*. United States Information Agency.

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills.* University of Michigan Press.

Wharton, S. (2012). Epistemological and interpersonal stance in a data description task: Findings from a discipline-specific learner corpus. *English for Specific Purposes, 31*(4), 261-270. https://doi.org/10.1016/j.esp.2012.05.005.

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. English for Spec*if*ic Purposes, 22(4), 365-385. https://doi.org/10.1016/S0889-4906(02)00026-1.

APPENDIX A

A list of 20 writing topics used in the study.

(1) The charts below show the results of a survey of adult education. The first chart shows the reasons why adults decide to study. The pie chart shows how people think the costs of adult education should be shared.

(2) The chart shows the number of mobile phones and landlines per 100 people in selected countries.

(3) The chart below shows estimated world illiteracy rates by region and by gender for the year 2000.

(4) The chart shows the percentage of their food budget the average family spent on restaurant meals in different years. The graph shows the number of meals eaten in fast food restaurants and sit-down restaurants.

(5) The bar chart shows the number of visitors to three London Museums between 2007 and 2012.

(6) The bar graph shows the annual number of rides taken by two forms of public transports in the city of Williamsville.

(7) The bar chart below shows the results of a survey conducted by a personnel department at a major company. The survey was carried out on two groups of workers: those aged from 18-30 and those aged 45-60, and shows factors affecting their work performance.

(8) The line graph below gives information on cinema attendance in the UK.

(9) The chart below shows the Higher Colleges of Technology graduates in the UAE.

(10) The chart below shows the amount of leisure time enjoyed by men and women of different employment status.

(11) The diagrams below show the changes that have taken place at West park Secondary School since its construction in 1950.

(12) The pie charts and the table show the types of living accommodation occupied by 25-year-olds in London during the 1990s and 2010s and the availability of different types of accommodation in London during the same two periods.

(13) The pie charts below show the share of Oscar winners by film genre for 2003 and 2008.

(14) The line graph below shows the percentage of tourists to Scotland who visited four different attractions in Edinburgh.

(15) The chart below gives information on the percentage of British people giving money to charity by age range for the years 1990 and 2010.

(16) The pie charts below show the online sales for retail sectors in New Zealand in 2003 and 2013.

(17) The chart below shows the changes that took place in three different areas of crime in New Port city center from 2003-2012.

(18) The chart below shows the annual number of rentals and sales (in various formats) of films from a particular store between 2002 to 2011.

(19) The chat below gives information about Southland's main exports in 2000 and future projections for 2025.

(20) The graph and table below show average monthly temperatures and average number of hours of sunshine per year in three major cities.

APPENDIX B

The precision and recall rates of the three target features

Target features	Precision	Recall
Adjectives	95.3%	95.9%
Nouns	96.2%	97.8%
Prepositions	98.9%	99.9%

Note: The precision and recall are generated based on 5% files of the corpus.

APPENDIX C

The detailed description of the coding scheme of rhetorical functions as well as examples taken from the datasets are presented.

Moves	Definitions	Examples from the Datasets
M1: Presenting visual information	Announcing the purpose(s) of visual prompts/forms	The bar graph representation provides data description on a survey of adult education with reference to the reasons why they have made their minds up to continue their studies. (GR)
M1S1: Providing an explanatory note to set the scene	Making a brief note to begin reporting visual information that follows	In today's report I am going to compare people's food budget in restaurant meal during 1970 until 2000...(UG)
M1S2: Indicating the location of the data	Showing the position of the non-verbal information	As bar chart shows, these factors have different effects on each range of age. (GR)
M2: Highlighting visual information; Comparing and contrasting key points	Revealing data at several levels of details including trends and regularities	From 2009 onwards, it has experienced an increase and in 2012 it records 14 million visitors in a year. Victoria and Albert Museum, on the other hand, shows a constant annual record from 2007 until 2008. The record is noticeable and is about 13 million visitors each year. However, this annual record undergoes a gradual decrease in the following two years....(GR)
M2S1: Describing the facts (with/without providing statistical evidence)	Stating information (with/without providing statistical evidence)	The least important fact among those mentioned, however, is meeting people. With respect to sharing the cost of the courses most probably is that it should be shared on the basis that tax payers pay 25%, employers 35% and individuals 40%. (GR)
M3: Commenting on visual information	Making comments on the non-verbal information	I believe landlines are not as important as mobile phones, we can carry them everywhere but landlines don't have this property... (UG). It is strange that the numerical distance between sexes in Latin America is very low (less than two percent)...(GR).
M3S1: Personal asides	Expressing personal opinions	I'm not trying to say that fast foods are disaster, but it is very harmful if it becomes your everyday meal. (UG)
M4: Concluding visual information	Affording a short account or summary of the key points	To conclude, we can say that Dubai and Abu Dhabi are the two most popular colleges in UAE and the women occupied a much greater role in UAE... (GR)

APPENDIX D

Shapiro-Wilk Test on Normality

	Statistics	*P*-value	Description
Attributive adjectives	0.993	0.081	normal
Appositive NPs	0.313	0.000	not normal
Noun-noun sequences	0.925	0.000	not normal
Nominalizations	0.912	0.000	not normal
Prepositional phrases (*of*)	0.958	0.000	not normal
Prepositional phrases (*others*)	0.969	0.000	not normal
M1	0.929	0.000	not normal
M1S1	0.684	0.000	not normal
M1S2	0.882	0.000	not normal
M2	0.992	0.052	normal
M2S1	0.904	0.000	not normal
M2S2	0.278	0.000	not normal
M3	0.526	0.000	not normal
M3S1	0.496	0.000	not normal
M4	0.605	0.000	not normal

https://doi.org/10.17323/jle.2023.14329

Exploring Translanguaging during Metacognitive Strategy Use on L2 Listening and Writing Skills

Roderick Julian Robillos

Khon Kaen University, Khon Kaen, Thailand

ABSTRACT

Background: The educational concept of translanguaging has garnered significant attention over the past decade. Its significance in fostering language acquisition in the English as a Foreign Language (EFL) classroom has been increasingly acknowledged. However, there is an emerging necessity to strategically implement this pedagogical approach to enhance learning outcomes and improve overall effectiveness.

Purpose: The present study sets out to investigate the influence of translanguaging during metacognitive strategy use and its impact on second language (L2) listening and writing abilities.

Method: The present study employed sequential mixed-method research involving a pre- and post-test design. A total of 16 college students was purposively selected as samples and underwent 11 sessions of applying translanguaging during metacognitive strategy use (the intervention used).

Results: Study results reveal a significant difference on participants' pre- ($\bar{x}$=9.19) and post- ($\bar{x}$=15.56) listening comprehension tests. An increasing trend of improvement on their quizzes in terms of writing components namely: grammar and structure, content, lexical resource, logical order, and supporting details was also found. In addition, the components on "grammar and structure" and "supporting details" have improved considerably. The participants, likewise, perceived translanguaging as a normal and not a disrespectful practice for them as EFL learners. Qualitative findings revealed that participants have welcomed the use of the intervention as it aids them to process their listening comprehension and writing skills in L2.

Conclusion: The pedagogical application of translanguaging approach during metacognitive strategy use is concluded as an agentive and facilitative pedagogical strategy that helps learners to not only improve their listening comprehension and writing skills but also promotes deeper cognitive fluency, improves L2 learning, and fosters them to become more involved in the learning processes of metacognitive planning, monitoring, and evaluating.

KEYWORDS

listening skill, metacognitive approach, second language (L2) learning, translanguaging, writing skill

Citation: Robillos R. J. (2023). Exploring Translanguaging during Metacognitive Strategy Use on L2 Listening and Writing Skills. *Journal of Language and Education, 9*(3), 110-128. https://doi.org/10.17323/jle.2023.14329

Correspondence:
Roderick Julian Robillos,
rodero@kku.ac.th

Received: May 02, 2022
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

The present study attempts to investigate whether a pedagogical application of translanguaging can be utilized in the EFL classroom, and how the L2 listening comprehension and writing skills of the students is impacted by using translanguaging during metacognitive strategy use, where no studies, to date, have been conducted in the EFL field. In the educational system of Thailand, EFL listening skill is put aside, meaning, it is just embedded in the instruction of speaking skill in the classroom, and thus, EFL learners do not have many opportunities to improve their listening skill (Robillos, 2019). In fact, listening activities are often used as a test of comprehension questions (*e.g.,* multiple choice, fill in the blanks) (Goh, 2008; Robillos, 2019; Robillos & Bustos, 2022). Thus, the process of explicitly teaching listening skills is neglected. Lack of exposure to listening tactics and strategies means students may be left with unresolved difficulties in their own understanding of the subtleties of linguistic and non-linguistic stim-

uli (Bozorgian, 2014; Graham, 2017). Furthermore, L2 learners should develop their communicative skills that can give them the confidence to use strategies in solving problems when their language skills are not sufficient (García, 2017) and since every successful communication starts from effective listening which is one most important skill in L2 learning (Robillos, 2019; 2020), this should be given utmost attention. One possible strategy of supporting listening in the EFL classroom which might help meet these demands is the incorporation of first language or L1 into an L2 instructional setting and this is often termed as 'translanguaging' (García, et al., 2017; Moody, Chowdhury, & Eslami, 2019)

Translanguaging is the act of using different languages interchangeably, in order to overcome language constraints, to deliver verbal utterances or written statements effectively (García & Li, 2014; García et al., 2017), to make - meaning (Rivera & Mazak, 2017) and, to effectively attain successful communication (Csillik & Golubeva, 2019b). Li (2018) describes translanguaging as the use of more than one linguistic repertoire and other meaning-making (semiotic) and cognitive resources in learning a target language (TL). The belief is, when it comes to languaging, if we can get our students to engage in ideas and let them use their linguistic repertoire or all the languages that they know, conceptual knowledge and language will blossom. Previous studies claimed that a strategic use of learners' L1 will not slow down or impede their process of learning a language, but rather scaffold both their L1 and their TL (Csillik & Golubeva, 2019b; García & Li, 2014). Translanguaging allows students to use their L1 as a first step to comprehend and enables them to generate new ideas and help promote their cognitive fluency. Furthermore, EFL learning is reframed as bilingual education (Turnbull, 2018) where both the teacher and the students can translanguage to generate learning opportunities, and as such, is still not given much attention in the EFL research field.

It is well-known that incorporating bilingual practices into the classroom (*e.g*. translanguaging) promotes language learning (Csillik & Golubeva, 2019b; García & Li, 2014; Garcia et al., 2017; Li, 2018; Otheguy et al., 2015; Sobkowiak, 2022). Many SLA researchers also claim that a tactical, metacognitive approach to second language instruction can provide learners with the proper guidance in learning their L2 (Oxford, 1993; O'Malley & Chamot, 1990; Robillos, 2019). For instance, the teaching of writing and listening skills is considered complex and requires a strategic approach. This approach has been proven to enhance the control, confidence, and overall proficiency of EFL learners (Bozorgian, 2014; Goh, 2008; Robillos, 2020; Robillos & Thongpai, 2022; Vandergrift & Tafaghodtari, 2010; Zheng, 2018). When students acquire the ability to plan, monitor, and evaluate their tasks, they develop metacognitive awareness. This awareness allows them to assume more responsibility for their learning, identify and address gaps in their understanding, and reflect on their performance (Vandergrift & Goh, 2012; Robillos & Phantharaphong, 2020).

The paper first describes the term "translanguaging" as a pedagogy, followed by the discussion of translanguaging during metacognitive strategy use in EFL listening and finally, summary of research findings and their contribution to the existing literature on translanguaging and L2 listening.

LITERATURE REVIEW

Application of TLAN to L2 Learning

The term "translanguaging" was first coined by Cen Williams (1994) -a Welsh scholar, to emphasize the role of L1 in the development of L2. Williams was in contrary to the norm of splitting named languages and believed that learners' L1 can be beneficially utilized in learning an L2. That is, in an L2 classroom that implements translanguaging, an "English-only" policy is abandoned, and learners are not thought of as deficient, non-native speakers of English, defined by what they lack, but as resourceful agents with multilingual repertoires, abilities, and talents (García & Kleifgen, 2018). They are encouraged to use all their linguistic resources, sustain their culture and identities while expanding their repertoire by learning features of the target language (Sobkowiak, 2022). This flexible use of linguistic resources promotes deep and critical thinking, and enables rigorous cognitive engagement with such tasks. Furthermore, the pedagogical application of translanguaging allows users to access their full linguistic repertoire without having to be constantly aware of socially and politically defined boundaries of named languages (Otheguy et al., 2015). Cumminns (2000) suggested that knowledge about the world is not bound to a specific language, and thus learners should use the language(s) in which they feel most comfortable." This de-emphasizes all knowledge and also emphasizes the point of translanguaging by including languages. This is supported by Cohen (1995) emphasizing that L2 leaners persistently shift between their various language resources. Therefore, L2 learners should be encouraged to use any of their linguistic repertoires for academic purposes (Garcia & Otheguy, 2020).

Moreover, the utilization of translanguaging acts as a support system for comprehending lessons, saving time, and maximizing students' linguistic abilities in problem-solving, constructing meaning, and acquiring knowledge (Tian et al., 2020). Numerous studies have provided empirical evidence regarding the effectiveness of translanguaging in EFL macro-skills. For instance, in the context of L2 reading, Mgijima & Makalela (2021) discovered a positive relationship between the application of translanguaging techniques and learners' performance in composing summary writing. In L2 writing, Chen et al. (2019) asserted that employing translanguaging techniques helped college students enhance

their ESP writing skills across three domains: content, literacy, and academic style. Regarding language improvement, Elashhab (2020) argued that translanguaging contributed to the improvement of medical students' four English language skills, highlighting its significance as an educational practice. This perspective is supported by Moody et al. (2019), who claimed that students' use of translanguaging is a natural linguistic resource that does not hinder successful communication. The students in their study expressed positivity towards the idea that implementing translanguaging techniques empowers them to enhance their proficiency in L2. This current study aims to expand on this research by examining the potential role of translanguaging pedagogy in improving listening skills, an area that has not yet been explored in previous studies.

Metacognitive Approach to L2 Listening

One of the most preeminent pedagogical approaches that are widely recognized in the education field is metacognition. As coined by Flavell (1979), metacognition refers to the knowledge about and regulation of the cognitive processes during learning. In short, metacognitively aware students recognize the gaps in their understanding and are willing to seek out strategies to help fill these gaps. Several researchers in second language acquisition (SLA) have acknowledged that learners' awareness of their learning processes, cognitive functions, and use of strategies can have positive effects on learning tasks (Bozorgian et al., 2021; Goh, 2008; Robillos, 2019; Vandergrift & Tafaghodtari, 2010). For instance, Bozorgian et al. (2021) conducted a study in Iran to explore the impact of metacognitive intervention in the learners' native language (L1) on the listening performance of EFL learners. The experimental group received an eight-week intervention conducted in L1, focusing on metacognitive strategies for the listening task. The results demonstrated a significant improvement in the overall listening performance of the learners. Similarly, Goh (2008) advocated for a metacognitive-based approach to teaching listening, as it enables the instruction of skills and processes related to listening comprehension. This approach enhances control, confidence, and proficiency among EFL listeners. Goh argues that solely testing listening without teaching the necessary tactics for gaining control over the listening process and improving listening skills only generates concerns for students without providing much assistance (Goh, 2008; Robillos, 2020; Vandergrift & Tafaghodtari, 2010).

The current study aims to build upon existing research by incorporating translanguaging pedagogy in conjunction with metacognitive strategy use to facilitate L2 listening comprehension and writing tasks. Specifically, metacognitive skills are instructed in the L2 language to support students' listening comprehension tasks. However, during paired or group activities that involve extensive discussions, students are allowed to utilize their linguistic resources freely. This enables them to deepen their understanding of the listening selections and engage in more meaningful discussions with their peers about listening strategies and their own insights and understanding. It is believed that metacognitively aware students are more likely to plan, monitor, and evaluate the strategies or effective tactics they use for learning, adapting these strategies based on their learning environment (Goh, 2008; Robillos, 2019; Robillos & Bustos, 2022; Vandergrift & Tafaghodtari, 2010). Implementing a metacognitive approach to listening has been shown to enhance control, confidence, and proficiency among EFL listeners (Bozorgian et al., 2021; Goh, 2008; Vandergrift & Tafaghodtari, 2010; Robillos, 2019; Robillos & Bustos, 2022). Therefore, translanguaging pedagogy is integrated into the stages of the metacognitive approach, which include (i) the planning stage involving relevant planning tactics such as advance organization and establishing background knowledge; (ii) the monitoring stage that entails collaborating with peers to compare, discuss differences, identify listening problems, and plan for subsequent listening processes; and (iii) the evaluation stage, which involves evaluation and reflection (Vandergrift & Goh, 2012). Listeners engage in translanguaging to express themselves creatively, connect with the given listening materials, and promote their own development as listeners. When listeners can express their ideas fully and meaningfully without language barriers, they can achieve a deeper understanding of the subject matter and enhance their L2 learning. Similarly, when they possess sufficient metacognitive knowledge and actively participate in planning, monitoring, and evaluating their listening comprehension, they can prepare themselves, check, assess, and reflect on their listening achievements (Vandergrift & Goh, 2012).

Although previous studies have demonstrated the effectiveness of translanguaging in various pedagogical approaches, such as its use in Content and Language Integrated Learning (Lin & He, 2017), Task-Based Language Teaching (Seals et al., 2020), and a Project-based approach (Carpenter & Matsugu, 2020), no prior research has explored the application of translanguaging during metacognitive strategy use to the teaching of listening comprehension and writing skills. Additionally, there is a lack of studies on the pedagogical application of translanguaging in the Thai educational context. Therefore, the present study aims to address this research gap by investigating the impact of implementing translanguaging during metacognitive strategy use to improve students' listening comprehension and writing skills in L2. Specifically, the study focuses on the following research questions (RQ's):

RQ1. Is there a significant difference between the participants' listening comprehension performance before and after implementing translanguaging during metacognitive strategy use (the intervention used in the study)?

RQ2. Is there an improvement on participants' listening comprehension and writing skills in L2 as manifested on their quizzes after the strategy intervention was provided?

RQ3. How do the participants perceive the use of translanguaging as practice, translanguaging for L2 learning, and translanguaging for listening comprehension task processes?

RQ4. What experiences have the participants gained in improving their listening comprehension and writing skills in L2 after the intervention?

METHOD

Research Design

The current study utilized a sequential mixed-method research design, which incorporated both quantitative and qualitative components (Creswell & Plano-Clark, 2011). The quantitative data were employed to examine the influence of implementing translanguaging during metacognitive strategy use on students' listening comprehension and writing skills in L2. On the other hand, the qualitative data were utilized to explore students' perspectives regarding the implementation of the intervention to facilitate their listening comprehension and writing tasks.

Table 1

Participants' Demographic Information

Demographic Profile		N	%
Gender	Male	3	18.75
	Female	13	81.25
	Total	16	100
Age	18 and below	15	93.75
	19-20	1	6.25
	21 and above	0	0
	Total	16	100
First Language Used	Thai	16	100
	Isarn	16	100
	Lanna	4	25
	Phu Thai	4	25
Language/s Proficiently Spoken	Thai	16	100
	English	8	50
	Chinese	5	31.25
	Japanese	7	43.75
	Khmer	2	12.5
Years of Experience in Using English language	1 year and below	0	0
	1-2 years	2	12.5
	3-4 years	4	25
	5-6 years	6	37.5
	7-8 years	4	25
	over 9 years	0	0
	Total	16	100

Note. The demographic information is self-reported and emphasized that language proficiency is solely based on speaking – not listening/reading/writing skills

Participants

In this study, the researcher used a single group of pre- and post-test design and a semi-structured interview for collecting data to explore the effect of applying the strategy intervention on the students' listening comprehension of short informative video items manifested on their writing products. Furthermore, the researcher used a total of 11 sessions constituted of 9 sessions for the implementation of the strategy intervention, and one session each for the administration of the pre- and post-listening tests. All the 16 First Year college students majoring in the program of Teaching English to Speakers of Other Languages (TESOL) at the study-university located in the Northeastern part of the country, were purposively selected as samples. These EFL participants are enrolled in the Intermediate Academic Listening and Speaking course aimed at developing their listening and speaking skills. Specifically, participants must take the course to improve both of their communicative and linguistic competencies which is the focus of the academic listening and speaking program. The course taken during their first year is generally one of the students' listening courses to expose them to learn how to listen and not only to listen to learn. However, the TESOL program required English language as the medium of instruction to almost all of the subjects the students are studying. Many of these students struggle to comprehend academic listening selections in English. This might be due to the traditional way of teaching listening (*e.g.,* focusing only on the listening product by testing listening and the strict English-only policy of the program) (Author 1; Authors 5). This has been a cause for concern at the study-university. Thus, the researcher as a lecturer in the program, decided to embark on this inquiry using translanguaging pedagogy within metacognitive approach to help students develop metacognitive skills that will aid them in dealing with listening comprehension problems and challenges.

It is worthy to note as well that all of the participants are simultaneous Thai-Isarn bilinguals with English as the third (or fourth) language. Albeit all the participants use Thai and Isarn as their first languages, some of them are not proficient to speak Isarn language. In order to yield more understanding with regard to participants' demographic information, table 1 below was presented.

Instruments and Data Collection

Short Informative Video Items

There were four short informative video items used throughout the intervention. The choice of this material is based on the fact that the learners participating in the study are exposed to this authentic material on a regular basis. These short informative videos focused on global issues and are aligned to the topics indicated in the subject "*Intermediate Academic Listening and Speaking*" they were studying. The videos were ranging from approximately 2.0 minutes to around 3.0 minutes long. The four (4) short informative video items used were : (1) *Living in the city or countryside* with 2.53 minutes long; (2) *Causes, Effects of Climate Change* with 2.48 minutes long; (3) *Noise Pollution: causes, effects and solution* with 2.40 minutes long; and (4) *Does Saving more Lives Lead to Overpopulation?* with 2.05 minutes long. Each of the 4 videos used as listening selections was divided into three segments. There were paired and group sharing activities to happen in each listening segment (*see:* the intervention programme) and they were permitted to use any of their linguistic repertoires to enable them to understand the listening segment/s more deeply and more meanifully before writing down their comprehension into paragraph/s. The informative videos were piloted to students similar to the current participants. The short informative videos received a mean rating of 4.37 out of 5.0 for cognitive appropriateness with a Cronbach alpha result of 0.89.

Scoring Rubric

The responses for the two tests as well as the participants' quizzes were scored based on the scoring rubric. There were two scoring rubrics utilized to score students' listening performances: one scoring rubric for their pre- and post-tests and another rubric for their listening quizzes. Whilst the scoring rubric for students' listening quizzes (*see* Appendix A) constituted of grammar and structure, content, lexical resource, logical order, and supporting details, the scoring rubric for their pre- and post-tests included 5 writing components such as main idea, supporting details, logical order, content, and paraphrasing (*see* Appendix C). The writing rubric is designed by the researcher himself, however, checked by the three English experts of the study-university for cognitive appropriateness. A total of 20 marks would be yielded with 4 marks as the highest, and 1 mark as the lowest.

Pre-and Post-Tests

Pre-and post-tests were conducted to assess the participants' listening comprehension performance for short informative video items. The pre-listening test was administered one week prior to the intervention, and activities such as posing questions and group brainstorming were conducted to activate their background knowledge before they wrote their comprehension of the listening selection. Participants were given 60 minutes to complete their written compositions. The post-test took place one day after the intervention, using the same short informative video as the pre-test. Prior to writing their compositions, participants were provided with the intervention. They were given 60 minutes to complete their written compositions. The topic for both the pre- and post-tests was "My family influences my life more than my peers/friends." Additionally, a separate scoring rubric (Appendix C) was utilized to assess and score the students' pre- and post-tests.

Students' Listening Comprehension Quizzes

There were four short informative video items used throughout the intervention program. The participants engaged in metacognitive stages to support their listening tasks. Subsequently, they completed a comprehension quiz by writing paragraphs based on the listening selection. Each written paragraph was evaluated using a writing rubric, considering factors such as grammar and structure, content, lexical resources, logical order, and supporting details. The students' written work was first scored and then returned to them. During group activities, consisting of skilled, unskilled, and average students, participants identified errors in their papers, which were highlighted, encircled, or underlined by their teacher. The purpose of the group activities was to allow students to identify and discuss the errors they made (e.g., grammar, structure, vocabulary, logical connections, content, supporting details) and share successful strategies for future use. However, the scores obtained from the written paragraphs remained unchanged to encourage students to practice evaluation sub-stages, such as problem identification and strategy evaluation. Furthermore, students were permitted to use any language they were proficient in to express their opinions about the errors in their papers more comprehensively. Examples of questions students might ask include: What errors did you identify? Why do you think the errors occurred? What do you believe is the correct version? Why do you think so?.

Students' Perceptions on TLAN during Metacognitive Strategy Use

This questionnaire was administered after the intervention was implemented to the students in order to determine the participants' perceptions towards translanguaging and requested the participants to rate the statements on a Likert scale ranging from 1=strongly disagree to 5=strongly agree. Item-level descriptive results are reported in Table 4. The first section focused on students' perceptions on "translanguaging as a practice" whilst the second section focused on "translanguaging for L2 learning" (Moody et al., 2019). The last section of the questionnaire, however, focused on "translanguaging during metacognitive strategy use to process their listening comprehension" is designed by the researcher himself. The questionnaire was then checked by the three English experts, piloted to 26 Second Year EFL college students to further identify any potential issues with the questionnaire items. The reported reliability value was 0.87. Minor adjustments were made to the order of the questionnaire items before its final administration.

Interviews

Interviews were conducted after the intervention was provided. This is to yield more in depth information from the participants in regards to how often and when they used the intervention in facilitating their listening comprehension tasks and be able to explore the planning, monitoring, evaluation, and reflection strategies that they implemented before, during, and after processing their listening comprehension and writing tasks.

Research Procedures and the Intervention Programme

The researcher used a total of 11 sessions constituted of: 1 session each for the administration of pre- and post- listening comprehension tests; 9 sessions (180 minutes/session) for implementing listening comprehension task/s on participants through metacognitive stages of planning, monitoring, and evaluation along with its subcomponents where they can translanguage every time they communicate whether in pairs or in groups to fully understand the listening selection as well as improve their writing skill. Figure 1 below displays the intervention programme implemented to the participants.

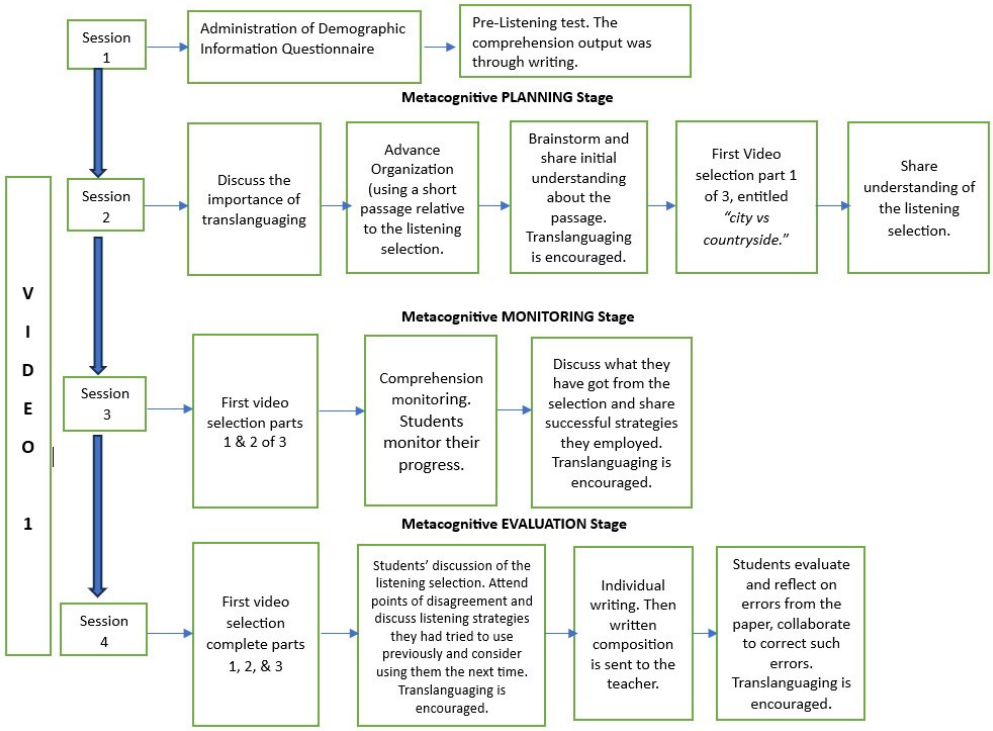
The procedures used in sessions 2, 3, and 4 for the 1st video were followed as similar procedures for videos 2, 3, and 4. However, the number of sessions was reduced to two sessions for each video. Thus, sessions 5 and 6 for video 2 activity, sessions 7 and 8 for video 3 activity, and sessions 9 and 10 for video 4. Videos 2, 3, and 4 focused on listening topics such as "*Causes and effects of climate change*", "*Noise pollution: causes, effects and solution*", and "*Does saving more lives lead to overpopulation*?" respectively. The last session (Session 11) was used for the administration of the post-listening comprehension test.

Analysis of the Data Gathered

The quantitative data were analyzed using descriptive and inferential statistics, while the qualitative data were subjected to thematic analysis. Descriptive statistics, such as mean, frequency, and standard deviation (SD), were calculated using the SPSS tool. The t-test was employed to compare the means of the pre- and post-listening comprehension tests, determining the effect of the strategy intervention and whether there was a significant difference. ANOVA was used to assess any significant differences among the scores in the four quizzes. In the analysis of qualitative data, a thematic analysis approach was utilized to examine the data obtained from semi-structured interviews. The data from these interviews were analyzed and coded through topical coding to label text, subsequently interpreted and modified to identify emerging themes. The following themes emerged: Theme 1 – students' perception of translanguaging practices, Theme 2 - benefits of applying translanguaging during metacognitive strategy use, and Theme 3 - improvement in L2 writing and L2 learning.

Figure 1

The Intervention Programme



RESULTS

Quantitative Analysis

Test of Difference on Participants' Listening Comprehension Performance (RQ1)

Figure 2 below compares the participants' listening comprehension scores before and after the intervention was implemented. The gray lines represent individual students, and the blue line represents the average. It can be gleaned that before the intervention was provided on participants, the majority of them performed below the passing score of 10 (68.75%) and only five (31.25%) of the participants got 10-11 marks. However, after the strategy intervention was implemented, five (31.25%) participants yielded a score ranging from 14-16 out of 20 marks, three of them got a score ranging from 11-13 (18.75%) marks, and eight or (50%) of them yielded a score ranging from 17-19 marks. Overall, students performed significantly better on the post-test ($\bar{x}$ = 15.56, SD = 2.34) than the pre-test ($\bar{x}$ = 9.19, SD = 1.11; t(15) = 13.23, p < .001, d = 3.30). The result is an indication that the participants performed better in their listening comprehension of short informative video items as none of them scored lower than 10 points. Figure 2 below shows the spaghetti plot of the participants' listening comprehension performance over time.

Listening Comprehension Quiz Results (RQ2)

Figure 3 displays the Mean results of the students' listening comprehension quiz scores as manifested on their written outputs by components over time. As noticed, the participants' quiz score began at a lower starting timepoint. However, the overall quiz scores they gained kept an increasing trend during the implementation of the programme as indicated by their overall mean scores of $\bar{x}$=2.12, $\bar{x}$=2.51, $\bar{x}$=2.76, and $\bar{x}$=3.02 for 1st, 2nd, 3rd, and 4th quizzes respectively. Furthermore, as gleaned specifically from the figure, albeit all the components showed an improvement over time, the component on "grammar" (mean score quiz #1, $\bar{x}$=1.89; mean score quiz #4, $\bar{x}$=3.00 = mean improvement of 1.31), indicated the most improved component as compared to the rest. It is also worthy to note that the component on "logical order' has just started to improve during the 2nd, 3rd, and 4th quizzes and was the least improved one. Notably, the students performed significantly differently across the four quiz timepoints: F (3,45) = 86.73, p<.001, partial η^2 = 0.85.

Further, since the students' performances were assessed on two dimensions for each quiz, *i.e.,* content and details for listening comprehension skill (orange line in figure 4), and grammar, lexical resources, and logical order for writing performance (solid green line in the figure), it is crucial to point out the distinctions between the two skills. As exhibited in the line graph, the students' listening performances

Figure 2.

Participants' listening comprehension tests over time

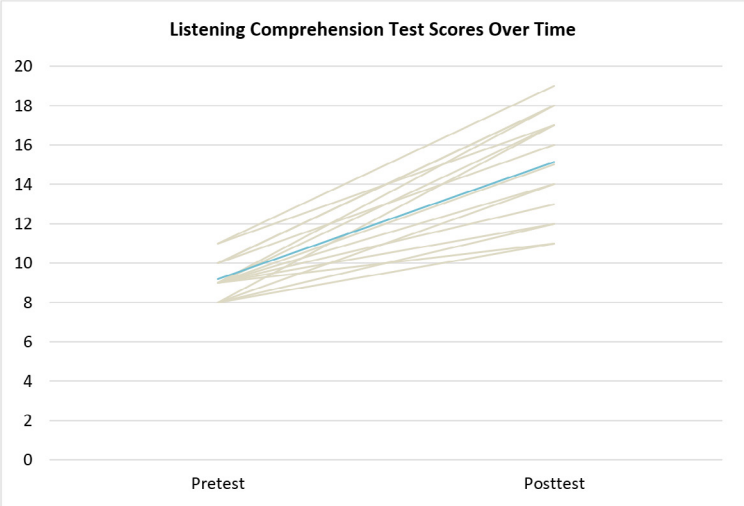
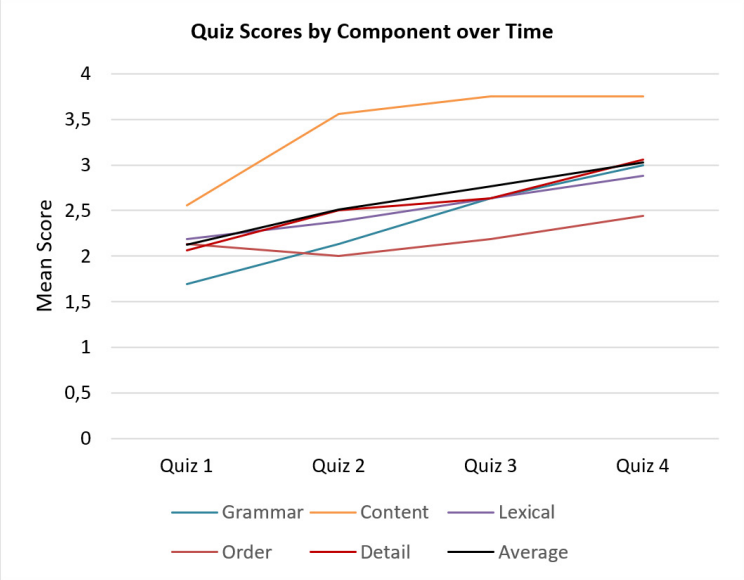


Figure 3

Participants' listening comprehension quizzes by component over time



displayed a gradual increase from $\bar{x}$ =2.31; $\bar{x}$ =3.03; $\bar{x}$ =3.19; and $\bar{x}$ =3.40 for their quiz 1, quiz 2, quiz 3, and quiz 4 respectively. However, the components under the writing part climbed slightly as evidenced by quiz 1 ($\bar{x}$ =2.17), quiz 2 ($\bar{x}$ =2.17), quiz 3 ($\bar{x}$ =2.48), and quiz 4 ($\bar{x}$ =2.77). The result in listening part might be attributed to the fact that when they use metacognitive strategies and allowed to translanguage during collaboration, they were able to obtain more ideas and information from their peers and thus, deeper understanding of the content is achieved albeit new listening selection is discussed the next class.

Table 2 reports all pairwise comparisons between means on the four quizzes. All comparisons were significant at $p < .001$ even after applying a bonferroni correction, indicating that scores between all quizzes were significantly different from one another. In particular, scores significantly increased over time (*i.e.* the mean Quiz 1 score was significantly lower than Quiz 2; the mean Quiz 2 score was significantly lower than Quiz 3; the mean Quiz 3 score was significantly lower than Quiz 4).

Descriptive Statistics for Participants' Perception on the Use of the Intervention (RQ3)

Table 3 presents the participants' perceptions regarding three aspects: Translanguaging as a practice, Translanguaging for L2 learning, and Translanguaging during metacognitive strategy use for facilitating listening comprehension tasks. Participants strongly agreed with statement #2, indi-

Figure 4

Students' Quiz Performances by Component and by Distinction between the Two Skills

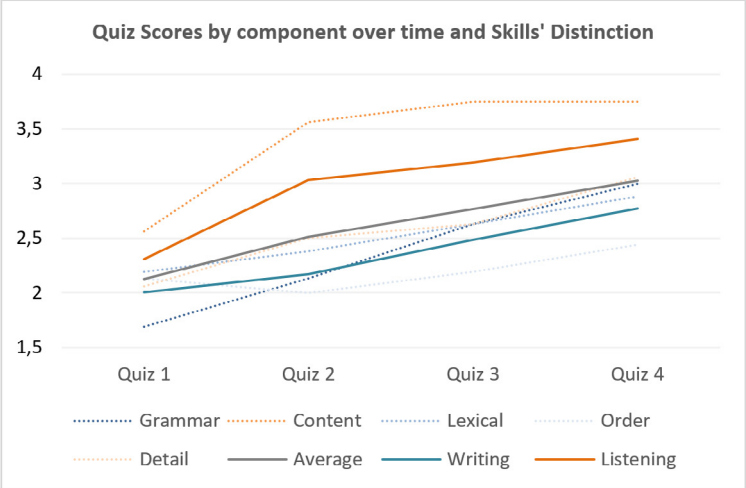


Table 2

Pairwise Comparisons of the Means of the Four Quizzes

First Quiz	Second Quiz	Mean Improvement	SE	95% CI for Mean Improvement	
1	2	1.81	0.31	0.88	2.74
1	3	3.25	0.35	2.20	4.31
2	3	1.44	0.22	0.76	2.12
1	4	4.44	0.39	3.26	5.61
2	4	2.63	0.18	2.08	3.17
3	4	1.19	0.25	0.44	1.93

Note. All comparisons are $p \leq .001$ after Bonferroni corrections for multiple comparisons applied.

cating that translanguaging is a normal practice (mean=4.75, SD=0.40). Conversely, they strongly disagreed with statement #4, suggesting that translanguaging is disrespectful (mean=1.00, SD=0.00). These results indicate that participants view translanguaging as a normal pedagogy that can be used without negatively impacting the language they are studying.

In terms of "Translanguaging for L2 learning," participants expressed positive views, strongly agreeing with statements #1, #3, and #4, which highlight the benefits of translanguaging in learning English. They strongly disagreed with statements #2 and #5, indicating that translanguaging is not a sign of low proficiency in the target language and that language instructors should not avoid using it in the classroom. This suggests that participants perceive translanguaging as an effective pedagogy for facilitating listening comprehension, writing tasks, and L2 learning.

Regarding the last aspect, participants strongly agreed that translanguaging during metacognitive strategy use aids in facilitating their listening comprehension and writing tasks. Their responses to all statements in this category were very positive. Statement #2 received the highest mean score of 4.81 (SD=0.40), indicating that translanguaging helps participants understand content and discuss listening difficulties with their peers. Statement #6 received the lowest mean score of 3.50 (SD=0.50), indicating that translanguaging provides opportunities to resolve discrepancies and adapt strategies.

Qualitative Analysis

Experiences the Participants Have Gained in Improving Their Listening Comprehension Performance and L2 Learning after the Strategy Intervention is Implemented (RQ4)

Theme 1: Students' Perceptions on translanguaging practices in the EFL classroom

This theme focuses on the participants' perspectives regarding the utilization of translanguaging in the EFL class-

Table 3

Mean and Standard Deviation for Students' Perception on the use of TLAN during Metacognitive Strategy Use

	Statements	Mean	SD	Description
	TLAN as a Practice			
1.	Translanguaging should be avoided by bilinguals.	1.19	0.40	Strongly disagree
2.	Translanguaging is a normal practice for bilinguals	4.75	0.45	Strongly agree
3.	Translanguaging indicates a lack of linguistic proficiency in our second language.	1.75	0.68	Strongly disagree
4.	Translanguaging is a disrespectful practice.	1.00	0.00	Strongly disagree
5.	Translanguaging is confusing for me.	1.25	0.45	Strongly disagree
6.	It is fine to apply translanguaging in listening tasks	4.00	0.73	Agree
	TLAN for learning Second Language (L2)			
1.	Translanguaging helped me learn the English language.	4.56	0.51	Strongly agree
2.	Translanguaging is a sign of low proficiency in L2	1.25	0.45	Strongly disagree
3.	Translanguaging is essential for learning both L1 and L2.	4.50	0.52	Strongly agree
4.	Translanguaging has assisted me in learning English.	4.75	0.45	Strongly agree
5.	Language instructors should avoid translanguaging because it will prevent L2 learning.	1.31	0.48	Strongly disagree
	TLAN during metacognitive strategy use			
1.	The use of translanguaging during metacognitive strategy use aids me link my background knowledge to the new topic.	4.06	0.25	Agree
2.	The use of translanguaging during metacognitive strategy use helps me discuss strategies with my peers.	4.81	0.40	Strongly agree
3.	The use of translanguaging during metacognitive strategy use provides an opportunity for us to resolve discrepancies, and adapt our strategies.	4.25	0.77	Strongly agree
4.	The use of translanguaging during metacognitive strategy use aids me link my background knowledge to the new topic.	4.13	0.81	Agree
5.	The use of tranlanguaging during metacognitive strategy use helps me discuss strategies with peers	4.44	0.51	Strongly agree
6.	The use of tranlanguaging during metacognitive strategy use provides opportunity for us to resolve discrepancies and adapt our strategies.	3.50	0.52	Agree
7.	The use of tranlanguaging during metacognitive strategy use assists us to share and compare what we have understood regarding the listening selection.	4.38	0.72	Strongly agree

room. The sub-themes identified were as follows: enhanced engagement with the concept, a more natural way of expressing ideas, and increased awareness of the structures and forms of their L1. When asked about their perception of using translanguaging in processing their listening tasks, participants expressed that translanguaging allowed for a deeper engagement with the topic compared to an English-only policy. It is evident that mental processes involved in L2 learning, such as comprehension, analysis, and reasoning, cannot occur solely in the L2 (Rivera & Mazak, 2017). By encouraging students to engage in ideas and utilize all the languages they know, deeper conceptual understanding and language proficiency can flourish. One respondent (R9) shared their experience as follows:

> *"Translanguaging allows us to engage in meaningful conversations and eases the process of clarifying and negotiating ideas. Solely depending on English as the language of thought can be quite challenging."* **R9**

Translanguaging showcases the multifaceted linguistic abilities of L2 learners and users (Cook & Li, 2016). When participants are given the freedom and encouragement to translanguage and utilize their entire linguistic repertoire while sharing and collaborating, they are able to tap into their full potential and engage in discussions in a more empowered and natural manner. Respondent 7 conveyed his view:

> *"Because we don't hesitate to convey our thoughts in English and use Thai, Isarn, Lanna, it has enabled us to engage in discussions more naturally and fluently."* **R7**

When students were given the opportunity to engage in translanguaging, they actively participated and perceived it as a valuable learning experience. Translanguaging facilitated a deeper understanding of the content, enabling them to engage in more critical thinking. As a result, they became more aware of the structures and forms of their L1 and attempted to connect and comprehend the new concepts in relation to their existing knowledge. This process provided them with a fresh perspective and allowed them to unpack their understanding of their L1. One participant expressed:

> *"Using Thai or Isarn language during the sharing activity made me acutely aware of my L1, including Thai idioms and complex expressions that are challenging to directly translate into L2. This self-examination has heightened my critical thinking compared to before."* **R11**

Theme 2: Benefits of applying translanguaging during metacognitive strategy use on their listening comprehension task processes

This theme encompasses the participants' feedback regarding the advantages of employing translanguaging during the utilization of metacognitive strategies in their listening tasks. The sub-themes identified are as follows: improved planning activities, heightened engagement in collaboration activities, and the effectiveness of monitoring, evaluation, and reflective strategies in achieving better listening outcomes. During the planning and predicting stage, students were exposed to activities such as reading related texts and making predictions, which allowed them to activate their background knowledge and share it with peers using their native language (L1). As a result, participants obtained more information about the topic and demonstrated increased involvement with the upcoming listening content. Respondents 3 and 8 provided their perspectives on this matter:

> *"Utilizing the reading of a relevant short text as a planning technique has assisted me in generating more ideas and information related to the upcoming selection. It also facilitates the connection of my prior knowledge to the current topic."* **R3**

> *"While employing the planning technique, I had the opportunity to translanguage, which greatly enhanced my comprehension of what to anticipate in the new topic."* **R8**

During the second phase of the initial verification stage, participants engaged in peer collaboration. They were encouraged to utilize their native language (L1) to share and discuss their comprehension of the listening selection. This led to increased enthusiasm in expressing ideas, discussing opinions, identifying listening problems, and making plans for future listening activities. Respondent 10 expressed their experience in the following manner:

> *"I gained a sense of empowerment and increased confidence when participating in monitoring activities. This was largely due to the freedom to use my L1 to express myself among peers, which facilitated more in-depth exchanges as we could better understand each other."* **R10**

Furthermore, during the final listening stage, the learners listened to the selection again in order to identify any discrepancies, make corrections to their written work, and take note of additional information they understood. In essence, they engaged in monitoring, evaluation, and problem-solving by collaborating with the entire group to discuss, reconstruct, and reflect on strategies that were effective in addressing specific listening challenges for future listening sessions. Translanguaging was encouraged in these monitoring and evaluating activities to enable them to freely express their ideas. Participants R4, R5, and R2 shared their experiences as follows:

> *"Monitoring strategies promoted collaborative peer engagement, performance assessment, and the identification of listening challenges. Translanguaging further facilitated the confident exchange of ideas and enhanced communication effectiveness."* **R4**

> *"The use of translanguaging in collaborative monitoring activities deepened our comprehension of the topic's context and boosted our proficiency in applying listening strategies."* **R5**

> *"During evaluation, regrouping enhanced problem identification and strategy discussion. Translanguaging eliminated language barriers, enabling meaningful communication." R2*

Theme 3: Improved L2 writing and L2 learning

The sub-theme that emerged from the study was related to gaining a deeper understanding of various aspects of language. The final phase of the listening comprehension activity involved students expressing their comprehension of the listening selection through written compositions. Once again, the students were grouped together and collaborated to identify and discuss errors in their papers. This included examining grammar and sentence structure, vocabulary usage, logical connections, supporting details, and overall content. To facilitate smooth communication, students were allowed to use their native language (L1) within the group. This activity heightened their awareness of mistakes and motivated them to avoid repeating them, resulting in improved L2 writing and learning, as evidenced by their quiz results. Two participants, specifically Respondents 7 and 13, expressed the following:

> *"My peers were instrumental in identifying and correcting errors in my paper during evaluation. Through collaborative discussions and translanguaging, we improved comprehension together."* **R7**

> *"Using our L1 greatly improved our English learning process. I gained a deeper understanding during monitoring and evaluation, making me more cautious about avoiding similar mistakes in the future."* **R13**

DISCUSSION

The present study explored the effects of employing metacognitive strategies and allowing translanguage use during listening comprehension tasks. Study findings revealed that utilizing translanguaging alongside metacognitive strategies improved learners' comprehension of listening material and enhanced their L2 writing skills. To illustrate:

Participants' Listening Comprehension Performance before and after the Implementation of Translanguaging during Metacognitive Strategy Use

The participants showed a positive improvement in listening comprehension performance based on their pre- and post-test results. This improvement is likely attributable to the utilization of translanguaging during the metacognitive strategy use, which aided in improving both their listening comprehension and writing skills. Translanguaging facilitated participants in expressing their ideas while comprehending listening selections with peers, making their listening tasks more enjoyable and productive. These findings align with previous studies by Chen et al. (2019), Elashhab (2020), and Sulaiman et al. (2020), which also supported the benefits of translanguaging in bilingual classrooms. These studies found that encouraging students to use their linguistic repertoires activated their prior knowledge, facilitated the exchange of information with peers, boosted confidence and communication fluency, and promoted higher levels of knowledge processing. Notably, the implementation of translanguaging in L2 classrooms supports sense-making and meaning negotiation among learners themselves and with their teachers.

Chen et al. (2019), Elashhab (2020), and Sulaiman et al. (2020) conducted studies that primarily focused on implementing translanguaging in specific areas like writing tasks, communication, and both oral and written exercises. In contrast, the present study employed translanguaging during metacognitive strategy use for both listening and writing tasks, encouraging participants to utilize their linguistic repertoires to express their comprehension of the listening selection, in line with findings by Chen et al. (2019), Elashhab (2020), and Sulaiman et al. (2020). Hence, it can be hypothesized that the participants' improved listening comprehension performance can be attributed to the effective use of translanguaging pedagogy during metacognitive strategy use, which enhanced their performance in both listening comprehension and writing tasks. Prior research has shown that structured procedures, encompassing planning, monitoring, and evaluation stages, enhance learners' control, confidence, and overall proficiency. These procedures allow learners to activate background knowledge, make comparisons, discuss differences, identify listening problems, and formulate plans for subsequent listening processes (Goh, 2008; Robillos, 2019; Robillos & Phantharakphong, 2020; Vandergrift & Tafaghodtari, 2010; Vandergrift & Go, 2012). However, it is essential to consider potential practice effects in this study. While students wrote essays on the same topic for both the pre-test and post-test, different topics were used for each quiz. Essentially, pre-test essay scores and quiz scores reflected students' initial attempts at composing on new topics, whereas the post-test essay did not.

Participants' Listening Comprehension and Writing Skills in L2 as Manifested on Their Quizzes

In terms of the participants' L2 writing skills, the findings of the study revealed a significant improvement in their written work. This improvement encompassed various aspects, including the use of academic vocabulary, sentence structure formation, appropriate deployment of transitional devices, content development, incorporation of supporting details, and adherence to correct grammar. In the present study, students' comprehension of the listening selection was assessed through written paragraphs. They were subsequently grouped with their peers and, using their native language (L1), collaborated to identify and discuss mistakes in their papers, with the aim of preventing them in future compositions. This evaluation, which allowed translanguaging, likely contributed to the participants' progress in L2 learning by enabling a more thorough and meaningful error identification process, fostering greater vigilance against repeating those errors in their writing. Translanguaging empowers learners to grasp and convey complex ideas, offer comprehensive explanations, persuade, and engage in persuasive arguments on various topics seamlessly. Sobkowiak (2022) supports this perspective, emphasizing that "...the free fusion and integration of languages in the L2 classroom can enhance students' cognitive engagement with learning activities and, consequently, enhance their linguistic performance (p. 1)."

Furthermore, findings of the study highlighted significant improvements in the "grammar" component of writing, with the least improvement observed in "logical order." This discrepancy can be attributed to the students' emphasis on discussing grammar errors during the collaborative metacognitive evaluation phase. Their prior exposure to grammar-focused instruction in elementary and secondary school English classes led to more meaningful discussions about grammar-related issues. However, the study also revealed that the "logical order" component exhibited the least improvement, indicating a challenge in the students' ability to establish logical connections between ideas, formulate strong topic sentences to initiate paragraphs, use effective transitions to link sentences, and employ diverse sentence structures. Despite their participation in collaborative activities and the allowance of translanguaging, the students still lacked the necessary skills to organize their ideas logically in their writing.

Interestingly, while the "logical order" component showed the least improvement, students did make gradual progress in their 2nd, 3rd, and 4th quizzes, though these improvements remained the least among the components. This may be attributed to the increased difficulty of the intervention videos, particularly in the 2nd (*Climate change*), 3rd (*Noise pollution*), and 4th (*Overpopulation*) informative video items, which presented unfamiliar vocabulary and terms that challenged the students in expanding their un-

derstanding of these concepts. Nevertheless, the opportunity to translanguage during collaborative activities in the monitoring and evaluation stages allowed them to enhance their vocabulary and gain better comprehension of these ideas and information. Sobkowiak (2022) emphasized that engaging students in discussions and utilizing their linguistic repertoire, encompassing all the languages they know, can foster the development of conceptual knowledge and language proficiency. Similarly, various researchers and practitioners in SLA, including Bozorgian et al. (2021), Goh (2008), Robillos & Bustos (2022), Robillos & Thongpai (2022), and Vandergrift & Tafaghodtari (2010), have recognized the positive impact of learners' metacognitive awareness of their learning processes, cognitive functions, and strategic use in learning tasks.

Participants Perceptions on the Use of Translanguaging as Practice, Translanguaging for L2 Learning, and Translanguaging for Listening Comprehension Task Processes

Furthermore, participants in the present study did not perceive the use of translanguaging as disrespectful, even though they were aspiring English language teachers. Instead, they viewed translanguaging as a strategic and helpful tool for effectively grasping concepts and acquiring knowledge in an L2. This perspective is supported by Moody et al. (2019), who argue that students' application of translanguaging is a natural linguistic resource that does not hinder successful communication. Additionally, the participants considered the use of translanguaging in listening comprehension activities to be a normal practice that did not hinder their ability to use the L2. Moody et al. (2019) similarly found that participants in their study held neutral opinions regarding whether translanguaging should be avoided when confusing or seen as a disrespectful practice. Interestingly, the students in the present study had a positive view of translanguaging, understanding that it does not indicate a lack of proficiency in a L2. Rivera and Mazak (2017) emphasized that even proficient L2 learners regularly employ their L1 in their thinking processes. In fact, using L1 is commonplace in bilinguals' writing when planning, monitoring, and evaluating their performance, and listening follows a similar pattern as writing in terms of these cognitive stages.

Encouraging students to embrace translanguaging, the practice of using both their L1 and L2 interchangeably, yielded several profound benefits. This approach not only fostered active student engagement but also facilitated a deeper and more critical understanding of the subject matter. While English and Thai served as the primary languages for translanguaging within the classroom, it was notable that students attempted to bridge language gaps by experimenting with other languages like Isarn, Lanna, and Phu Thai, even if they weren't fully fluent in these secondary languages. When faced with language barriers, students employed various communication strategies, including hand gestures, fragmented phrases, and word root exploration, all in an effort to convey their thoughts, especially when confronted with unfamiliar or intriguing concepts. Beyond mere communication, translanguaging brought about increased linguistic awareness among students. They became more attuned to the structures and nuances of their native language (L1) and began making meaningful connections between new ideas introduced in the target language (L2) and their existing knowledge in their native language. This approach transcended the conventional notion of language learning solely as a means of overcoming language barriers; it became a way for students to delve into the intricate linguistic aspects of their L1 within diverse social contexts. Furthermore, this perspective resonates with García's (2017) view, emphasizing that bilingual individuals don't compartmentalize languages as separate entities. Instead, they seamlessly draw upon their entire linguistic repertoire, encompassing all the languages they know, to construct meaning and communicate effectively. In essence, languages cease to exist as separate entities and become integral components of a unique linguistic repertoire, enriching the learning process and deepening the students' understanding of language and culture.

CONCLUSION

In conclusion, our study sheds light on the promising potential of utilizing translanguaging during metacognitive strategy use in the Thai EFL classroom. The researcher has witnessed firsthand how this approach can significantly enhance students' comprehension and processing of listening and writing tasks in their L2. Importantly, it elevates their metacognitive abilities, empowering them to think, understand, monitor, and evaluate at a higher level.

What stands out from the study findings is the resounding consensus among participants: incorporating their linguistic repertoires during metacognitive strategy use is a valuable and effective tool for English language learning. It amplifies their grasp of grammar, vocabulary, content, logical connections, and supporting details in L2 writing. It's time for EFL educators to reconsider the dogmatic adherence to English-only policies and recognize the rich resource that students' L1 and other languages represent. However, we must tread carefully; overreliance on the L1 is a legitimate concern. To address this, teachers must become discerning guides, knowing when to permit students to leverage their linguistic diversity and when to steer them towards the target language. With this balance, learners will come to view translanguaging as a strategic learning aid, not a crutch.

Furthermore, the study underscores the role of translanguaging in facilitating metacognitive strategies within the EFL classroom. By integrating metacognitive practices like planning, monitoring, and evaluation into listening tasks

and encouraging the use of linguistic resources, teachers not only offer valuable support for task completion but also foster students' confidence and meaningful expression. The inclusion of translanguaging leads to skill development in planning, monitoring, evaluation, and reflection, nurturing metacognitive awareness and giving students greater control over their learning. In this supportive language environment, students can freely articulate their ideas without the constraints of language anxieties, resulting in deeper and more meaningful learning experiences.

Despite the insights gained, the present study has limitations. The small sample size, consisting of only 16 participants, restricts the generalizability of our findings. To strengthen the conclusions and validate our claims, future research should seek a larger and more diverse participant pool. Additionally, it would be worthwhile for forthcoming studies to explore the fusion of translanguaging with digital innovations, such as Web 2.0 applications and blended learning. In an age where technology is integral to education, investigating the synergy between translanguaging and digital tools could unveil new dimensions of language learning potential.

In conclusion, our research underscores the transformative power of embracing linguistic diversity in the EFL classroom. It beckons educators to reevaluate their language policies and encourages students to harness their multilingual resources effectively. Translanguaging, when judiciously applied, not only enhances metacognitive skills but also empowers students to take charge of their language learning journey. As we look to the future, our findings beckon further exploration into the dynamic relationship between translanguaging and digital tools, promising an even brighter horizon for language education.

DECLARATION OF CONFLICTING INTERESTS

None declared.

REFERENCES

Bozorgian, H. (2014) Less-skilled learners benefit more from metacognitive instruction to develop listening comprehension. *International Journal of Research Studies in Language Learning, 4*(1), 3-12 https://dx.doi.org/10.5861/ijrsll.2014.748

Bozorgian, H., Fallahpour, S., & Alinasab Amiri, M. (2021). Listening for young-adult EFL learners: Metacognitive intervention through L1. *International Journal of Listening*, 1-14. https://doi.org/10.1080/10904018.2021.1923499

Carpenter, J., & Matsugu, S. (2020). Translanguaging in project-based language learning. *Global perspectives on project-based language learning, teaching, and assessment: Key approaches, technology tools, and frameworks* (pp. 49-68). Taylor and Francis.

Chen, F., Tsai, S. C., & Tsou, W. (2019). The application of translanguaging in an English for specific purposes writing course. *English Teaching & Learning, 43*(1), 65-83. https://doi.org/10.1007/S42321-018-0018-0

Cohen, A. D. (1995). In which language do/should multilinguals think? *Language, Culture and Curriculum, 8*(2), 99-113. https://doi.org/10.1080/07908319509525194

Cook, V., & Li, W. (Eds.). (2016). *The Cambridge handbook of linguistic multi-competence*. Cambridge University Press.

Creswell, J. W., & Plano-Clark, V.L. (2011). *Designing and conducting mixed method research* (2nd ed.). Sage Publications, Inc. https://doi.org/10.1080/15424065.2022.2046231

Csillik, É., & Golubeva, I. (2020). Translanguaging practices in early childhood classrooms from an intercultural perspective. *In International perspectives on modern developments in early childhood education* (pp. 15-39). IGI Global. https://doi.org/10.4018/978-1-7998-2503-6.ch002

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Multilingual Matters.

Elashhab, S. (2020). The impact of translanguaging on the EFL competence development of Arabic speaking learners. *The Asian EFL Journal, 27*(3.1), 393-413.

Flavell, J. H. (1979) Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34. 10.906

García, O. (2017). Translanguaging in schools: Subiendo y bajando, bajando y subiendo as afterword. *Journal of Language, Identity, and Education, 16*(4), 256-263. https://doi.org/10.1080/15348458.2017.1329657

García, O., Johnson, S.L., & Seltzer, K. (2017). *The translanguaging classroom: Leveraging student bilingualism for learning*. Caslon. https://doi.org/10.21283/2376905X.9.165

García, O., & Li, W. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.

García, O., & Otheguy, R. (2020). Plurilingualism and translanguaging: Commonalities and divergences. *International Journal of Bilingual Education and Bilingualism, 23*(1), 17-35. https://doi.org/10.1080/13670050.2019.1598932

Goh, C. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implications. *RELC Journal: A Journal of Language Teaching and research in Southeast Asia, 39*(2), 188-213. https://doi.org/10.1177/0033688208092184

Graham, S. (2017). Research into practice: Listening strategies in an instructed classroom setting, *Language Teaching, 50*(1), 107–119. https://dx.doi.org/10.1017/S0261444816000306

Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39*(1), 9–30. https://doi.org/10.1093/applin/amx039

Lin, A. M. Y., & He, P. (2017). Translanguaging as dynamic activity flows in CLIL Classrooms. *Journal of Language, Identity and Education, 16*(4), 228-244.

Mgijima, V. D., & Makalela, L. (2021). Developing summary writing skills through translanguaging, *South African Journal of African Languages, 41*(2), 196-206. https://doi.org/10.1080/02572117.2021.1948223

Moody, S., Chowdhury, M., & Eslami, Z. (2019). Graduate students' perceptions of translanguaging. *English Teaching & Learning, 43*(1), 85-103. https://doi.org/10.1007/S42321-018-0019-Z

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524490

Oxford, R.L. (1993). Research update on teaching L2 listening. *System, 2*(2), 2005-2011.

Otheguy, R., García, O., & Reid, W. (2015). Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review, 6*(3), 281-307. https://doi.org/10.1515/applirev-2015-0014

Robillos, R. J. (2019). Crossing metacognitive strategy instruction in an EFL classroom: Its impact to Thai learners' listening comprehension skill and metacognitive awareness. *Asian EFL Journal, 21*(2), 311-336.

Robillos, R.J. (2020). Instruction of metacognitive strategies: Its role on EFL learners' listening achievement and awareness of their metacognitive listening strategies and selfregulation. *Asian EFL Journal, 27*(3.2), 442-452.

Robillos, R.J. & Phantharakphong, P. (2020). Enhancing EFL learners' argumentative abilities in written composition and critical thinking dispositions through argument mapping within metacognitive approach. *Asian EFL Journal, 27*(3.3), 181-208.

Robillos, R. J., & Thongpai, J. (2022). Computer-aided argument mapping within metacognitive approach: Its impact on students' argumentative writing performance and self-regulated learning. *LEARN Journal: Language Education and Acquisition Research Network*, *15*(2), 160-186. . https://so04.tci-thaijo.org/index.php/LEARN/index

Robillos, R. J., & Bustos, I. G. (2022). Learners' listening skill and metacognitive awareness through metacognitive strategy instruction with pedagogical cycle. *International Journal of Instruction*, *15*(3), 393-412. https://doi.org/10.29333/iji.2022.15322a

Rivera, A. J., & Mazak, C. M. (2017). Analyzing student perceptions on translanguaging: A case study of a Puerto Rican university classroom. *How, 24*(1), 122–138. http://dx.doi.org/10.19183/how.24.1.312.

Rost, M. (2001). Listening. In R. Carter, & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp 7-13). Cambridge University Press.

Seals, C. A., Newton, J., Ash, M., & Nguyen, B. T. T. (2020). Translanguaging and Task- based language teaching: Crossovers and challenges. In *Envisioning TESOL through a translanguaging lens* (pp. 275-292). Springer.

Sobkowiak, P. (2022). Translanguaging practices in the EFL classroom- the Polish context. *Linguistics and Education, 69*, 1-10. https://doi.org/10.1016/j.linged.2022.101020

Tian, Z., Aghai, I., Sayer, P., & Schissel, J. (2020). Envisioning TESOL through a translanguaging lens in the era of post-multilingualism. In Z. Tian, L. Aghai, P. Sayer, & J. Schissel (Eds.), *Envisioning TESOL through a translanguaging lens- global perspectives* (pp. 1-20). Springer

Turnbull, B. (2018). Reframing foreign language learning as bilingual education: Epistemological changes towards the emergent bilinguals. *International Journal of Bilingual Education and Bilingualism, 21*(8), 1041-1048. https://doi.org/10.1080/13670050.2016.1238866

Vandergrift, L., & Tafaghodtari, H.M. (2010). Teaching L2 students how to listen does make a difference: An empirical study. *Language Learning: A Journal of Research in Language Studies. 60*(2), 470-497. https://doi.org/10.1111/j.1467922.2009.00559.x

Vandergrift, L., & Goh, C (2012). Teaching and learning second language listening: Metacognition in action**.** *ELT Journal, 68(*2), 211–213. https://dx.doi.org/10.1093/elt/ccu003

Williams, C. (1994). *An evaluation of teaching and learning methods in the context of bilingual secondary education* [Unpublished doctoral dissertation]. University of Wales.

Zheng, J. (2018). The metacognitive strategy in English listening comprehension. *Theory and Practice in Language Studies, 8*(2) 226-231. http://dx.doi.org/10.17507/tpls.0802.07

APPENDIX A

Writing Rubric for Quizzes

Criteria	4 points	3 points	2 points	1 point
Grammar and Structure	Maintain a high degree of grammatical and structural accuracy; errors are difficult to be spotted	Use compound grammatical and structural forms to express ideas. Does not make errors which cause misunderstanding	Use simple grammatical and structural forms to express ideas but still make basic mistakes	Limited control of a few simple and structural structure
Content	Demonstrate clear understanding of information in the short informative video clip	Demonstrate adequate understanding in the short informative video clip	Demonstrate basic information in the short informative video clip	Demonstrate little or no understanding in the short informative video clip
Lexical Resource	Use a range of appropriate, relevant, and innovative vocabulary to express ideas	Use a range of appropriate and related vocabulary to present the ideas	Use appropriate vocabularies to express ideas; some irrelevant vocabularies may be spotted	Limited vocabulary using isolated words and phrases
Logical order	Details are in logical order	Ideas are in logical order	Ideas are in random order and not logical	Ideas are not in a logical order
Supporting Details	All important details are included	Important details are included but some might be missing	Some critical information is missing	Contains only some details

APPENDIX B

Questionnaire (part 1)

Name (optional)_____

Please put a check mark (✓) on the space next to the answer of your choice or write in the space provided as the case may be.

Age

( ) 18 and below ( ) 19-20 ( ) 21 and above

1. What is your First Language?

( ) Thai ( ) Phu Thai

( ) Isarn ( ) others, pls specify

( ) Lanna

2. Languages you speak proficiently. Mark as many if you speak those languages proficiently.

( ) Thai ( ) French

( ) English ( ) Japanese

( ) Chinese ( ) others, pls specify_____

3. How much experience do you have in using English?

( ) below 1 year ( ) 5-6 years

( ) 1-2 years ( ) 7-8 years

( ) 3-4 years ( ) over 9 years

APPENDIX C

Writing Rubric for Pre- and Post- Tests

Writing Components	Scales			
	4 points	**3 points**	**2 points**	**1 point**
Main Idea	Clearly states the overall main idea of the video clip	Clearly states the main idea of the video clip	Main idea is unclear – not specifically stated in the writing	The main idea is not present
Supporting Details	All important details are included	Important details are included but some might be missing	Some critical information is missing Contains only some details	Contains only some details
Logical Order	Details are in logical order	Ideas are in logical order	Ideas are in random order and not logical	Ideas are not in a logical order
Content	Demonstrate clear understanding of information in the video clip	Demonstrate adequate understanding in the video clip	Demonstrate basic information in the video clip	Demonstrate little or no understanding in the video clip
Paraphrasing	Is characterized by paraphrasing of the main idea and significant details	Is characterized by paraphrasing of the main idea and significant details	Is characterized by substantial copying of key phrases and minimal paraphrasing	Is characterized by substantial copying of indiscriminately selected phrases or sentences

https://doi.org/10.17323/jle.2023.13245

Perceptions of Situational Factors of Willingness to Communicate Inside and Outside the Classroom: Thai EFL First-Year University Students' Reflections

Satima Rotjanawongchai

Prince of Songkla University, Trang Campus, Thailand

ABSTRACT

Background: Although previous studies have reported WTC variables in the EFL context, limited studies have investigated learners' perceptions of WTC outside the classroom. In addition, insights into learners' perceptions from qualitative data have rarely been presented in this study area.

Purpose: This study investigated the perceptions of willingness to communicate (WTC) of Thai English as a Foreign Language (EFL) first-year university students. It focused on their perceptions of situational factors of WTC inside and outside the classroom. Also, it investigated the reasons behind the perceptions.

Methods: As a mixed-method study, a questionnaire adapted from Baghaei's (2013) and Peng and Woodrow's (2010) was used to collect quantitative data, while reflective reports and semi-structured interviews were used to reveal explanations for the quantitative data.

Results: The present study found that the students were more willing to speak outside the classroom than inside the classroom. Interlocutors were significant factors affecting WTC both inside and outside the classroom. Also, speaking topics that are suitable for students' perceived level of proficiency but still pose a challenge for their language development have the potential to increase WTC in the classroom. At the same time, a stimulating environment was powerful for WTC outside the classroom due to a lack of an English-speaking environment in the EFL context. The qualitative data revealed that foreign language anxiety concerning the interlocutor's competence, familiarity with the interlocutor, and language classroom experiences, as well as social support from friends, were the rationale behind the impact of the situational factors.

Conclusion: Teachers can apply the results of this present study to enhance WTC in the classroom and increase students' opportunities to speak inside and outside the classroom through pedagogical support.

KEYWORDS

willingness to communicate (WTC), perceptions, situational variables, Thai EFL context, students' reflections

Citation: RotjanawongchaiS. (2023). Perceptions of Situational Factors of Willingness to Communicate inside and outside the classroom: Thai EFL First-Year University Students' Reflections. *Journal of Language and Education, 9(*3), 128-142. https://doi.org/10.17323/jle.2023.13245

Correspondence:
Satima Rotjanawongchai,
katibbchan@gmail.com

Received: October 25, 2021
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

Student reticence to speak in the English language classroom has been reported as one of the most common classroom problems. It can be considered a normal phenomenon, especially in Asian contexts such as Japan (Donald, 2010; Talandis & Stout, 2015), China (Li & Liu, 2011; Zhang & Head, 2010), Taiwan (Chang, 2011), Hong Kong (Jackson, 2002), Iran (Doqaruni, 2015; Riasati, 2014), and Thailand (Pattapong, 2015). Recent studies have suggested that silence in the classroom may contribute to the development in language learning as it might be considered as space for "attentive listening, thinking, and reformulating ideas" (Harumi, 2020, p. 39) and as a way to "keep the classroom dynamics in harmony" (Chung, 2021, p.79). However, encouraging language output in the classroom is still essential since only language input may not be sufficient for learners to produce clear and communicative language, which is

the aim of language learning and communication (Birkner, 2021).

The problem of reticence seems to deteriorate in EFL contexts in which English is learned and used mainly in the classroom, and opportunities to communicate in English are rare. Research on willingness to communicate (WTC) has investigated the reasons behind this phenomenon. Widely-referenced WTC models, such as those of MacIntyre & Charos (1996) and MacIntyre et al. (1998), as well as the findings of studies on WTC in various EFL contexts, have highlighted the intertwined state of variables affecting WTC. However, few studies have reported WTC outside the classroom from learners' perspectives, especially in the Thai context. Considering the context when adopting a teaching idea or suggestion is essential because a sound methodology is one that is suited to the context (Kumaravadivelu, 2001). In addition, having a better understanding of the teaching context, especially the learners and situational variables, can fill the gap between educational research and practice (Vanderlinde & van Braak, 2010). For this reason, investigating learners' perceptions of situational factors of WTC both inside and outside the classroom may shed light on a potentially more effective way to deal with the problem of reticence and unwillingness to speak.

The present study aimed to address the research gap by answering the following questions:

(1) What are the Thai first-year university EFL learners' perceived WTC inside and outside the classroom and their perceptions of WTC according to the four constructs (receivers, topics, task types, and speaking environment)?

(2) What are the rationales behind those perceptions?

LITERATURE REVIEW

L2 Use and Willingness to Communicate (WTC)

WTC is defined as the probability of taking opportunities to initiate conversations (McCroskey, 1992). Studies on WTC initially focused on L1 communication. The literature on L1 WTC highlights personality traits as significant factors of WTC (MacIntyre, 2007). Regarding L2 WTC, however, it is found that personality traits affect L2 WTC, albeit indirectly (MacIntyre & Charos, 1996). This is also demonstrated in MacIntyre et al.'s (1998) definition of L2 WTC, where they describe it as a state of readiness to engage in conversation with particular individuals or groups using a second language (L2) (p. 547).

While the concept of WTC initially centred on L1 communication and highlighted personality traits as significant factors (MacIntyre, 2007), the understanding of L2 WTC reveals a more complex interplay of variables, as demonstrated in MacIntyre et al.'s (1998) six-layered model. MacIntyre et al. (1998) integrated linguistic, communicative, and social psychological variables affecting L2 WTC into a six-layered model with twelve constructs. On the first layer, at the bottom, are intergroup climate and personality, which are the most remote influences. On the second layer are intergroup attitudes, social situations, and communicative competence. The third layer is composed of motivation and self-confidence. The fourth layer illustrates the most immediate influences on WTC: the desire to communicate with a specific person and state communicative self-confidence to that person. Above this, WTC is posited as the most immediate influence on communicative behaviours (L2 use). It was hypothesised that these variables influence WTC, which, in turn, predicts L2 use. This construct shows that situational factors such as interlocutors exert more influence on L2 communication than trait-like factors such as personality. This accords with many empirical studies such as Yashima (2002) and Clément et al. (2003). This model was widely used as the basic conceptualisation for later related studies such as Pattapong (2015) and Peng and Woodrow (2010), confirming the vital contribution of situational factors.

Situational WTC

WTC as a situational construct views situational variables as more powerful than individual variables (Kang, 2005). In the FL context, situational WTC has been increasingly highlighted probably because situational variables tend to significantly affect FL learners more than individuals' personality. According to MacIntyre et al. (1998) and MacIntyre et al. (1999), FL learners may experience more language anxiety due to the unfamiliarity of the culture and the language and have a lower level of proficiency due to their limited exposure. These differences highlight the importance of considering the unique context of FL learning when studying WTC and emphasise the potential of increasing FL WTC through situational variables. In addition, paying attention to situational variables is likely to promote WTC more efficiently because the situational WTC can drive a decision to initiate a conversation, while trait WTC tends to only push individuals to find a chance to communicate (Cao & Philp, 2006; Kang, 2005).

In accordance with situational antecedents in the L2 WTC model of MacIntyre et al. (1998), situational factors such as social environment, the task, and the communication partner are significant to FL WTC. Previous studies have shown that the interlocutor's characteristics such as familiarity with the interlocutors and the interlocutors' cooperation (Pawlak et al., 2016; Riasati, 2012) affect WTC. Focusing on WTC in the classroom, teacher support, student cohesiveness, and task orientation are found to directly influence individuals' factors i.e., communication confidence, motivation, and learner beliefs (Aomr et al., 2020; Peng & Woodrow, 2010; Weda et al., 2021). Teacher support encompasses the ways in which the teacher assists, offers encouragement, demon-

strates trust, builds relationships, and shows interest in their students (Dorman et al., 2006). Teachers' influence on WTC was also reported in terms of their attitude and teaching style (Zarrinabadi, 2014). Student cohesiveness refers to the level of mutual acquaintance, assistance, and support among students, while task orientation refers to the emphasis placed on finishing tasks and maintaining focus on the topic at hand (Dorman et al., 2006). These situational variables can result in FL classroom anxiety, which can be the strongest predictor of WTC in FL classrooms (Barrios & Acosta-Manzano, 2021; Dewaele & Dewaele, 2018) and also WTC outside the classroom (Lee & Hsieh, 2019).

While the significance of situational factors in FL WTC has been well-documented, particularly within classroom settings, it is essential to examine how these factors manifest in the specific context of Thai EFL learners, where the interplay of situational elements may differ from those in other FL contexts. Current research on WTC in the Thai EFL context has predominantly focused on classroom settings, with studies emphasizing the significance of interlocutors, interest in the topic, and other classroom-related factors. However, limited attention has been given to WTC outside the classroom, and the impact of situational factors in this context remains largely unexplored. So far, WTC research relating to situational factors in the Thai EFL context has been mostly restricted to WTC in the classroom. For example, through WTC questionnaires and semi-structured interviews, Karnchanachari (2019) explored WTC in the classroom in Thai and international programs. The study reported that interlocutors and interest in the topic were essential for WTC. Similarly, Pattapong (2010) explored factors affecting university students' WTC in the classroom and reported the effects of interlocutors, classroom management, and tasks on WTC. According to the study, familiarity with classmates played an essential role in WTC. For students with high WTC, the English competence of the interlocutor was another vital factor. It was reported that learners were more willing to talk with friends of a higher level of English competence because they valued corrective feedback from those peers. Although WTC outside the classroom has been explored in many studies in other FL contexts (i.e., Iran (Baghaei, 2013), Turkey (Basoz & Erten, 2018), Belgium (Denies et al., 2015)), research on WTC in the Thai context to date has not yet investigated WTC outside the classroom. Among the limited studies on WTC outside the classroom, social support and teacher teaching style have been found to be significant factors (Tanaka, 2007; MacIntyre et al., 2001). MacIntyre et al. (2001) studied L2 French immersion students and reported social support, especially from friends, as an important factor.

METHOD

This article presents an investigation into the WTC of first-year students at a university in Thailand. The data collection extended over a full 16-week term and followed a mixed-method design, combining qualitative and quantitative data to provide a comprehensive understanding of the subject (Creswell et al., 2011).

Participants

Forty-six first-year English majors (38 female, 8 male) enrolled in a speaking-focused course voluntarily participated in this study. According to their English scores on the Ordinary National Education Test (O-NET), which is used for university admission in Thailand, all participants could be considered at a low level of proficiency. They aged between 17 and 19 years old and averaged over ten years of learning English in the national education system prior to university admission. Since this research collected the data for a full 16-week term, which might intrude on the privacy and time of the participants, convenience sampling and volunteer sampling were adopted. The participants were drawn from the course that consisted of a single class, with no additional tutorial sessions. To implement convenience sampling, the questionnaires were distributed to all students enrolled in the course during their regular class time. Data collection took place on days when the majority of students were expected to be present, minimising potential selection bias. They received brief information about the study and the consent form during the first hour of the course. After four weeks of data collection, four out of the 50 students withdrew from the study due to personal reasons, leaving 46 students who voluntarily participated. With convenience sampling, the sample lacks clear generalisability. However, it can be useful for demonstrating the credibility of relationships among variables (Clark, 2017). To minimise the disadvantages of convenience sampling, homogenous convenience sampling in the frame of sociodemographic factors (i.e., Thai first-year university students of low English proficiency) could be used to create clearer generalisability when compared to conventional convenience sampling (Jager et al., 2017). The independent-sample t-test result indicated no significant difference between the WTC of male and female participants in the contexts of inside the classroom ($t(44)= 0.36$, $p= .724$) and outside the classroom ($t(44)= -2.20$, $p= .653$). The participants were informed about the aims of the study, the tasks they would be expected to perform, the potential consequences of participating in the research, the extent to which answers would be confidential, their right to withdraw from the study at any point, and their right to have any questions about the procedures answered (Cohen et al., 2000).

Instruments

Three data sources were adopted to investigate students' perceptions of WTC inside and outside the classroom: questionnaires investigating WTC inside and outside the classroom, reflective reports investigating topics being discussed and obstacles to speaking inside and outside the classroom,

and semi-structured interviews exploring information gleaned from the questionnaire and the reflective reports.

The questionnaire had two main parts. The first part was designed to measure WTC in the classroom in 10 situations, categorised by three receivers (teachers, close friends, and the whole class) and three task types (one-turn, probably more than one-turn, and more than one-turn conversations). The second part focused on WTC outside the classroom in 13 situations with five receivers (native speakers, non-native speakers, friends, teachers, and self). The WTC scale was taken from Baghaei's (2013) and Peng and Woodrow's (2010) studies, with slight modifications for this present study. The two scales are proposed to better fit with FL contexts since limited contact with native speakers of the target language in a foreign language learning environment contributes to differences in WTC compared to an L2 WTC context (Baker & MacIntyre, 2000). The 10-item WTC scale of Peng and Woodrow (2010) focuses on the classroom environment captured by the teacher, learners, and tasks (α = .88). Proposed to investigate FL WTC both inside and outside the classroom, the scale of Baghaei (2013) is a 22-item scale with three subscales categorised by communication with three receivers – native speakers, foreign non-native speakers, and classmates/instructors (a correlation of 0.39 with an integrative English language proficiency test, the separation reliability of 0.99). Nine items (Items 12,13, and Items 16-22) of Baghaei's (2013) scale were deleted as they repeated those in Peng and Woodrow's (2010) and some of the items in Baghaei's scale were not included (Item 6 and Item 13 involving the experiences of foreigners) as they were found to not influence WTC (Baghaei, 2013).

The questionnaire was translated into Thai by a professional translator for a better understanding of the participants who were of low English proficiency. To enhance the validity of the questionnaire, the translated version was reviewed by three ELT professionals (Nemoto & Beglar, 2014; Sudina, 2023). Their feedback was communicated to the translator who made the necessary revisions. After that, the questionnaire was piloted and revised. A reliability analysis was carried out on the WTC scale modified for this present study. Cronbach's alpha showed the questionnaire reached acceptable reliability, with α = 0.83 (WTC in the classroom) and α = 0.93 (WTC outside the classroom).

Instead of using a 6-point Likert scale and a 2-point agree/disagree scale as in Peng and Woodrow (2010) and Baghaei (2013) respectively, a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) was used to measure the participants' WTC inside and outside the classroom. As this present study focuses on the students' perceptions, it is important to provide participants with an option to express neutrality rather than forcing them to choose a positive or negative response. In this context, odd-numbered Likert scales (such as 5-point or 7-point scales) are often preferred because they provide a middle or neutral option. By doing so, participants who do not perceive something as strictly positive or negative can select the neutral option and accurately reflect their true perception (Croasmun & Ostrom, 2011). Compared to a 7 or 11-point scale, the use of a 5-point rating scale may lead to more thoughtful and differentiated responses, which can result in higher-quality data. (Revilla et al., 2014).

Apart from the questionnaire, a self-reflective report was used to collect the data. With ten open-ended questions, the report was intended to collect additional data about the students' speaking problems inside and outside the classroom. The items in the report, developed especially for this study, focused on speaking problems and WTC perceptions. The self-reported records were expected to reveal more in-depth information in the context of the speaking. It was evaluated by three ELT professionals and edited before use. To develop a fuller understanding of perceptions about WTC, six participants were invited to individual semi-structured interviews at the end of the term. To roughly group the participants according to WTC scores for the interviews, the mean scores were categorised into three groups: 4.00-5.00, 3.00-4.00, and below 3.00 since there is no one-size-fits-all approach to interpreting Likert scale data and the method of interpretation can be based on the purpose of using the data (Todd, 2011). This made for six interviewees in total. All interview questions were translated into Thai to avoid misunderstandings between the interviewers and the interviewees.

Procedure

The study was conducted during a 16-week term. The questionnaires were distributed to the students on the first day of the course and collected on the same day at a time of the students' choosing. The course was in the second term of the academic year and all of the participants had experiences of speaking English inside and outside the classroom from two compulsory speaking courses they enrolled in the first term. The instructions for the reflective reports were also given on that day. The students were told to record the problems related to speaking English they faced inside and outside the classroom and to submit a self-reflective report every four weeks throughout the term. The semi-structured interviews, based on the questionnaire results and the reflective reports, took place at the end of the term. Each interview, which was conducted in Thai and audio recorded, lasted about 40-60 minutes.

Data Analysis

The quantitative data from the questionnaire were analysed using SPSS Version 16. The mean score of each item and subscale were obtained to show the degree of WTC in each situation, focusing on different variables. T-tests and one-way ANOVAs were applied to test whether the difference in the mean scores of each variable was significant. The qualitative data obtained from the interviews and the

reflective reports were analysed using thematic analysis, which involved identifying themes and searching for commonalities and relationships. It was useful for reducing data while preserving the context (Mills et al., 2010). To enhance the credibility of the research, this study used triangulation from different data sources and member-checking, in which the participants were asked to verify the interpretation of the interviews.

RESULTS AND DISCUSSION

In this section, the descriptive statistics will be reported to answer the first research question, followed by the qualitative data to answer the second research question.

RQ1: How do Thai first-year university EFL students perceive their willingness to communicate (WTC) inside and outside the classroom, considering different factors?

Table 1 shows the WTC scores inside the classroom (the details of all items can be found in the Appendix). The data indicated that the participants were the most willing to speak when asking close friends the meaning of an English word (Item 2; *M*= 4.35, *SD*= 0.73). The second rank was when talking to close friends about a personal topic (Item 1; *M*= 4.29, *SD*=0.80) and talking to close friends about how to say an English phrase (Item 3; *M*= 4.29, *SD*= 0.74). They were the least willing to speak when doing a role-play in front of the class without notes (Item 7; *M*= 3.60, *SD*= 0.93). Table 1 also shows the frequency of each option on the Likert scale (strongly agree, agree, neutral, disagree, and strongly disagree).

In accordance with the mean score, half of the participants (50%) chose 'strongly agree' to 'I am willing to ask my close friends in English the meaning of an English word' (Item 2). Another item that most of them (50%) were strongly willing to speak was when talking to their close friends in English about a personal topic' (Item 1). The fewest number of responses of 'strongly agree' (14.6%) was 'doing a role-play in English without notes at their desk with their close friends on topics like ordering food' (Item 9). Interestingly, this is not in line with the lowest mean score, which was 'doing a role-play standing in front of the class without notes' (Item 7). It might be roughly assumed here that speaking partners (close friends in Item 9 and the class in Item 7), speaking topics (a personal topic in Item 1, the meaning of a word in Item 2, and other more challenging subjects like ordering food in Item 9) and task types (asking a simple question in Items 1 and 2, and doing a role-play without notes in Item 9) played an important role. The mean scores of these subscales (receivers, speaking topics, task types) are illustrated in Table 2.

As shown in Table 2, the participants' WTC when speaking with their close friends (*M*= 4.07, *SD*= 0.84) was higher than when speaking with the class (*M*= 3.68, *SD*= 0.88) and with teachers (*M*=3.87, *SD*= 0.85). A repeated measures ANOVA with a Sphericity correction determined that mean WTC scores differed significantly across the three subscales of interlocutors, $F(2, 90) = 6.91$, $p<0.005$. Post-hoc pairwise comparisons using the Bonferroni correction indicated a significant difference in WTC between speaking with close friends (*M*= 4.07, *SD*= .838) and speaking with the class (*M*=

Table 1

Willingness to Communicate in the Classroom

			Percentage of responses				
Item	**Mean**	**SD**	**Strongly agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
1	4.29	0.80	50	29.2	20.8	0	0
2	4.35	0.73	50	35.4	14.6	0	0
3	4.29	0.74	43.8	43.8	10.4	2.1	0
4	4.02	0.76	29.2	43.8	27.1	0	0
5	3.71	0.92	22.9	33.3	35.4	8.3	0
6	3.73	0.84	18.8	41.7	33.3	6.3	0
7	3.60	0.93	25.0	22.9	45.8	6.3	0
8	3.71	0.85	18.8	41.7	31.3	8.3	0
9	3.63	0.79	14.6	37.5	43.8	4.2	0
10	3.81	0.89	22.9	43.8	25.0	8.3	0
Total	**3.91**						

Table 2

WTC Scores in the Classroom with the Subscales

Item	Receivers	Mean	SD
1-3, 9,10	Close friends	4.07	0.84
4, 5	Teachers	3.87	0.85
6-8	Class	3.68	0.88
Item	**Topics**	**Mean**	**SD**
1-4	One-turn	4.24	0.76
5-6	Probably more than one turn	3.72	0.88
7-10	More than one turn	3.68	0.87
Item	**Task types**	**Mean**	**SD**
7, 9	Without notes	3.62	0.86
8, 10	With notes	3.76	0.88

3.68, *SD*= .878), p<0.001. It can be assumed here that talking to more familiar people resulted in a higher level of WTC. The impact of interlocutors is supported by the pyramid model of MacIntyre et al. (1998) and other studies such as those by Karnchanachari (2019) and Liu and Jackson (2008). The influence can be described by the feeling of security and less anxiety (Barjesteh et al., 2012; Kang, 2005).

As for the subscale of topics, topics that did not require them to extend a conversation, such as asking about the meaning of an English word (Items 1-4; *M*= 4.24, *SD*= 0.76), gained the highest WTC scores. The WTC scores dropped when they probably needed to 'converse' such as when asking the teacher about the lesson and when expressing opinions (Items 5-6; *M*= 3.72, *SD*= 0.88) and when they indeed needed to talk more than one turn, such as when doing a role-play (Items 7-10; *M*= 3.68, *SD*= 0.87). This is in accordance with the WTC scores of each item, as shown in Table 1. With the same interlocutor (close friends), they were likely to be more willing to ask about the meaning of a word (one-turn conversation) (Item 2; *M*= 4.35, *SD*= 0.73) rather than doing a role-play (Item 9; *M*= 3.63, *SD*= 0.79 and Item 10; *M*= 3.81, *SD*= 0.89). Across the subscales of the topic as shown in Table 2, the results of a repeated measures ANOVA with a Sphericity correction indicated a significant effect of speaking topics on WTC in the classroom, $F(2, 90) = 24.43$, p<0.001. Post-hoc pairwise comparisons using the Bonferroni correction showed that WTC when talking one turn (*M*=4.24, *SD*=0.76) was significantly higher than when engaging in a conversation that may require multiple turns (*M*=3.72, *SD*=0.88), p<0.001, and also significantly higher than when talking more than one turn (*M*=3.68, *SD*=0.87, p<0.001. This shows that although they were to speak with their close friends, their WTC tended to drop when they needed to talk in more than one turn or sentence.

Regarding task types, doing a role-play with notes received a higher mean score than doing a role-play without notes. However, a paired t-test was conducted to compare the mean between the two subscales (*M*=3.62, *SD*=0.86; *M*=3.76, *SD*=0.88) and the results indicated no significant difference, $t(45)=-1.82$, p>0.05. Whilst low-proficiency students are likely to use talk scripts during classroom-based speaking activities (Tantiwich & Sinwongsuwat, 2021), this present study found that their inclination to engage in verbal communication may not be exclusively determined by the deployment of such scripts. Rather, this willingness may be influenced by various factors, including but not limited to the conversation partners and the topics being discussed.

Table 3 shows the descriptive statistics of WTC scores outside the classroom (the details of the items can be found in the Appendix). A paired-sample t-test was conducted to compare WTC in the classroom (*M*=3.91, *SD*=0.53) and outside the classroom (*M*=4.10, *SD*=0.62). The results indicated a significant difference in WTC inside and outside the classroom, $t(45)= 1.89$, $p < 0.05$. The reasons why the present study's participants had higher WTC outside the classroom were investigated in the interviews.

Considering the receiver subscales of WTC outside the classroom, they were most willing to speak with their friends and least willing to speak with strangers (*M*= 4.27, *SD*= 0.77 compared to *M*= 4.07, *SD*= 0.85 and *M*= 4.06, *SD*= 0.84, see Table 4). However, the repeated measures ANOVA with a Greenhouse-Geisser correction result indicated no significant difference in WTC with different receivers outside the classroom, $F(3.13, 140.74) = 2.11$, p> 0.05. This shows that although the factor of interlocutors played an essential role in WTC inside the classroom, its effect was not significant outside the classroom.

Whether the interlocutor was a native speaker or a non-native speaker did not matter. A paired t-test was conducted to compare the WTC scores for native and non-native speakers. The results indicated that there was no significant dif-

Table 3

Willingness to Communicate Outside the Classroom

			Percentage of responses				
Item	**Mean**	**SD**	**Strongly agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
1	4.10	0.86	37.5	39.6	18.8	4.2	0
2	3.69	0.81	14.6	45.8	33.3	6.3	0
3	4.17	0.78	39.6	37.5	22.9	0	0
4	3.98	0.98	39.6	25.0	29.2	6.3	0
5	4.42	0.68	52.1	37.5	10.4	0	0
6	3.98	0.84	29.2	43.8	22.9	4.2	0
7	3.75	0.86	22.9	33.3	39.6	4.2	0
8	4.25	0.70	39.6	45.8	14.6	0	0
9	3.94	1.02	37.5	29.2	22.9	10.4	0
10	4.38	0.61	43.8	50.0	6.3	0	0
11	4.27	0.77	45.8	35.4	18.8	0	0
12	4.08	0.85	35.4	41.7	18.8	4.2	0
13	4.23	0.91	45.8	37.5	12.5	2.1	2.1
Total	**4.10**						

Table 4

WTC Scores Outside the Classroom with the Subscales

Item	**Receivers**	**Mean**	**SD**
1-5	Native speakers (strangers)	4.07	0.85
6-10	Non-native speakers (strangers)	4.06	0.84
11	Friends	4.27*	0.77
12	Teachers	4.08*	0.85
13	Self	4.23*	0.91
Item	**Stimulation**		
1, 2, 4, 6, 7	Without stimulation	4.00	0.89
3, 5, 8, 10	With stimulation	4.29	0.70

Note. **There was only one item in this subscale*

ference in the WTC scores between speaking with native (*M*=4.07, *SD*=0.85) and non-native speakers (*M*=4.06, *SD*=0.84), $t(45)=0.28$, $p > 0.05$. This shows that whether the interlocutor was a native speaker did not significantly affect their WTC. Regarding the speaking environment outside the classroom, the mean score of WTC in a stimulating environment (*M*= 4.29, *SD*= 0.70) was higher than WTC without any stimulus (*M*= 4.00, *SD*= 0.89), see Table 6. The paired t-test results indicated a significant difference in WTC outside the classroom with external stimulus (such as when a foreigner needs help) and without external stimulus (such as when only encountering a foreigner somewhere), $t(45)=-4.57$, $p < 0.001$.

Generally, the data indicated that the level of WTC outside the classroom was higher than WTC inside the classroom. Willingness to communicate in the classroom significantly varied according to interlocutors and speaking topics, but not whether speaking scripts were allowed, while the identities of interlocutors did not make a significant difference in WTC outside the classroom. Instead, a stimulating environment seemed to be an important influence on WTC outside the classroom. The following section will report the data from the reflective reports and interviews according to the emerging themes to answer the second research question and provide more insights into the questionnaire data.

RQ 2: What are the rationales behind those perceptions?

This section will report the data from the reflective reports and the interviews about the participants' perceptions of WTC inside and outside the classroom and the subscales in the questionnaire, i.e., interlocutors, speaking topics, task types, and a stimulating environment. The themes that emerged from the qualitative data are 'language anxiety', 'relevant topics', and 'stimulating environment'.

Foreign Language Anxiety

As discussed, the questionnaire data showed that the participants were more willing to speak outside the classroom than inside the classroom. When asked why they were more willing to talk outside the classroom, anxiety in the classroom emerged in their answers. The relationship between WTC and anxiety has also been reported in Jackson (2002) and Liu and Jackson (2008). According to the qualitative data of this present study, the anxiety was implied as a result of interlocutors' communicative competence, relationships with interlocutors and language classroom experience caused by teachers.

If the participants viewed their conversation partners as lacking in competence, they exhibited a reduced willingness to speak because they perceived conversations with such individuals as likely to cause anxiety and less beneficial for the enhancement of their own speaking proficiency. In the classroom context, their interlocutors were mostly their classmates, whom they perceived as not competent enough for their speaking development. Outside the classroom, however, they believed the probability of encountering individuals of greater competence, such as foreigners, was higher. They said:

> *"When talking with a friend in the classroom, they didn't know if my grammar was correct. I was afraid I'd remember a wrong model sentence from my friend. But outside the classroom, I can seek a chance to talk to foreigners who can correct me." (Interviewee 6)*

> *"In the classroom, the interlocutor is usually the teacher or my classmate. They are non-native speakers. I'd prefer to talk with native speakers to get a decent accent in a learning setting like the classroom." (Interviewee 3)*

> *"My friends didn't understand what I was saying, so they didn't say anything back." (Student 16, Report 3)*

In the classroom setting, the participants preferred to converse with native speakers due to their perceived higher level of competence. However, this preference did not extend to outside of the classroom. When asked about the potential impact on their WTC outside of the classroom, Interviewee 3 and Interviewee 5 declined. They explained that within the classroom environment, they were more focused on speaking accurately and thus preferred interacting with a competent interlocutor. However, they expressed that outside of the classroom, any opportunity to engage in English conversation with others would be greatly appreciated, no matter what their first language was.

Language anxiety also played a crucial role when the participants discussed a preference to talk to friends over strangers. This is because they considered the degree of intimacy with the interlocutor important. As student 26 revealed in Report 3, "I'm shy when speaking English with others who are not my friends." Student 36 elaborated on this point in Report 1, saying, "When I speak to my friends, it makes me relaxed and want to speak more." In the interview data, Interviewee 6 also pointed out, "I'd be more confident when I talked to my close friends."

Apart from interlocutors' competence and familiarity with interlocutors, negative classroom experience was also discussed in relation to classroom anxiety. As the interviewees said:

> *"In the classroom, I'm a bit anxious, but outside...it was like a chance for trial and error. In the classroom, I'm afraid I'd make mistakes.*
>
> *When I was in high school, the teacher criticised me for not speaking correctly.*
>
> *So, I don't dare to speak (in the classroom)." (Interviewee 1)*

> *"Sometimes I felt embarrassed if I made mistakes. My English teachers used to scold my classmates when they made mistakes and I witnessed it with fear." (Interviewee 2)*

The data above emphasises the influence of classmate competence and teacher competence on WTC in the classroom. Also, it pointed out that negative past experiences in the classroom and teacher-centred classroom, where the teacher primarily acts as an authority rather than a facilitator (Karnchanachari, 2019) can negatively affect WTC. Fear of negative evaluation, as found in this present study, is considered part of language learning anxiety (Horwitz et al., 1986). These components are classroom dynamics that teachers need to pay attention to (Mai & Fan, 2021).

Relevant Topics

In line with the questionnaire results indicating a significant effect of speaking topics on WTC inside the classroom, the reflective report results of the open-ended question "What have you talked about in the past few weeks?" showed that daily routines and study were the topics the participants talked about the most during the 16 weeks (31.6% and 30.6%, respectively). The third most popular topic was hobbies (30.3%). Interestingly, the fourth most discussed topic was not about personal matters but about what was happening in the country and the world (12.1%), such as the spread of COVID-19 and economic regression, which required a higher vocabulary level and could be considered more complex. Table 5 shows the list of subjects the participants chose to talk about.

Table 5

Speaking Topics Reported in the Self-Reflective Reports

Speaking topics	N	Percentage
Daily routines	72	31.6%
Study	71	30.6%
Hobbies	47	20.3%
News	28	12.1%
Love	4	1.7%
Weather	4	1.7%
Politics	3	1.2%
Dream jobs	2	0.8%

The reasons why they preferred to talk about simple topics like routines and studies were reflected in the reflective reports. They revealed that they would be more willing to communicate if the topics were relevant to their language proficiency and their content knowledge. They revealed:

> *"I'd rather talk about hobbies, movies, or things that don't require difficult vocab and grammar." (Student 33, Report 2)*

> *"When I don't know about that topic, I don't know what to say, and it'd be the end of the conversation." (Interviewee 2)*

In addition to assessing their proficiency with the topic, the participants also considered the potential benefits they could gain from engaging in conversations. Some of them were interested in a challenging topic, as they believed it would be beneficial for their language development. This is in line with the reflective report data which indicates a more complicated matter such as news and economics was ranked fourth among the most discussed topics. Interviewee 5 pointed out, "If that topic is useful for my development (of English), I'd love to talk." Also, Interviewee 6, who rated WTC higher on a situation relating to a more complicated topic like being a tour guide for a foreigner, stated "I'd be more willing to speak with a foreigner when offering help to be his tour guide for free because I think it's challenging." This could be interpreted as WTC according to the topic of interest since interest consists of both intrinsic emotions and value-related factors (Schiefele, 1991), which can enhance intrinsic motivation (Deci & Ryan, 1985).

Previous research also highlights the impact of interest in the speaking topic on WTC (Peng, 2014; Riasati, 2012; Weda et al., 2021; Yashima et al., 2016). Unlike the previous research, the present study discovered that the topic also affected L2 WTC through a sense of achievement, as Interviewee 6 referred to it as a challenge. Although the participants were of low proficiency, they did not always want an easy task. Some of them might evaluate the value of the conversation based on a sense of improvement they could gain. This can be linked to the reasons behind their preference for interlocutors of higher proficiency.

Stimulating Environment

As discussed earlier, the questionnaire data indicated that WTC in a stimulating environment was higher than WTC without a stimulus. The reflective report and interview data explained the additional reasons behind this: no English-speaking environment and no pressure to speak. A lack of an English-speaking environment was not considered only from the perspective of the EFL environment, but also collaborative learning with their peers. They reflected on the reports about obstacles to speaking English:

> *"People around me don't speak English. I don't know who to talk with." (Student 32, Report 1)*

> *"My friends speak Thai. There are only a few foreigners around here." (Student 9, Report 2)*

> *"If I have a chance to speak (English), I will. But when I spoke English with my friends, they didn't cooperate." (Student 23, Report 3)*

> *"With no forced circumstances, I won't dare to talk with a foreigner. I need a push. When my friends spoke English, I'd jump into the conversation." (Interviewee 2)*

As can be seen here, it is about orientations for friendship that could drive the participants to speak outside the classroom and how the people around them, especially their friends, could support them by creating an English-speaking environment through collaborative learning. Although 'environment' has been widely discussed in previous research, most of them were related to the environment in the classroom (e.g., Mai & Fan, 2021; Peng, 2019). This present study showed that collaborative English speaking with friends was expected by learners both inside and outside the classroom, and was believed to enhance WTC. It is a form of social support (MacIntyre et al.,1998). The findings from the interview data and the reflective reports of this present study revealed insights into what social support EFL learners might expect outside the classroom.

Pedagogical Implications

This study indicated that the students would be more willing to communicate outside the classroom than inside the classroom, possibly due to foreign language anxiety relating to interlocutors and a lack of teacher support. They perceived the classroom as a setting with limited opportunities to talk with a competent speaker, and then they might end up talking with classmates of lower proficiency who could not help correct their speaking mistakes. This could lower their motivation to speak and WTC. Since talking to a more competent learner could increase WTC, teachers might consider mixing higher-level students into lower-level groups when doing classroom activities. However, to increase the frequency of L2 use outside the classroom, teachers may want to try to lessen their anxiety by first assigning them to talk to their friends in L2. This could enhance the social support that FL students expect outside the classroom. Then, once they feel more comfortable communicating, teachers

may assign them to speak to foreigners or someone they are less familiar with. This could be useful for simulating speaking in a broader range of situations where they cannot speak only with people they are close to.

To create a positive classroom environment, teachers may pay more attention to careful feedback on students' spoken language. This is because teacher feedback could deteriorate their willingness to speak, and different types of corrective feedback could affect students' WTC (Tavakoli& Zarrinabadi, 2018). With limited exposure to English in the EFL context, alternative ways of raising student awareness of their incorrect output might be considered. To increase the amount and quality of students' spoken output and reduce foreign language anxiety both inside and outside the classroom, teachers may integrate mobile-assisted language learning (MALL), especially automatic speech recognition (ASR), into classroom activities (Ahn & Lee, 2016). Previous studies, such as Jung (2011) and Evers and Chen (2022), have found that ASR can be useful for the fluency development of EFL learners through individualised feedback on pronunciation and the integration of peer feedback.

Another finding of this present study was that speaking topics could play a pivotal role in students' WTC in relevance to their foreign language proficiency. To assign a speaking task appropriate for their proficiency yet still challenging enough for them to pursue the task, teachers might try incorporating scaffolding into challenging tasks. Scaffolding, which is assistance given by the teacher or more competent peer to guide learners to advance their skills within their level of capacity can be applied to the speaking classroom by guiding them with pre-task planning and comprehensible input (Krashen, 1985; Skehan, 1998; Wood et al.,1976). Also, scaffolding can be adopted as part of peer feedback to accommodate the need for social support found in this present study. Since familiarity with the interlocutor and stimulating environment could affect WTC, teachers may consider allowing them to select the speaking partner(s) themselves when giving students a speaking activity. When assigned to discuss a challenging topic with someone who makes them feel comfortable and supported by both the teacher and peers, it is more likely that they will be motivated to speak and, as a result, more willing to do so.

Limitations of the Study and Future Research

The study was limited to 46 first-year university students in Thailand. Therefore, it was not possible to generalise the findings to other university students in the same context or, especially, in different contexts. Despite this limitation, it is expected that the insights into WTC generated by this study might shed some light on WTC in other EFL contexts. Another limitation was that the study focused only on interlocutors, speaking topics, task types, and a stimulating environment. It did not cover various situational factors. For a more in-depth understanding of learners' perceptions of WTC, future research might explore other situational variables and add observational data to explore if learners' perceptions and actions match.

CONCLUSION

This study has shown that the students had WTC outside the classroom more than inside the classroom. This phenomenon could potentially be ascribed to foreign language anxiety stemming from perceptions of inadequately skilled conversation partners, the selection of topics unsuitable for their language proficiency, and negative assessments of the classroom environment overseen by the teacher. Interlocutors made a significant contribution to WTC, both inside and outside the classroom. They tended to consider the interlocutor's communicative competence rather than their race and nationality. Outside the classroom, a drive from their friends to create an English-speaking environment seems to be more prominent when compared to the classroom context. This could be because of a lack of English speakers in the EFL contexts. Although previous studies have revealed several possible factors affecting WTC, most of them focused on WTC inside the classroom. Among the limited studies of WTC outside the classroom, qualitative insights into learners' perceptions relating to WTC are rare. As expected, based on prior research, this present study found that the classroom can be a place that causes foreign language anxiety. With support from the teacher and peer, the students may successfully cope with the causes of reticence both in class and in real life, the frequency of L2 use could be increased, and the development of speaking skills would be promoted.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*(4), 778-786. https://doi.org/10.1111/bjet.12354

Aomr, J. A., Seng, G. H., & Kapol, N. (2020). Relationship between willingness to communicate in English and classroom environment among Libyan EFL Learners. *Universal Journal of Educational Research*, *8*(2), 605–610. https://doi.org/10.13189/ujer.2020.080232

Baghaei, P. (2013). Development and psychometric evaluation of a multidimensional scale of willingness to communicate in a foreign language. *European Journal of Psychology of Education*, *28*(3), 1087–1103. https://doi.org/10.1007/s10212-012-0157-y

Baker, S. C., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and Second language orientations. *Language Learning*, *50*(2), 311-341. https://doi.org/10.1111/0023-8333.00119

Barjesteh, H., Vaseghi, R., & Neissi, S. (2012). Iranian EFL Learners' willingness to communicate across different context- and receiver-types. *International Journal of English Linguistics*, *2*(1). https://doi.org/10.5539/ijel.v2n1p47

Barrios, E., & Acosta-Manzano, I. (2021). Factors predicting classroom WTC in English and French as foreign languages among adult learners in Spain. *Language Teaching Research*, *0*(0). https://doi.org/10.1177/13621688211054046

Basoz, T., & Erten, I. H. (2018). Investigating tertiary level EFL learners' willingness to communicate in English. *English Language Teaching*, *11*(3), 78. https://doi.org/10.5539/elt.v11n3p78

Birkner, V. (2021). Determination of revisiting input and output hypotheses in second language learning. *Modern Perspectives in Language, Literature and Education, 9*, 22–26. https://doi.org/10.9734/bpi/mplle/v9/4158f

Cao, Y., & Philp, J. (2006). Interactional context and willingness to communicate: A comparison of behavior in whole class, group and dyadic interaction. *System, 34*(4), 480-493. https://doi.org/10.1016/j.system.2006.05.002

Chang, F. (2011). The causes of learners' reticence and passivity in English classrooms in Taiwan. *The Journal of Asia TEFL*, *8*(1), 1-22.

Chung, J. H. (2021). "We Participate, Silently": Explicating Thai university students' perceptions of their classroom participation and communication. *Qualitative Research in Education*, *10*(1), 62–87. https://doi.org/10.17583/qre.2021.7159

Clark, R. (2017). Convenience Sample. In G. Ritzer (Ed), *The Blackwell Encyclopedia of Sociology* (pp. 1–2). John Wiley & Sons.

Clément, R., Baker, S.C., & MacIntyre, P.D. (2003). Willingness to communicate in a second language: The effects of context, norms, and vitality. *Journal of Language and Social Psychology*, *22*(2), 190-209. https://doi.org/10.1177/0261927x03022002003

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education*. Routledge.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.

Croasmun, J. T., & Ostrom, L. (2011). Using Likert-type scales in the social sciences. *Journal of Adult Education*, *40*(1), 19-22.

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Denies, K., Yashima, T., & Janssen, R. (2015). Classroom versus societal willingness to communicate: Investigating French as a second language in Flanders. *The Modern Language Journal*, *99*(4), 718–739. https://doi.org/10.1111/modl.12276

Dewaele, J.M., & Dewaele, L. (2018) Learner-internal and learner-external predictors of willingness to communicate in the FL classroom. *Journal of the European Second Language Association*, *2*(1), 24-37. https://doi.org/10.22599/jesla.37

Donald, S. (2010). Learning how to speak: Reticence in the ESL classroom. *The Annual Review of Education, Communication, and Language Sciences*, *7*, 41-58.

Doqaruni, V. R. (2015). Increasing Confidence to Decrease Reticence: A Qualitative Action Research in Second Language Education. *The Canadian Journal of Action Research*, *16*(3), 42-60. https://doi.org/10.33524/cjar.v16i3.227

Dorman, J. P., Fisher, D. L., & Waldrip, B. G. (2006). Learning environments, attitudes, efficacy and perceptions of assessment: A LISREL analysis. In D. L. Fisher & M. S. Khine (Eds.), *Contemporary approaches to research on learning environments* (pp. 1-28). World Scientific.

Evers, K., & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, *35*(8), 1869-1889. https://doi.org/10.1080/09588221.2020.1839504

Harumi, S. (2020). Approaches to interacting with classroom silence: the role of teacher talk. In J. King and S. Harumi (Eds.), *East Asian Perspectives on Silence in English Language Education* (pp. 37–59). Multilingual Matters.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*, 125–132.

Jackson, J. (2002). Reticence in second language case discussion: Anxiety and aspirations. *System*, *30*, 65-84. https://doi.org/10.1016/s0346-251x(01)00051-3

Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). II. More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development*, *82*(2), 13-30. https://doi.org/10.1111/mono.12296

Jung, H. K. (2011). The correction of learner's English pronunciation errors through speech recognition reading program. *Multimedia-Assisted Language Learning*, *14*(3), 291–314.

Kang, S. J. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System*, *33*(2), 277-292. https://doi.org/10.1016/j.system.2004.10.004

Karnchanachari, S. (2019). An investigation into learners' willingness to communicate in English in the classroom: A study of Thai EFL students in the Thai and international programs. *rEFLections*, *26*(2), 84-106. https://doi.org/10.61508/refl.v26i2.241757

Krashen, S. D. (1985). *The input hypothesis: Issues and implications.* Addison-Wesley Longman Limited.

Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly, 35*(4), 537-560. https://doi.org/10.2307/3588427

Lee, J. S., & Hsieh, J. S. (2019). Affective variables and willingness to communicate of EFL learners in in-class, out-of-class, and digital contexts. *System*, *82*, 63-73. https://doi.org/10.1016/j.system.2019.03.002

Li, H., & Liu, Y. (2011). A brief study of reticence in ESL class. *Theory and Practice in Language Studies*, *1*(8), 961-965. https://doi.org/10.4304/tpls.1.8.961-965

Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, *92*, 71-86. https://doi.org/10.1111/j.1540-4781.2008.00687.x

MacIntyre, P. D., Baker, S. C., Clément, R., & Conrod, S. (2001). Willingness to communicate, social support, and language-learning orientations of immersion students. *Studies in Second Language Acquisition*, *23*(3), 369–388. https://doi.org/10.1017/s0272263101003035

MacIntyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal*, *82*(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

MacIntyre, P. D., & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology*, *15*, 3-26. http://dx.doi.org/10.1177/0261927X960151001

MacIntyre, P.D. (2007). Willingness to Communicate in the second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal*, *91*, 564-576. https://doi.org/10.1111/j.1540-4781.2007.00623.x

Mai, X., & Fan, Y. (2021). An empirical study of the willingness to communicate in college English classes from an ecological perspective. *Creative Education*, *12*(9), 2056–2065. https://doi.org/10.4236/ce.2021.129157

McCroskey, J. C. (1992). Reliability and validity of the willingness to communicate scale. *Communication Quarterly*, *40*(1), 16-25. https://doi.org/10.1080/01463379209369817

Mills, A. J., Durepos, G. & Wiebe, E. (2010). *Encyclopedia of case study research*. Sage Publications.

Nemoto, T., & Beglar, D. (2014). Developing Likert-scale questionnaires. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings* (pp. 1-8). JALT.

Pattapong, K. (2015). Complex Interactions of Factors Underlying Thai EFL Learners' willingness to communicate in English. *PASAA: Journal of Language Teaching and Learning in Thailand*, *49*, 105-136.

Pattapong, K. (2010). *Willingness to communicate in a second language: A qualitative study of issues affecting Thai EFL learners from students' and teachers' point of view* [Doctoral dissertation, University of Sydney]. Sydney Digital Thesis. https://ses.library.usyd.edu.au/handle/2123/9244

Pawlak, M., Mystkowska-Wiertelak, A., & Bielak, J. (2016). Investigating the nature of classroom willingness to communicate (WTC): A micro-perspective. *Language Teaching Research*, *20*(5), 654–671. https://doi.org/10.1177/1362168815609615

Peng, J., & Woodrow, L. J. (2010). Willingness to communicate in English: A model in Chinese EFL classroom context. *Language Learning*, *60*(4), 834 876. https://doi.org/10.1111/j.1467-9922.2010.00576.x

Peng, J. (2019). The roles of multimodal pedagogic effects and classroom environment in willingness to communicate in English, *System*, *82*, 161-173. https://doi.org/10.1016/j.system.2019.04.006

Peng, J. E. (2014). *Willingness to communicate in the Chinese EFL university classroom: An ecological perspective* (vol. 76). Multilingual Matters. https://doi.org/10.21832/9781783091560

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree–disagree scales. *Sociological Methods & Research*, *43*(1), 73-97. https://doi.org/10.1177/0049124113509605

Riasati, M. J. (2012). EFL learners' perception of factors influencing willingness to speak English in language classrooms: A qualitative study. *World Applied Sciences Journal*, *17*(10), 1287-1297.

Riasati, M. J. (2014). Causes of reticence: Engendering willingness to speak in language classrooms. *International Journal of Research Studies in Language Learning*, *3*(1), 115-122. https://doi.org/10.5861/ijrsll.2013.410

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, *26*(3-4), 299-323.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Sudina, E. (2023). Scale quality in second-language anxiety and WTC: A methodological synthesis. *Studies in Second Language Acquisition*, 1-29. https://doi:10.1017/S0272263122000560

Talandis, G., & Stout, M. (2015). Getting EFL students to speak: An action research approach. *ELT Journal*, *69*(1), 11-25. https://doi.org/10.1093/elt/ccu037

Tanaka, K. (2007). Japanese students' contact with English outside the classroom during study abroad. *New Zealand Studies in Applied Linguistics*, *13*(1), 36-54.

Tantiwich, K., & Sinwongsuwat, K. (2021). Thai university students' problems of language use in English conversation. *LEARN Journal: Language Education and Acquisition Research Network*, *14*(2), 598-626.

Tavakoli, M., & Zarrinabadi, N. (2018). Differential effects of explicit and implicit corrective feedback on EFL learners' willingness to communicate. *Innovation in Language Learning and Teaching*, *12*(3), 247-259. https://doi.org/10.1080/17501229.2016.1195391

Todd, R. W. (2011). Analyzing and interpreting rating scale data from questionnaires. *rEFLections*, *14*, 69-77. https://doi.org/10.61508/refl.v14i0.114230

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, *36*(2), 299–316. https://doi.org/10.1080/01411920902919257

Weda, S., Atmowardoyo, H., Rahman, F., Said, M. M., & Sakti, A. E. F. (2021). Factors affecting students' willingness to communicate in EFL classroom at higher institution in Indonesia. *International Journal of Instruction*, *14(*2), 719-734. https://doi.org/10.29333/iji.2021.14240a

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Child Psychiatry*, *17*, 89–100.

Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal*, *86*(1), 54-66.

Yashima, T., MacIntyre, P. D., & Ikeda, M. (2016). Situated willingness to communicate in an L2: Interplay of individual characteristics and context. *Language Teaching Research*, *22*(1), 115–137. https://doi.org/10.1177/1362168816657851

Zarrinabadi, N. (2014). Communicating in a second language: Investigating the effect of teacher on learners' willingness to communicate. *System*, *42*, 288-295. https://doi.org/10.1016/j.system.2013.12.014

Zhang, X., & Head, K. (2010). Dealing with learner reticence in the speaking class. *ELT Journal*, *64*(1), 1-9. https://doi.org/10.1093/elt/ccp018

APPENDIX A

Table A

Questionnaire of Willingness to Communicate in the Classroom

Item	WTC situations	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
1	I am willing to talk to my close friends in English about a personal topic (such as their well-being, their family, etc.).					
2	I am willing to ask my close friends in English the meaning of an English word.					
3	I am willing to ask my close friends in English how to say an English phrase to express the thoughts in my mind.					
4	I am willing to ask my teachers in English how to pronounce a word in English.					
5	I am willing to ask my teachers in English when I don't understand the lesson they teach.					
6	I am willing to express my opinions or answer questions in English to the class.					
7	I am willing to do a role-play standing in front of the class in English (e.g., ordering food in a restaurant) without notes.					
8	I am willing to do a role-play standing in front of the class in English (e.g., ordering food in a restaurant) with notes.					
9	I am willing to do a role-play in English without notes at my desk, with my close friends (e.g., ordering food in a restaurant).					
10	I am willing to do a role-play in English with notes at my desk, with my close friends (e.g., ordering food in a restaurant).					

Table B

Questionnaire of Willingness to Communicate outside the Classroom

Item	WTC situations	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
1	If I encountered some native speakers of English in the street, restaurant, hotel etc., I hope an opportunity would arise and they would talk to me.					
2	If I encountered some native speakers of English in the street, restaurant, hotel etc., I would find an excuse to talk to them.					
3	If I encountered some native speakers of English who were facing problems in Thailand because of not knowing the Thai language, I would take advantage of this opportunity and would talk to them.					
4	I am willing to accompany some native speakers of English and be their tour guide for a day free of charge.					
5	If someone introduced me to a native-speaker of English, I would like to try my abilities in communicating with him/her in English.					
6	If I encountered some non-native speakers of English in the street, restaurant, hotel etc., I hope an opportunity would arise and they would talk to me.					

Item	WTC situations	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
7	If I encountered some non-native speakers of English in the street, restaurant, hotel etc., I would find an excuse to talk to them.					
8	If I encountered some non-native speakers of English who were facing problems in Thailand because of not knowing the Thai language, I would take advantage of this opportunity and would talk to them.					
9	I am willing to accompany some non-native speakers of English and be their tour guide for a day free of charge.					
10.	If someone introduced me to a non-nativespeaker of English, I would like to try myabilities in communicating with him/her in English.					
11.	In order to practice my English, I am willing to talk in English with my friends outside the classroom.					
12.	In order to practice my English, I am willing to talk in English with my teachers outside the classroom.					
13.	In order to practice my English, I am willing to talk in English with myself outside the classroom.					

https://doi.org/10.17323/jle.2023.16604

Topic Modeling for Text Structure Assessment: The case of Russian Academic Texts

Valery Solovyev [1], **Marina Solnyshkina** [1], **Elena Tutubalina** [2]

[1] Kazan Federal University, Kazan, Russia

[2] Ivannikov Institute for System Programming of the RAS, Moscow, Russia

ABSTRACT

Background: Automatic assessment of text complexity levels is viewed as an important task, primarily in education. The existing methods of computing text complexity employ simple surface text properties neglecting complexity of text content and structure. The current paradigm of complexity studies can no longer keep up with the challenges of automatic evaluation of text structure.

Purpose: The aim of the paper is twofold: (1) it introduces a new notion, i.e. complexity of a text topical structure which we define as a quantifiable measure and combination of four parameters, i.e. number of topics, topic coherence, topic distribution, and topic weight. We hypothesize that these parameters are dependent variables of text complexity and aligned with the grade level; (2) the paper is also aimed at justifying applicability of the recently developed methods of topic modeling to measuring complexity of a text topical structure.

Method: To test this hypothesis, we use Russian Academic Corpus comprising school textbooks, texts of Russian as a foreign language and fiction texts recommended for reading in different grades, and employ it in three versions: (i) Full Texts Corpus, (ii) Corpus of Segments, (iii) Corpus of Paragraphs. The software tools we implement include LDA (Latent Dirichlet Allocation), OnlineLDA and Additive Regularization Of Topic Models with Word2vec-based metric and Normalized Pairwise Mutual Information.

Results: Our findings include the following: the optimal number of topics in educational texts varies around 20; topic coherence and topic distribution are identified to be functions of grade level complexity; text complexity is suggested to be estimated with structural organization parameters and viewed as a new algorithm complementing the classical approach of text complexity assessment based on linguistic features.

Conclusion: The results reported and discussed in the article strongly suggest that the theoretical framework and the analytic algorithms used in the study might be fruitfully applied in education and provide a basis for assessing complexity of academic texts.

KEYWORDS

text structure, topic modeling, school textbooks, text complexity, Russian language

Citation: Solovyev V., Solnyshkina M., & Tutubalina E. (2023). Topic Modeling for Text Structure Assessment: The case of Russian Academic Texts. *Journal of Language and Education, 9*(3), 143-158. https://doi.org/10.17323/jle.2023.16604

Correspondence:
Valery Solovyev,
maki.solovyev@mail.ru

Received: January 04, 2023
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

Approaches to Determining Text Complexity

Numerous attempts have been made to explain what text complexity is and recently it has become a feature of importance (Si and Callan 2001; McNamara et al., 2014; Gatiyatullina et al., 2023). The reason for this is obvious: a text is supposed to correspond to the proficiency of the target audience in all possible areas including education, publishing, legislation, science, medicine, etc. In the modern world where educators are committed to providing high quality personalized teaching and distance education, validated assessment of text complexity has become of particular importance for textbooks writers.

A broad definition of text complexity as "a measure of how easy or difficult a text is" (Bailin and Grafstein, 2016) though

universally accepted but of no practical relevance as it provides no algorithm of its assessment. Similarly, analysis and assessment of the topical structure complexity seems potentially useful for identifying structural patterns which could distinguish between texts of different grade levels. Text topical structure as "a way of indicating the relationship between the progression of sentence topics and the topical depth which indicates the semantic hierarchy» (Chuang 1993, p.2) is especially in demand in education. The reason for this is that features of topical structure reveal qualitative differences among texts.

As a research problem, text complexity has been studied for over a century and, as a notion, it has given rise to numerous definitions (for complexity of Russian texts see Ivanov et al., 2018; Oborneva, 2006; Solnyshkina et al., 2014; Solovyev et al., 2019). However, all its descriptive definitions, as the one above, are of little practical use when it comes to measure its level. Many definitions of text complexity tend to lack operationality since they do not provide order of the procedures to be applied to determine how comprehensible a text is.

Similarly with the text-reader multi-criteria approach: although psycholinguistics seeks to develop sophisticated methods of matching texts with specific categories of readers (Crossley et al., 2014), conducting experiments of the kind is not only time- and effort-consuming, applicability of their results is always questionable due to differences in cognitive and linguistic abilities of the respondents (readers) involved. Therefore, automation of text complexity assessment based on computing text metrics, not involving readers, is viewed by the authors as a research niche.

However, feasibility of the task to a great extent depends on (1) availability of a representative corpus annotated by experts for certain categories of readers and (2) adequacy of the applied methods.The history of the latter started with the first complexity formulas proposed in (Flesch, 1948) which relied on two simple variables of a text: the average word and sentence length. Over the years the formulas, known as readability formulas, gained popularity and, due to their simplicity and ease of calculation, have been ubiquitously used and even installed with Microsoft Word. However, a natural question asked by a number of researchers is whether these formulas fully reflect all aspects of text complexity or only some of them. An exemplary experiment of contrasting two versions of the same text presented in (Thorndyke, 1977) illustrates limitations of Flesch-Kincaid readability formulas: complexity levels estimated with readability formulas of a text and its jumbled version are the same, though it is obvious that the jumbled text is much more difficult to read and comprehend.

Back in the 1970s – 1990s, a number of formulas were developed and validated. They had more than those in Flesch-Kincaid variables and were expected to provide a much higher accuracy. Unfortunately, they also proved limited and were criticized in (Crossley et al., 2008; McNamara et al., 1996) for their ineffectiveness to take into account other parameters such as text informational capacity and discourse structure. While discriminating complexity levels in Wikipedia texts, standard readability formulas do not offer accuracy achieved with machine learning methods[1] (Eremeev & Vorontsov, 2020; Martinc et al., 2021).

In the 2000s – 2010s, hundreds of features were introduced and validated as text complexity predictors. E.g., Coh-Metrix (cohmetrix.com) computes 108 parameters, ReaderBench (Toma et al., 2021) calculates over 200 parameters, and TAALES (Kyle et al., 2018) calculates over 400 parameters. The majority of the parameters are interdependent thus forming the so-called clusters. For example, the average word length measured in letters/syllables is co-dependent with the average number of long words, i.e. words of two or more syllables. In fact, all readability formulas estimate the same, i.e. formal text complexity, and duplicate measurements using slightly different parameters. In (McNamara et al., 2014), all the parameters they estimate are streamed into five groups: narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion. The streaming is based on substantially different aspects of text complexity. However, despite a significant increase in the number and groups of parameters measured, there is always doubt if the list is exhaustive. Apparently, text complexity is a multifaceted phenomenon that encompasses a range of different aspects which is difficult to exhaust.

The study dataset comprises textbooks used in Russian schools, the quality, sophistication and complexity of which have widely been discussed over decades (see Solnyshkina et al., 2020 for a review). Recent reductions in the Federal List of Textbooks recommended for use in mainstream schools[2] did not make the situation less challenging: although educators in Russia have fewer choices, they still need reliable tools either to assist in choosing or modifying available texts for the target audience. Moreover, the quality of school textbooks language is another important issue in Russian education. At the moment, the quality of textbook language is evaluated quite subjectively by experts, the existing expertise tests almost nothing but compliance with grammar, vocabulary and spelling norms. With the advent of online resources for the Russian language and modern natural language processing models, there is an opportunity to develop an algorithm to assist in addressing these challenges.

[1] Although, the authors clarify that the machine learning models are supposed to be trained on representative text samples and their complexity levels are to be graded by qualified experts (Eremeev & Vorontsov, 2020).

[2] Publication.pravo.gov.ru/document/0001202307280015

Complexity of Text Topical Structure

In this paper, we introduce a new notion, i.e. complexity of text topical structure, which, to the best of our knowledge, has not been previously recognized as a quantifiable measure estimated with a limited number of parameters. We refer to these parameters as 'predictors of topical structure complexity' which include the number of topics, topic coherence, topic distribution, and topic weight.

The types of topic development identified by researchers indicate that while conveying information, writers do not limit themselves to one topic but may divert to other topics, boundaries of which are sometimes hard to identify (Watson, 2016). Besides, as topics are developed differently in various text types and genres, e.g. instructional and expository texts may develop different types of topic progression following which may present additional difficulty for a reader (Ninio & Snow, 1999). Moreover, if topics are implicit or intertwined, the text is also hard to comprehend. The latter makes us conclude that structural organization of topics is directly related to text complexity.

Topic coherence is viewed in the article as semantic similarity of the words forming a topic (McNamara et al., 2014; McNamara et al., 1996; Balyan et al., 2018), whereas text coherence is referred to as "sense relations between single units (sentences or propositions) of a text" (www.glottopedia.org/index.php/Coherence) and manifests itself in repetitions or synonyms, as well as cohesion devices. As for comprehensive definition of 'complexity of a text topical structure', as a notion and an aspect of text complexity, it is, in our view, to exhibit and be based on precise mathematical formalism and specific features designating it. For the purpose indicated, we suggest implementing topic modeling apparatus developed by distinguished scholars in recent decades (Boyd-Graber et al., 2017; Mulunda et al., 2018; Rehurek & Sojka, 2010).

As it was mentioned earlier, similar numerical experiments require collections of texts annotated for complexity. For this purpose, researchers tend to use either collections of school textbooks, foreign language tests or collect corpora for specific research goals. As a rule of thumb in discourse complexology, suitability and 'complexity' of the dataset is to be evaluated by experts (Solovyev et al., 2022).

Present Study

Following the tradition developed in the area, for the current study we compiled four subcorpora: (1) a subcorpus of textbooks on Social Science, (2) a subcorpus of textbooks on Biology used in secondary and high schools of the Russian Federation, (3) a subcorpus of literary fiction read by students in secondary and high schools of the Russian Federation; (4) a subcorpus of texts used in tests for learners of Russian as a foreign language (A2-C1, CEFR).

In comparison with our previous conference papers (Sakhovskiy et al., 2020a, 2020b), we significantly extended experimental exploration of text complexity and focused on text complexity correlation with semantic and statistical properties of its topics. More specifically, in this research we have conducted new experiments, moving from one collection of educational texts on Social Science to four collections, including collections of school textbooks on Biology, fiction texts, and texts of Russian as a foreign language. We also extended evaluation of topic models and investigated effectiveness of previously suggested topic-based complexity features using texts of different domains. Finally, we investigated the correlation between text complexity and topic variety measured by the entropy of a topic distribution.

For the current analysis, we employ Latent Dirichlet Allocation (LDA) (Blei et al., 2003), and additive regularization of topic models (ARTM) (Vorontsov & Potapenko, 2015). The revealed topics are evaluated based on a number of several standard quality measures. Importantly, we show correlation of the revealed topics with the textbooks grade levels using Spearman's rank correlation.

Our objectives in this paper include (1) introducing a new concept of topical structure complexity as a quantifiable measure assessed with a limited number of parameters, and (2) testing the hypothesis that topic modeling methods enable to estimate complexity of a text topical structure. More specifically, we study the following questions:

(1) What is the reference range of topics in a collection of thematically related academic texts? We hypothesize that in order to achieve its goals and ensure readability, a text is supposed to have a certain optimum number of topics. A text typically contains more than one topic but topics excess hampers comprehension.

(2) Is topic coherence related to text complexity? Does topic coherence increase or decrease across grades?

(3) What topic parameters are co-dependent with the grade level?

(4) What type of topic progression is characteristic of textbooks? How do types of topic progression change across grades? Can these types and changes be quantified?

LITERATURE REVIEW

It's worth mentioning that the majority of research in the field of computer linguistics is often published in conference proceedings to establish precedence swiftly.

Text structure is viewed as an essential aspect of its complexity and has been explored as a part of reading comprehension and writing processes (McNamara et al., 2019;

Williams, 2005). In the majority of investigations, researchers focus on one or more specific types of text structures. (Kendeou & Broek, 2007; Diakidoy et al., 2003) present experimentally validated influence of text structure in refutation and non-refutation texts types. (Williams, 2005) reports on differences of respondents comprehending compare vs contrast text structures. In (Roehling et al., 2017), the authors identify text structure as one of the aspects of its complexity. A number of works explore ways of improving "text coherence and text information structure" for students with learning disabilities (cf. Arfe and Mason, 2018).

The impact of text structure on understanding and storing information has been a focus of a number of psychological and discourse studies. E.g. intensive studies on a sentence level have been conducted to test Chomsky's theory. (McBride & Cutting, 2015) offer a thorough overview of classical works on the influence of text structure on understanding individual sentences or short texts.

One of the main theories accepted by the majority of researchers working in the area, is W. Kintsch's theory of macro-propositions emphasizing a hierarchical organization of texts and revealing the structure and mechanisms of text comprehension (Kintsch, 1998). In (Kintsch & Vipond, 1979), W. Kintsch's theory was applied to contrast complexity of speeches delivered by US presidential candidates, D.D. Eisenhower and A.E. Stevenson. The authors argue that D.D. Eisenhower's speeches, if estimated with standard readability formulas, are more complex than those of A.E. Stevenson's, but much simpler to comprehend which ultimately explains his victory in the election. The modern paradigm of text complexity studies also implements theory of macro-propositions to demonstrate that structural organization is manifested in features reflecting referential and deep cohesion of texts (McNamara et al., 2014; McNamara et al., 2010; Balyan et al., 2018). The general conclusion of these studies is that texts of higher coherence are easier to comprehend. However, these studies were conducted at sentence and paragraph levels, not the topical level of a corpus of texts, thus leaving a niche for our research.

Another aspect to emphasize is that highly cohesive texts are not necessarily coherent, i.e., cohesive ties per se do not constitute quality well-structured texts, but embody numerous deliberate repetitions, which may make texts boring and unattractive. Well-written coherent texts, on the opposite, are not repetitive, but can usually quite "uncontroversially be divided into successively smaller segments down to the level of the clause, yielding a hierarchical structure" (Hobbs, 1990, p. 111).

Higher thematic coherence manifested in semantic proximity of the most significant terms of topics is likely to contribute to better comprehension of a text. However, if the topics are too close, a reader faces an interference effect as retroactive and proactive inhibition (Loftus, 1983). The latter is proved to hamper text comprehension. All the above establishes a new area to research in the area and highlights the importance of topic level of text complexity as a focus of new studies. Within topic modeling as one of the most frequently implemented approaches aimed at defining the content of text collections with the helpof automated tools, researchers accumulated a number of methods (Boyd-Graber et al., 2017) with a potential to be applied in a wide range of spheres (Mulunda et al., 2018). In this study we offer unsupervised learning methods enabling automatic extraction of those text features that affect text topic complexity.

METHOD

Background

The modern research paradigm requires that parameters selected to measure complexity of a text topical structure are to be tested on a representative corpus able to provide reproducibility of results. As it was mentioned earlier, to achieve the stated objectives, i.e. to test the algorithm proposed and apply topic modeling to assess the nature of a text structure and its complexity, we use Corpus of school textbooks. We view school textbooks as a suitable type of texts for the research as they are sequenced from lower to higher grade levels based on their assigned complexity indices. K. Berendes and S. Vajjala test (2018) this assumption and refer to it as 'systematic complexification'. They also contrasted textbooks from different German publishers and validated the systematic complexification assumption with varying degrees of consistency (Berendes & Vajjala, 2018). Collections of school textbooks have also been widely used to teach and test various models assessing text complexity in different languages (cf. Al-Tamim et al., 2014; Chen et al., 2013; Chen & Daowadung, 2015; Si & Callan, 2001; Santucci et al. 2020; Gazzola et al., 2022). Specifically, the tradition in the area is that a typical size of training collections of books is approximately a million tokens. Researchers use a corpora limited either by the number of subjects or grades. E.g. in Chen et al. (2013), these are textbooks on Mandarin, Social Studies, and Life Science for Grade 6 only. Si & Callan (2001) use a collection of Mathematics textbooks. Tanaka-Ishii et al. (2010) qualify a corpus of school textbooks as desirable and emphasize numerous challenges in compiling one. Another possible source of texts with assigned complexity levels is a corpus of foreign language tests described by Laposhina et al. (2018).

In this paper, we report on the algorithm to identify, extract and match topics exemplified in 4 sets of text collections, i.e. textbooks on Social Studies, Biology used in the 5th - 11th grades of secondary and high schools of the Russian Federation, fiction texts selected to read by schoolchildren in secondary and high schools, and Russian texts used to teach Russian as a foreign language. It is noteworthy that

the research is conducted on a battery of textbooks written by one author or a collective of authors for all grades. This fact eliminates any influence of authors' style or pedagogical concepts implemented in textbooks of different grades and enables to focus on text complexity only. Contrasting textbooks of the allegedly same complexity level written by different authors provides the possibility to identify the impact of authors' style.

In our experiments, we fit each topic model on the whole text collection. We have D documents and each document is described by frequencies of W words from the vocabulary elicited from the texts. The documents we have are of three types: full textbooks, segments, or paragraphs. The whole collection is viewed as an W by D matrix and the goal of topic modeling is decomposition of a large matrix into smaller matrices: a document-topic matrix and the word-topic matrix. Thus, we do not operate on the level of separate documents but instead fit a topic model on the full collection in one step. As for our experiments on correlation analysis, we first elicit the topic distribution of full textbooks using the observed word frequencies and then conduct our experiments on these distributions. For details refer to (Sakhovskiy et al., 2020b) where the authors provide mathematical foundations of topic modeling calculations.

Topic Models Quality: Metrics of Assessment

To assess the quality of a topic model, we utilized the following metrics: word2vec-based metric (Nikolenko, 2016) and normalized pairwise mutual information (NPMI) (Bouma, 2009). For our experiments, we employed 300-dimensional Rusvectores (Kutuzov & Kuzmenko, 2016) skip-gram models trained on the Russian National Corpus (RNC) (ruscorpora.ru/) and the Taiga corpus (Shavrina &Shapovalova, 2017).

We use standard approaches based on the distributive hypothesis which implies that semantics of a word is identified based on its contexts. Contexts are set by the adjacent words frequency vectors, and the metrics are calculated as distances between vectors. The normalized pairwise mutual information metric between two words indicates how likely the words are to occur together in a corpus.

A list of topic words is viewed as coherent if these words frequently occur in the same documents. To quantify the degree of topic coherence, we utilized the NPMI. In this work, we calculated the frequencies using RNC. The larger the NPMI measure, the more often the words of the topic occur together in the texts, i.e. the topic is more coherent.

The NPMI and word2vec-based metric are known to correlate well with human estimates of topic interpretability (Nikolenko, 2016; Newman et al., 2010a; Newman et al., 2010b), but they characterize topics from slightly different points of view. In Topic modeling studies, interpretability refers to subjective evaluation of experts to what extent a topic corresponds to any specific topic.

The word2vec-based Q-metric characterizes topics by how semantically close their words are, regardless of their relative location in the texts. The smaller the Q-metric, the closer is the semantics of the words, i.e. the topic is more coherent. NPMI is a more complex measure. It takes into account 'words joint occurrence' and reflects both the semantic proximity of words and their syntactic properties. This measure is a function of two factors. On the one hand, the closer semantically are the words, i.e. the more coherent is the topic, and the higher is the NPMI. On the other hand, high NPMI values can also be due to the fact that topic keywords form stable combinations, thus reflecting not so much text coherence but its stereotyped style and occasionally the writer's desire to facilitate text perception. If the collection of texts is stylistically homogeneous, and designed for similar categories of readers, then the second factor is leveled. However, in our case, i.e. with the collection of textbooks for different grades, the second factor is not leveled. Findings in (Nikolenko, 2016) indicate that the word2vec-based Q-metric reflects topic coherence better than NPMI.

Besides we introduce a new parameter, i.e. topic weight, which is defined as the average frequency of topic words in an auxiliary corpus.

To increase the stability and robustness of topic evaluation (cf. Lau and Baldwin, 2016), we computed all metrics using 5, 10, 15, 20 top frequency words of each topic and took the mean over the four values as a topic estimate. The model quality is the average topic quality over all its topics.

Data and Preprocessing

When constructing topic models, we used the following corpora: (i) Corpus of Full Texts, (ii) Corpus of Segments, (iii) Corpus of Paragraphs. Corpora (ii) and (iii) were obtained by splitting full book corpus documents into smaller documents. The corpus of segments was compiled by the algorithm designed to get a maximum possible number texts so that the trained model was of excellent quality. The algorithm included numerous interdependent steps: sentences were sequentially added to the segment until the end of the book was reached, or until situations when adding a sentence resulted in an excessive size of the segment. And in that case, the sentence was added to a new 'empty' segment. The maximum segment size was set at 1000 tokens. We also removed punctuation, rare words (i.e. words registered in fewer than 3 documents), stop-words and auxiliary parts of speech. The Stop-word list was adopted from github.com/stopwords-iso/stopwords-ru. Ultimately the segments' length appeared to be shorter than 1000 words. At the final stage of preprocessing we also used UDPipe library for lemmatization and POS tagging. Table 1 below shows the corpora statistics received after preprocessing the dataset.

Table 1

Number of documents and document lengths after preprocessing the collection

Collections	Number of full texts	Average document length (tokens)		
		full texts	segments	paragraphs
Social Studies	16	29214	576	26
Biology	25	22523	607	25
Russian as a Foreign Language	199	133	127	20
Fiction	111	19809	390	28

The Table above shows that the average length of full texts in "Russian as a Foreign Language" subcorpus is small and does not differ much from the average length of a segment. Thus, splitting this collection into segments did not significantly increase the number of documents, and was not used in further work.

The Topic models we used included the LDA model from Mallet library (McCallum, 2002), OnlineLDA (Hoffman et al. 2010). and ARTM model from BigARTM library (Vorontsov, 2015). While constructing LDA models, we performed 1000 iterations in the training corpus. To tune hyperparameters of ARTM model, we introduced regularizers of selection and decorrelation of topics, regularizers of thinning distributions $(d|t)$ and $p(w|t)$, the effect of which was gradually increased. The models were trained with 280 training iterations for the corpus of full books and 175 iterations for smaller documents. 25 training iterations were performed in each document. We constructed models with the range of 5 to 100 topics and 1 step.

RESULTS

Quality Assessment of Topic Models

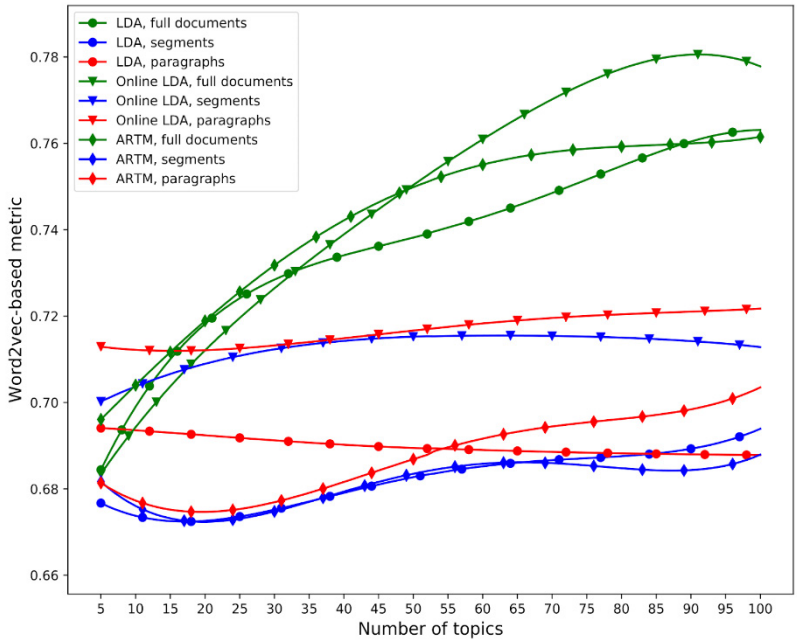
The quality of the constructed topic models was assessed with word2vec-based metrics and NPMI. While assessing the quality of each metric, we used 5, 10, 15, 20 most probable words of each topic.

The results of assessing the quality of topic models showed that LDA and ARTM models achieve the best quality of the selected metrics when trained on segments and paragraphs of the source texts. Thus, splitting the source texts into smaller sets of documents results in receiving more interpretable and distinguishable topics.

As an example, Fig. 1 presents quality assessments of the models trained on the collection of texts on Social Studies.

Figure 1

Quality Assessments of Topic Models Trained on Social Studies Texts: Word2vec Metric with RNC[3]



[3] In Fig. 1, metrics and quality of the model are inversely dependent: the smaller is the metric, the higher is the quality of the model.

The received estimates of the topic models were used to define the number of topics and the type of the document split for further experiments. The resulting graphs demonstrate the main findings: the best quality values are observed for the LDA and ARTM segment models, and the best quality value of these models is achieved with a number of topics close to 20. The diagram also shows calculation results for the OnlineLDA model, but as they turned out to be the worst of the three models considered, the OnlineLDA was excluded from further analysis.

For further experiments of correlation tests of texts on Social Studies, we used the segment model. For all other collections we implemented paragraph-trained models. The number of topics assigned are 40 for fiction texts and 20 for all other collections of texts.

Correlation Analysis of Topic Properties and Text Complexity

The topics obtained as a result of topic modeling were later used to definethe topic properties able to identify the type of relations between a text and a certain level of complexity. To this end we implemented distributions of topics in Corpus of Full Texts in the collections of Social Studies, Biology, Fiction, and Russian as a Foreign Language. Each document in this collection had an assigned complexity either by Common European Framework of Reference, i.e.A1-C2 (www.coe.int/en/web/common-european-framework-reference-languages), or a grade level. The relationship between complexity levels and topic distribution metrics were assessed with Spearman's correlation tests.

With the topic defined as 20 most probable words in a text, we conducted a number of experiments aimed at assessing the relationship between text complexity and the following text parameters: the average NPMI of topics and average distance between word vectors in the topic in the word-2vec space. The research was performed on corpus Taiga and two subcorpora of RNC, i.e. Subcorpus "Sovremennyye pis'mennyye teksty" of the Main Corpus and the Spoken Subcorpus. We applied two approaches to assess texts. The first implies (1) selecting a subset of topics which exceeded a certain established threshold of a topic probability in the text and (2) identifying a new text parameter as a certain subset of topics in a text. For our experiments, we use a threshold equal to $1/|T|$, where $|T|$ is the number of topics in the topic model. For example, the probability threshold for a topic in a book is 0.05 for a topic model with 20 topics and 0.025 for a topic model with 40 topics. The disadvantage of this approach is information loss of the topics below the established threshold.

Another approach requires assessment of each topic contribution to the book topics and is conducted based on the probability values of each topic in a certain book. In our work, this idea is realized in the cosine distance between the vector of topics probabilities in the document andthe vector of assessments of individual topics. We also compared both approaches presented above to assess a document and implemented the average topic score as a parameter for a subset of topics selected based on the probability threshold.

Table 2 shows the values of Spearman's correlation coefficients for the complexity levels of texts and the topics weight. The results obtained allow us to make the following main observations. Firstly, the cosine distance between the vectors of estimates and topics probabilities in the text enables higher absolute values of correlation coefficients. The results obtained are of little statistical significance if the probability threshold is employed as a method of selecting topics. Second, for all the collections of books, except for the collection of Russian as a foreign language, there is a negative correlation between the cosine distance and text complexity. Third, the phenomena are observed for both models. Thus, with text complexity increasing, the vector of topics probabilities in texts acquires similarities with the vector of topics weights.

Spearman's correlation indices for text complexity and NPMI are provided in Table 3 below.

As in the previous series of experiments on interdependency between the topic weight and text complexity level, we observed negative correlations while using the cosine distance as a way to calculate a text metric. However, for

Table 2

Correlations between the Average Topic Weight (RNC Frequencies) and Text Complexity Level

Collection ot text	ρ, q, where ρ is Spearman's correlation coefficient, $\rho < 10^{-q}$			
	probability threshold		cosine distance	
	ARTM	LDA	ARTM	LDA
Social Studies	0.55, 1	0.51, 1	-0.84, 3	-0.72, 2
Biology	-0.13, 0	-0.18, 0	-0.54, 2	-0.60, 2
Fiction	-0.07, 0	0.14, 0	-0.36, 3	-0.34, 3
Russian as a foreign language	0.11, 0	0.18, 2	-0.00, 0	-0.20, 2

Note. *In Table 2 and below, q index is preceded by a comma, and statistically significant if ≥ 2.*

NPMI metric, the absolute values of the correlation coefficient are smaller, and the revealed dependencies have a lower level of statistical significance.

We also identified few values inconsistencies when the negative correlation alternates with positive. A probable reason for this has to do with the influence of two opposing factors on NPMI metric noted above.

Table 4 presents results of correlation tests for text complexity level and the mean cosine distance between word2vec of the topic words. The results obtained are similar to the results of correlation tests for topic weight and text complexity level: we observe a negative correlation for both types of topic models when using the cosine distance for all text collections, except for the collection of texts of Russian as a foreign language.

Dependencies between Text Complexity and Topic Structure

The next series of correlation tests was aimed at testing the hypothesis of interdependency between text complexity and topic distribution structure in texts. We define topic distribution entropy as Shannon entropy of document-topic distribution p(t|d) of a pre-trained topic model. Since Shannon entropy reaches its maximum on uniform probability distribution, its smaller values would indicate that a text is focused on a small number of topics. Vice versa, higher entropy indicates a text focused on a wide range of topics with generally sparse and shallow coverage.

The results of the correlation tests between text complexity and topic distribution entropy are provided in Table 5 below.

Spearman's correlation tests indicate statistically significant positive correlation between topic distribution entropy of texts and their complexity level in Biology and Fiction collections. We also observe maximum of topic distribution entropy in cases when probabilities of all topics in texts are equal. That is true about polythematic texts with equally distributed coverage of topics. On the contrary, topic distribution entropy is minimal, if probability of one of the topics is close to 1, hence the text is monothematic.

Based on the results obtained, we conclude that more complex Biology and Fiction texts cover more topics of comparable sizes. As texts of Russian as a foreign language are much shorter and their complexity does not grow across texts, their results differ dramatically: a short text cannot cover numerous topics and embodies monothematicity.

In texts of Social Studies collection, as in Biology and Fiction texts, we observe a positive correlation for both types of models. However, statistical significance of the results is lower, which is either caused by a smaller size of the sample or by the fact that text complexity of this collection is largely

Table 3

Correlations between NPMI Average Number of Topics and Text Complexity Level

Collection oftexts	ρ, q, where ρ is Spearman's correlation coefficient, ρ <10^{-q}			
	probability threshold		cosine distance	
	ARTM	LDA	ARTM	LDA
Social Studies	0.41, 0	0.63, 1	-0.68, 2	-0.61, 1
Biology	-0.01, 0	0.33, 0	-0.32, 0	-0.46, 1
Fiction	-0.17, 1	-0.16, 1	-0.25, 2	-0.16, 1
Russian as a foreign language	0.13, 1	-0.07, 0	-0.13, 1	-0.09, 0

Table 4

Correlations between Mean Topic Word2vec (RNC Model) and Text Complexity

Collection oftexts	ρ, q, where ρ is Spearman's correlation coefficient, ρ <10^{-q}			
	probability threshold		cosine distance	
	ARTM	LDA	ARTM	LDA
Social Studies	-0.69, 2	-0.78, 3	-0.62, 1	-0.43, 0
Biology	-0.22, 0	-0.13, 0	-0.58, 2	-0.59, 2
Fiction	0.16, 1	0.04, 0	-0.39, 4	-0.28, 2
Russian as a foreign language	0.12, 1	-0.01, 0	0.09, 0	-0.17, 1

Table 5

Correlations between Text Complexity Level and Topic Distribution Entropy

Collection of texts	ρ, q, where ρ is Spearman's correlation coefficient, $\rho < 10^{-q}$	
	ARTM	LDA
Social Studies	0.59, 1	0.38, 0
Biology	0.76, 4	0.59, 2
Fiction	0.42, 5	0.38, 4
Russian as a foreign language	-0.10, 0	0.21, 2

dependent on different parameters. The findings may also be caused by frequency and semantic properties of the most probable topics which correspond to the observations obtained in the previous experiments (see Tables 2-5 above). Thus, the foregoing suggests that, as a function of the subject area, properties of different types of topics affect text complexity diversely. Consequently, a more accurate definition of complexity requires taking into account numerous properties of various texts including semantics, frequency, text topic structure.

Topic structures of individual texts were scrutinized based on the topic distributions of ARTM models trained on Social Science texts and a segment model with 20 topics. The results of the experiment are presented in Fig. 2.

Graphs in Fig. 2 show that textbooks of higher grades demonstrate lower probability of the most probable topic. On the opposite, the probability of the most probable topics in texts of the 5th and 6th grades is exponentially higher than probability of any other topic. Therefore, text complexity growth results in texts becoming polythematic, while lower complexity texts predominantly maintain their monothematicity. This observation is consistent with the findings of the study on interdependency of entropy and complexity: text complexity growth leads to the increase of topic distribution entropy.

Topics Qualitative Analysis

Analysis of words representing topics is a mandatory step in the Topic Modeling algorithm. To this end, we selected several most interpreted topics of ARTM segment model with 20 topics, trained them on a collection of Social Studies texts. Topics interpretability was assessed with word2vec scores. Table S1 in Appendix shows examples of the selected topics and their corresponding word2vec scores. In addition, for each topic selected we identify 5 texts, in which the share of the topic is maximum.

Analysis of the above topics reveals, that most of the topics, represented by the words semantically close in the word2vec, have the highest weight in the texts of the 9th – 11th grades. This observation is consistent with the results of the correlation tests presented in Table 4. However, there are cases inconsistent with this observation, for example, topics 1 and 4, which have a high weight in texts of the 6th and 7th grades, respectively. This observation indicates that text complexity estimate implies assessment of more than one topic metric, i.e. topic interpretability word2vec.

DISCUSSION

In this research, we investigate applicability of the two state-of-the-art topic models, i.e. Latent Dirichlet Allocation (LDA), and Additive regularization of topic models (ARTM), for assessment of text complexity in schoolbooks. We adopt three training strategies for representing books to train topic models: (i) full-length textbooks, (ii) segments with a maximum size of 1000 words, (iii) paragraphs. When assessing topic coherence, we used two metrics (word2vec, NPMI) and two corpora (RNC, Taiga).

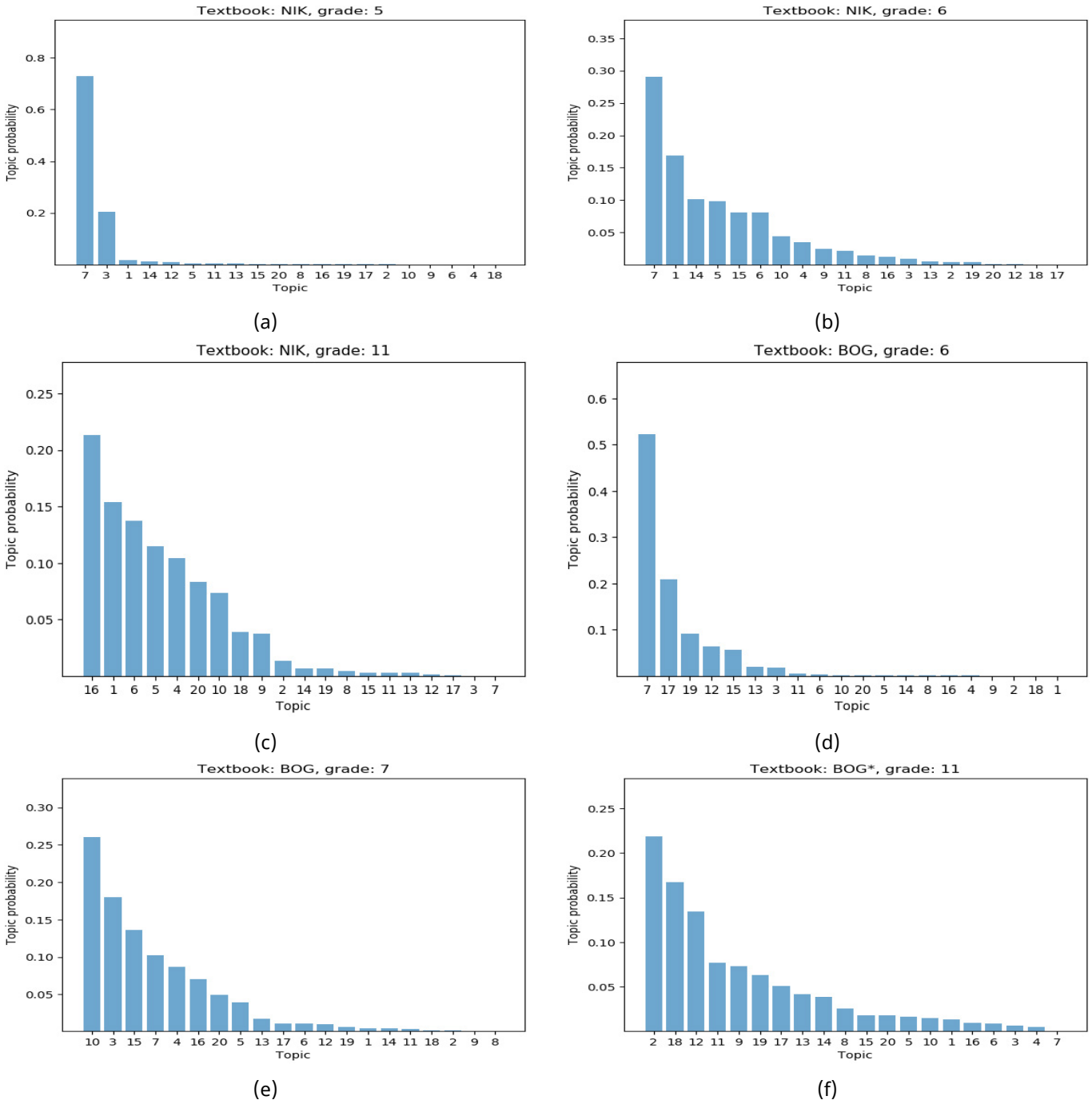
We also validated the topical pattern of Russian school textbooks: typically, there are 15-20 topics in a collection of academic texts. This result does not depend either on the chosen method for assessing the quality of topics (word2vec or NPMI) or the text corpus (RNC or Taiga). Based on Social Studies textbooks, we presume that the selected 20 topics are well interpreted by experts.

ARTM demonstrated better evaluation results compared to LDA in terms of topic coherence metrics. Based on the trained topic models, we revealed correlation between book grades and properties of the highest weights topics.

Spearman rank correlation results demonstrate the following statistically significant dependencies: (i) higher-grade texts, i. e. more complex texts, are characterized by higher topics coherence; (ii) higher-grade texts contain more topics and the topics are equally distributed in the texts (distribution entropy of topics correlates with text complexity); (iii) higher grades texts contain fewer topic words of average frequency.Specifically, (i) implies that narrower, more specialized topics are taught in senior grades, i.e. there is a

Figure 2

Distribution of Topics in Social Science Texts



Note. (a) - (c) refer to textbooks by Nikitin A.F., (d) - (f) refer to textbooks by Bogolyubov L.N.
* marks advanced complexity levels of books

transition from general to more specific in introduction of educational material. This finding is also supported by (iii), which confirms that rarer, i.e. more specific terms are used in senior grades.

As for (ii), it implies that author's attention in lower grades texts is focused on a limited, much lower number of topics than in the senior grades (Figure 2). Thus, throughout the school, textbooks acquire more topics, widening world view from narrow to broad in senior grades. Although this statement may seem self-evident, the methods we have developed make it possible to obtain a quantitative assessment of this process for each subject and each series of textbooks of different grades.

The correlation coefficient of grade level received in the study is moderate, which indicates that we observe a lack of systematic complexification of textbooks across grades.

The latter is similar to the results obtained in (Berendes & Vajjala, 2018) for German school textbooks. We have also established correlation of two additional parameters, i.e. topic weight and distribution entropy of topics with text complexity. In addition to this we confirmed the previously obtained results (Sakhovskiy et al., 2020a, Sakhovskiy et al., 2020b) and replicated them on a larger representative balanced collection of texts.

CONCLUSION

The article offers a consistent description of experimental application of topic modeling algorithms to evaluating a text structure. The dataset used in the study was compiled of graded texts of four academic subcorpora, i.e. textbooks of Social studies and Biology for Russian schools, tests materials for Russian as a foreign language and texts for extracurricular reading. We revealed and described patterns of structural change in educational texts as they become more complex from grade to grade. The study confirmed the hypothesis that topic properties change systematically across grade levels. We also offer a list of parameters discriminating various educational texts structures, and present the latter as a set of topics designed of keywords. We validated the list of introduced parameters, i.e. number of topics, topic coherence, topic distribution, and topic weight as predictors of (1) a topics change from specific to general; and (2) an increasing text complexity.

We conclude that topic models can be used to assess text structure dynamics. Due to the ease of computing values of these parameters with available software programs, they can be used along with traditional text complexity assessment tools. We also emphasize that the studied parameters characterize only one aspect of text complexity, i.e. structural organization of text topics.

In the algorithm suggested, complexity is assessed not by commonly used linguistic parameters (length of sentences, number of long words, TTR, etc.), but by computational parameters related to textbook topics. The proposed approach offers new insights into the problems of text complexity and methods of presenting educational material in a textbook. Automatically obtained metrics of the introduced parameters, i.e. number of topics, topic coherence, distribution, and weight, enable to evaluate sustainability of strategy and presentation techniques across a text/textbook. The algorithm designed and developed for Russian texts can be further extrapolated to other languages and texts, provided the language is well-resourced and a representative corpus is available to compute word2vec and NPMI.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Higher Education of the Russian Federation, agreement No. 075-15-2022-294 dated 15 April 2022 and by the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030).

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS CONTRIBUTIONS

Valery Solovyev: Conceptualization, Methodology, Supervision, Project administration.

Marina Solnyshkina: Resources, Writing - Original Draft.

Elena Tutubalina: Resources, Writing - Original Draft.

REFERENCES

Al Tamimi, A. K., Jaradat, M., Al-Jarrah, N. & Ghanem, S. (2014). AARI: Automatic Arabic readability index. *International Arab Journal of Information Technology, 11*(4), 370-378.

Arfe, B., Mason, L. & Fajardo, I. (2018). Simplifying informational text structure for struggling readers. Reading and Writing, 31, 2191-2210. https://doi.org/10.1007/s11145-017-9785-6

Bailin, A., & Grafstein, A. (2016). *Readability: Text and context*. Palgrave Macmillan.

Balyan, R., McCarthy, K.S., & McNamara, D.S. (2018). Comparing machine learning classification approaches for predicting expository text difficulty. In *The Thirty-First International Flairs Conference* (FLAIRS-31) (pp. 421-426). AAAI press.

Berendes, K., & Vajjala, S. (2018). Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track? *Journal of Educational Psychology, 110*(4), 518-543. https://doi/10.1037/edu0000225

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

Borda, M. (2011). *Fundamentals in information theory and coding*. Springer.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL* (pp. 31-40). Gunter Narr Verlag Tubingen.

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

Chen, Y.-H., & Daowadung, P. (2015). Assessing readability of Thai text using support vector machines. *Maejo International Journal of Science and Technology, 9*(3), 355-369.

Chen, Y.-T., Chen, Y.-H., & Cheng, Y.-C. (2013). Assessing Chinese readability using term frequency and lexical chain. *International Journal of Computational Linguistics and Chinese Language Processing*, 18(2), 1-18.

Chuang, Hsiao-yu (1993). Topical structure and writing quality: A study of students' expository writing. *Theses Digitization Project, 686*. California State University.

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. T*ESOL Quarterly*, 42(3), 475-493.

Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014).What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language, 26*(1), 92-113.

Diakidoy, I.-A. N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology, 28*(3), 335-356. https://doi.org/10.1016/S0361-476X(02)00039-5

Eremeev, M. A., & Vorontsov, K. V. (2020). Quantile-based approach to estimating cognitive text complexity. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue", 19*, 256-269. РГГУ.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.

Gatiyatullina, G.M., Solnyshkina, M.I., Kupriyanov, R.V., & Ziganshina, C.R. (2023). Lexical density as a complexity predictor: The case of science and social studies textbooks. Research Result. *Theoretical and Applied Linguistics, 9*(1), 11–26. 10.18413/2313-8912-2023-9-1-0-2

Gazzola, M., Leal, S., Pedroni, B., Theoto Rocha, F., Pompéia, S., & Aluísio, S. (2022). Text complexity of open educational resources in Portuguese: mixing written and spoken registers in a multi-task approach. *Language Resources and Evaluation, 56*(2), 621-650.

Hobbs, J. (1990). *Literature and cognition*. Stanford.

Hoffman, M.D., Blei, D., & Bach, F. (2010). Online inference for latent Dirichlet allocation. In *Neural Information Processing Systems* (pp. 856-864). Curran Associates, Inc.

Ivanov, V.V, Solnyshkina, M.I., & Solovyev, V.D. (2018). Efficiency of text readability features in Russian academic texts. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2018"* (pp. 267-283). RGGU.

Kendeou P., & Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition, 35*(7), 1567-1577. https://doi.org/10.3758/BF03193491

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. Nilsson (Ed.), *Perspectives on memory research* (pp. 329-365). Psychology Press.

Kutuzov, A., & Kuzmenko, E. (2017). Web vectors: A toolkit forbuilding web interfaces for vector semantic models. In *International Conference on Analysis of Images, Social Networks and Texts* (pp. 155-161). Springer Cham.

Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods, 50*(3), 1030-1046.

Laposhina, A. N., Veselovskaya, T. V., Lebedeva, M. U., & Kupreshchenko, O. F. (2018). Automated text readability assessment for Russian second language learners. In C*omputational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2018"* (pp. 396-406). RGGU.

Lau, J. H. & Baldwin, T. (2016). The sensitivity of topic coherence evaluation to topic cardinality. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologie* (pp. 483–487). Association for Computational Linguistics.

Loftus, G.R. (1983) The continuing persistence of the icon. *Behavioral and Brain Science, 6*(1), 28. https://doi.org/10.1017/S0140525X00014461

Martinc, M., Pollak, S., & Robnik-Šikonja, M. (2021). Supervised and unsupervised neural approaches to text readability. *Computational Linguistics, 47*(1), 141-179.

McBride, D. M., & Cutting, J. C. (2015). *Cognitive psychology: Theory, process, and methodology.* Sage.

McCallum, A.K. (2002). *Mallet: A machine learning for language toolkit*. University of Massachusetts Amherst.

McNamara, D.S., Graesser, A. C., McCarthy, P. M., & Cai, Zh. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43. https://doi.org/10.1207/s1532690xci1401_1

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*(4), 292-330. https://doi.org/10.1080/01638530902959943.

McNamara, D. S., Roscoe, R., Allen, L., & Balyan, R., & McCarthy, K.S. (2019). Literacy: From the perspective of text and discourse theory. *Journal of Language and Education, 5*(3), 56-69. https://doi.org/https://doi.org/10.17323/jle.2019.10196.

Mulunda, C.K., Wagacha, P.W., & Muchemi, L. (2018). Review of trends in topic modeling techniques, tools, inference algorithms and applications. In *2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMI)* (pp. 28-37). IEEE.

Newman, D., Jey, H. L., Karl, G., & Timothy, B. (2010a). *Automatic evaluation of topic coherence. Human language technologies*. (pp. 100-108). Association for Computational Linguistics.

Newman, D., Noh, Y., Talley, E., Karimi, S., & Baldwin, T. (2010b). Evaluating topic models for digital libraries. *Proceedings of the 10th annual joint conference on Digital libraries* (pp. 215-224). Association for Computing Machinery.

Nikolenko, S.I. (2016). Topic quality metrics based on distributed word representations. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 1029-1032). Association for Computing Machinery.

Ninio, A., & Snow, C. (1999). The development of pragmatics: Learning to use language appropriately. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of language acquisition* (pp. 347-383). Academic Press.

Oborneva, I. V. (2006). *Automated assessment of the complexityof educational texts based on statistical parameters* [Unpublished doctoral dissertation,]. Institute of Contents and Methods of Training RAO.

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45-50). University of Malta.

Roehling, J. V., Hebert, M., Ron Nelson, J., & Bohaty, J. J. (2017). Text structure strategies for improving expository reading comprehension. *The Reading Teacher, 71*(1), 71- 82. https://doi.org/10.1002/trtr.1590

Sakhovskiy A., Solovyev V., & Solnyshkina M. (2020a). Topic modeling for assessment of text complexity in Russian textbooks. *Proceedings of Ivannikov ISPRAS Open Conference* (pp. 102-108). IEEE.

Sakhovskiy, A., Tutubalina, E., Solovyev, V., & Solnyshkina, M. (2020b). *Topic modeling as a method of educational text structuring. DeSE* (pp. 399-405). IEEE.

Santucci, V., Santarelli, F., Forti, L., & Spina, S. (2020). Automatic classification of text complexity. *Applied Sciences, 10*(20), 7285.

Shavrina, T., & Shapovalova, O. (2017). To the methodology ofcorpus construction for machine learning: "Taiga" syntax tree corpus and parser. In *Proceedings of the International Conference "Corpus Linguistics–2017"* (pp. 78-84). St Petersburg State University.

Si, I., & Callan, J. (2001). A statistical model for scientificreadability. In *Proceedings of the tenth international conference on Information and knowledge management* (pp. 574-576). Association for Computing Machinery.

Solnyshkina, M.I., Harkova, E.V., Kazachkova, M.B. (2020) The structure of cross-linguistic differences: Meaning and context of 'readability' and its Russian equivalent 'chitabelnost'. *Journal of Language and Education, 6*(1), 103-119. https://doi.org/10.17323/jle.2020.7176

Solnyshkina, M.I., Harkova, E.V., & Kiselnikov, A.S. (2014). Unified (Russian) state exam in English: Reading comprehension tasks. *English Language Teaching, 7*(12), 1-11. https://doi.org/10.5539/ELT.V7N12P1

Solovyev, V., Solnyshkina, M., Ivanov, V., & Batyrshin, I. (2019). Prediction of reading difficulty in Russian academic texts. *Journal of Intelligent & Fuzzy Systems, 36*(5), 4553-4563. https://doi.org/10.3233/JIFS-179007

Solovyev, V. D., Solnyshkina, M. I., & McNamara, D. S. (2022). Computational linguistics and discourse complexology: Paradigms and research methods. *Russian Journal of Linguistics, 26*(2), 275-316.

Tanaka-Ishii, K., Tezuka, S., & Terada H. (2010). Sorting texts by readability. *Computational Linguistics, 36*(2), 203-227. https://doi.org/10.1162/coli.09-036-R2-08-050

Thorndyke, P.W. (1977). Cognitive structures in comprehension and memory in narrative discourse. *Cognitive Psychology, 14*, 560-589. https://doi.org/10.1016/0010-0285(77)90005-6

Toma, I., Marica, A. M., Dascalu, M., & Trausan-Matu, S. (2021). Readerbench–automated feedback generation for essays in Romanian. *University Politehnica of Bucharest Scientific Bulletin Series C-Electrical Engineering and Computer Science, 83*(2), 21-34.

Vorontsov, K., Frei, O., Apishev, M., Romov, P., & Dudarenko, M. (2015). Bigartm: Open source library for regularized multimodal topicmodeling of large collections. In *International Conference on Analysis of Images, Social Networks and Texts* (pp. 370-381). Springer Cham.

Vorontsov, K., & Potapenko, A. (2015). Additive regularization of topic models. Machine Learning, 101(1-3), 303-323.

Watson T.R. (2016). Discourse topics. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Williams J.P. (2005). Instruction in reading comprehension for primary-grade students: A focus on text structure. *The Journal of Special Education, 39*(1), 6-18.

APPENDIX

Examples of the selected topics and their corresponding word2vec scores

Table S1

ARTM Segment Model Topics and Texts with Maximum Topic Weights. A Lower Word2vec Qt Score Corresponds to a Better Topic

#	Qt	Topic	Most probable words of the topic	Books with the highest weight	Weights of topics in texts
1	0.61	Political system	federation power Russian statelaw state organ Constitution of RF federal	Nik-6	0.17
				Nik-10-11	0.16
				Nik-11	0.15
				Bog-9	0.11
				Nik-8-9	0.08
4	0.64	Law and order	crime law court criminal administrative punishment offense law responsibility authority	Nik-10-11	0.23
				Nik-7	0.17
				Bog-9	0.14
				Nik-9	0.12
				Nik-11	0.10
13	0.65	Science	science knowledge scientific education human cognition scientist research activity truth	Bog-10*	0.17
				Nik-10	0.10
				Bog-10	0.08
				Bog-8	0.05
				Nik-8-9	0.05
18	0.65	Development of society	society country development economic life social production economy modern social	Bog-11*	0.17
				Nik-10	0.09
				Bog-10	0.09
				Bog-8	0.09
				Bog-10*	0.08
16	0.65	Economy	economy money commodity state country price market income market economic	Nik-11	0.21
				Nik-9	0.18
				Bog-8	0.18
				Bog-7	0.07
				Nik-8-9	0.05
8	0.66	Religion	religion religious person society philosopher life history century state god	Nik-10	0.12
				Nik-8-9	0.09
				Bog-10*	0.08
				Nik-7	0.05
				Nik-8	0.04
14	0.66	National identity	people person country national state Russia culture Russian conflict language	Nik-6	0.10
				Nik-10	0.07
				Nik-8-9	0.07
				Nik-8	0.06
				Bog-8	0.05

#	Qt	Topic	Most probable words of the topic	Books with the highest weight	Weights of topics in texts
19	0.67	Personality	person activity society personality social life consciousness need social spiritual	Bog-10*	0.25
				Bog-10	0.12
				Nik-10	0.12
				Bog-9	0.09
				Bog-11*	0.06
11	0.68	Culture	culture art spiritual mass society human value artistic cultural work	Bog-10	0.10
				Bog-11*	0.08
				Nik-8	0.05
				Nik-10	0.04
				Bog-10*	0.04

https://doi.org/10.17323/jle.2023.17632

A Bibliometric Analysis of English for Specific Purposes from 2011 to 2023 Using Citespace: Visualizing Status, Themes, Evolution, and Emerging Trends

Sining Tan [1], Madhubala Bava Harji [2], Xiaogang Hu [3]

[1] Canvard College, Beijing Technology and Business University, Beijing, China
[2] SEGI University & Colleges, Kota Damansara, Selangor, Malaysia
[3] Xinyu College, Jiangxi, China

ABSTRACT

Introduction: Research on English for Specific Purposes (ESP) emerged in the 1960s. A few researchers conducted reviews on ESP literature. However, there is currently a lack of up-to-date and comprehensive bibliometric analysis covering the last decade from an international perspective, particularly covering the last decade's developments in ESP research.

Purpose: Combining bibliometric analysis and systematic review of the literature on English for Specific Purposes, this study serves to objectively maps the knowledge area, and aims to identify the current status, major research themes, evolution, and the emerging trends in this field.

Method: To maintain the objectivity and transparency, the review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocol. Based on literature retrieved from the Web of Science core collection, a total of 1657 bibliometric records published from 2011 to 2023 were visualized and analyzed via Citespace.

Results: The current research status through publication distribution and co-country network shows that the research is undergoing a steady increase and international authorship. Clusters and systematic review of citing articles indicate four themes, i.e., "linguistic inquires", "teaching pedagogy", "student learning", and "teacher development". Through the timezone view of keyword co-occurrence network, four features, such as the predominance of genre, corpus and pedagogy, varied pedagogies, varied research methodologies, and technology-assisted teaching, were identified. The keyword and citation burst detection, as well as systematic review of citing articles, were conducted to identify the research trends. It is found that EAP, teacher development, needs analysis in under-researched settings, as well as EMI and Corpus-based teaching pedagogies are the new frontiers in this field.

Conclusion: The field of English for Specific Purposes continues to experience exponential growth and development, indicating an ongoing expansion and advancement of research in this area. This paper provides references for researchers to understand the status, hidden structure, evolution, and emerging trends of research on English for Specific Purposes.

KEYWORDS

English for Specific Purposes, bibliometric analysis, trends, citation, visualization

Citation: Tan S., Harji M. B., & Hu X. (2023). A Bibliometric Analysis of English for Specific Purposes from 2011 to 2023 Using Citespace: Visualizing Status, Themes, Evolution, and Emerging Trends. *Journal of Language and Education, 9*(3), 159-175. https://doi.org/10.17323/jle.2023.17632

Correspondence:
Sining Tan,
ansining@126.com

Received: July 16, 2023
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

English for Specific Purposes (ESP) primarily caters to non-native English speakers in academic and professional contexts, with a particular emphasis on the learning, teaching, and researching specialized variations of the English language (Belcher, 2006). It is a language teaching approach in which all decisions as to the teaching content and methods are determined by the academic, professional, social and cultural needs of the learners (Hutchinson & Waters, 1987). The main aim of ESP is to facilitate the communication or the social and psycho-

logical integration of learners into "the target professional or academic community in which communication is fostered mainly in English" (Vesna, 2022, p.118). Due to its interdisciplinary nature and continuous evolution, ESP has emerged as a highly dynamic and vibrant field within foreign language education and research (Hu, 2019).

The 1960s marked the commencement of ESP research, spurred by significant advancements in science, technology, economics, and global trade. Johns (2012) conducted a historic overview of ESP development, and divided it into four stages of development. The foci of interest during the Early Years (1962-1981) were English for science and technology in academic settings, and rhetorical and discourse analysis. The second stage is the Recent Past (1981-1990), when the research scope broadened and some central concepts, including genre and rhetorical moves, were introduced. This period saw a shift in research focus towards ESP practitioner training, vocational English, and interlanguage. Modern Era (1990-2011) witnessed a surge in ESP-related publications in international journals, and a prominent focus on genre analysis research, accompanied by thorough exploration of corpus studies. According to Johns (2012), the anticipated developments in the fourth stage, i.e., the Future (2011 plus), included international authorship, the different roles of ESP practitioners in teaching, curriculum design, material creation, collaboration, research, and assessment. Additionally, this stage was expected to feature diverse methodologies, multimedia course materials, a broad spectrum of ESP teaching contexts, and a teaching focus on genre awareness.

LITERATURE REVIEW

To understand the existing repository of knowledge of a specific field, literature review is a prerequisite (Jia & Harji, 2023). More recent reviews of ESP research (Basturkmen, 2021, 2022; Dou et al., 2023; Hu, 2019; Hyland & Jiang, 2021; Liu & Hu, 2021) have provided insights into the current development in this field. Basturkmen (2021) drew upon her twenty years of experience as an academic in ESP and English for Academic Purposes (EAP), and explored key research themes in ESP literature, including well-established topics and emerging areas of interest. Linguistic analysis has been a significant focus, driven by the need for linguistic descriptions to inform ESP pedagogy. The ESP literature also contains many case studies on teaching practices. The review highlights underexplored areas such as learning ESP, diverse teaching practices, and the educational needs of ESP practitioners, suggesting these as promising avenues for future research. In another study, Basturkmen (2022) randomly selected 23 articles from 2018 to 2020 issues of the field's flagship journal *English for Specific Purposes* for empirical studies in the Asia Pacific region to identify recent themes in ESP research. Four broad themes emerged from the sample, including teacher/student identity issues, teaching/learning applications, under-researched contexts, and linguistic inquiries. This conclusion supported her findings in the previous study.

Dou et al. (2023) conducted a review on ESP literature from the 1960s to 2023, and summarized the distinctive characteristics of various stages in the development of ESP. They concluded that needs analysis remained central to ESP, but its conceptualization has evolved from purely target situation analysis to also considering learning processes and contexts. Technology was transforming ESP through data analysis, materials design, and online learning platforms. They predicted that the future would likely see more research on intercultural communication, continued technology impacts, and needs analysis that incorporates learner motivation and perspectives.

In general, these studies provide valuable insights into the research trends of ESP literature. However, these studies are conventional reviews that adopt a descriptive and impressionistic approach, relying on manual analysis or personal experiences. A few quantitative reviews (Hu, 2019; Hyland & Jiang, 2021; Liu & Hu, 2021) used a more objective and statistical method, i.e., Bibliometric Analysis (BA). BA is a robust approach to delve into extensive literature sets in a quantitative manner, aiming to unveil the intellectual framework and emerging patterns within a specific field (Diao et al., 2022). BA addresses the limitations of conventional or narrative reviews that depend on the experiential and intuitive viewpoints of their authors (Liu & Hu, 2021). It aids researchers in uncovering concealed trends and concepts within the literature that manual reviews may inadvertently miss (Diao et al., 2022; Zhong et al., 2019).

Hu (2019) conducted BA of ESP research in China from 2000 to 2017, mapping out the research hotspots and trends, including ESP teaching methodologies and models, needs analysis, and ESP curriculum development and design. They suggested that future exploration should focus on areas such as developing an ESP evaluation system, investigating ESP instructors, building ESP corpora, and enhancing ESP textbook development. Their BA primarily focused on the Chinese context rather than on a global scale. Hyland and Jiang (2021) also adopted the bibliometric analysis approach to examine the EAP literature from 1980 to 2020, to analyze shifts in research themes, influential publications, author contributions, and geographical sources. They found a sustained interest among researchers in teaching, learning, and classroom methodologies. From 2001, a significant portion of research has shifted its focus towards examining the contexts, discourses, identity, educational practices, and genre, and more involvement of Asian countries in the field. The focus of this study is on one branch of ESP, i.e., EAP, instead of ESP.

By looking into the ESP field from an international perspective, Liu and Hu (2021) used bibliometric and visual analysis techniques to systematically map and quantify the research

topics, influential publications, and historical trends of based on citation data. 1092 articles published from 1980 to 2018 in two flagship journals, *English for Specific Purposes* and *Journal of Journal of English for Academic Purposes*, were analyzed. They provided a theme-based three-staged characterization of ESP development, which corresponds mostly to that of ESP's historical phases in the study of Johns (2012), but presented a more comprehensive and objective evolution map of ESP literature. However, this review has certain limitations. It relied solely on articles from the two lead journals as its source data, utilized only the co-citation analysis technique, and did not include articles published from 2019 to 2023. Thus, there is a need for a more up-to-date and systematic review that includes the latest ESP research and its sub-branches with more analysis techniques.

In view of the above, this study adopted Citespace, one of multiple tools for BA, to analyze the literature on ESP research from 2011 to 2023. Citespace is a now a widely used tool in BA and data visualization to unfold research patterns, hotspots and evolution trends in fields such as medicine, demography, sociology, geography, education, and related fields (Jiang et al., 2022; Sun & Asmawi, 2022; Tu et al., 2022). Through visualizing some knowledge networks in ESP literature, this study aims to address the following research questions:

(1) What is the current status of ESP research from 2011 to 2023?

(2) What are the major themes of ESP research from 2011 to 2023?

(3) How did ESP research evolve from 2011 to 2023?

(4) What are the emerging trends in ESP research from 2011 to 2023?

The findings presented in this paper provide ESP researchers and practitioners with several research networks and insights to complement conventional narrative reviews, facilitate the understanding of ESP, and promote further research in this field.

METHOD

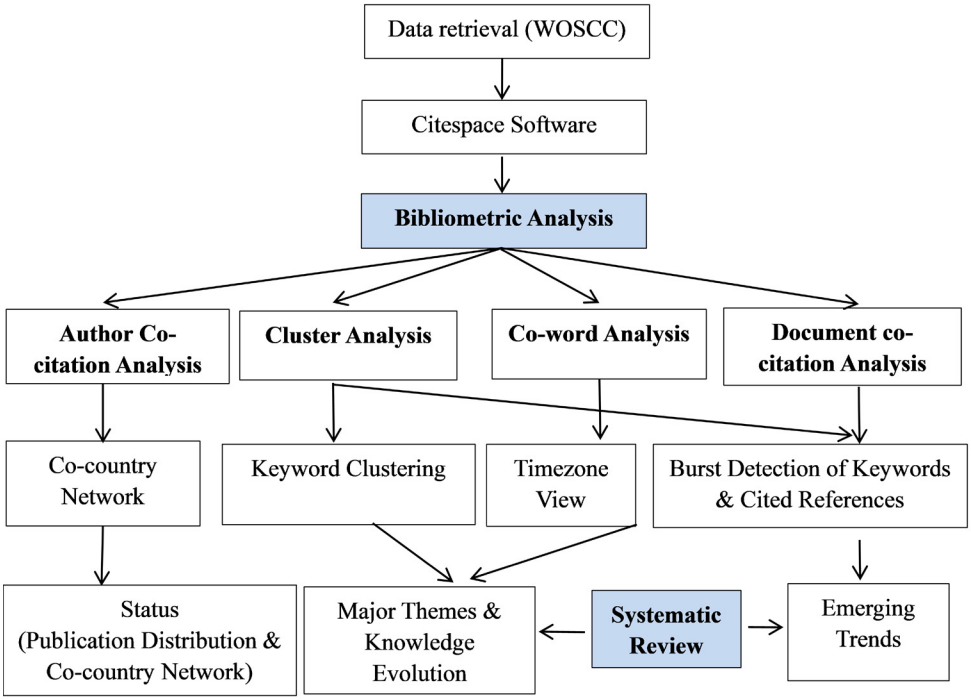
Using BA, this study provides a quantitative, objective, and systematic overview of literature on ESP. The methodology of systematic review was also adopted. The research design is presented in Figure 1. Publication distribution and co-country network were analyzed to identify the research status, cluster analysis was conducted to examine the major themes, timezone view of keyword co-occurrence network and burst detection, along with systematic review, were carried out to explore major themes, knowledge evolution, and emerging trends.

Data Collection and Processing

The literature quantity with the time distribution figure can shed light on the research status and development of the discipline or the field, which works as a prediction for future research trends (Diao et al., 2022; Fu et al., 2021). This review

Figure 1

Research Design



adapted a modified diagram of Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA).

This review obtained data from the Web of Science Core Collection (WOSCC) on September 25, 2023, because the WOSCC included high-quality journal literature widely recognized and influential worldwide, and is suited for extensive and bibliometric analysis on a large scale (Jia &Harji, 2022). The ESP classification of Dudley-Evans and St. John (1998) was adopted in the process of the search queries. ESP can be classified into two categories: EAP and English for Professional Purposes (EOP). Both EAP and EOP can be further divided into several branches, as seen in Figure **2**.

Topic terms comprised all the sub-branches of ESP, including English for Science and Technology, English for Medical Purposes, English for Legal Purposes and English for Finance and Economics, etc., Therefore, publications that mention these ESP branches in titles, abstracts, and/or keywords were included in the data.

Inclusion and Exclusion Criteria

As suggested by Jia and Harji (2022), and Wang et al. (2023), some inclusion and exclusion criteria were adopted in data collection and screening. The literature database was obtained from the WOSCC. During this stage, the publication time, language, topic, and document types were set as inclusion and exclusion criteria.

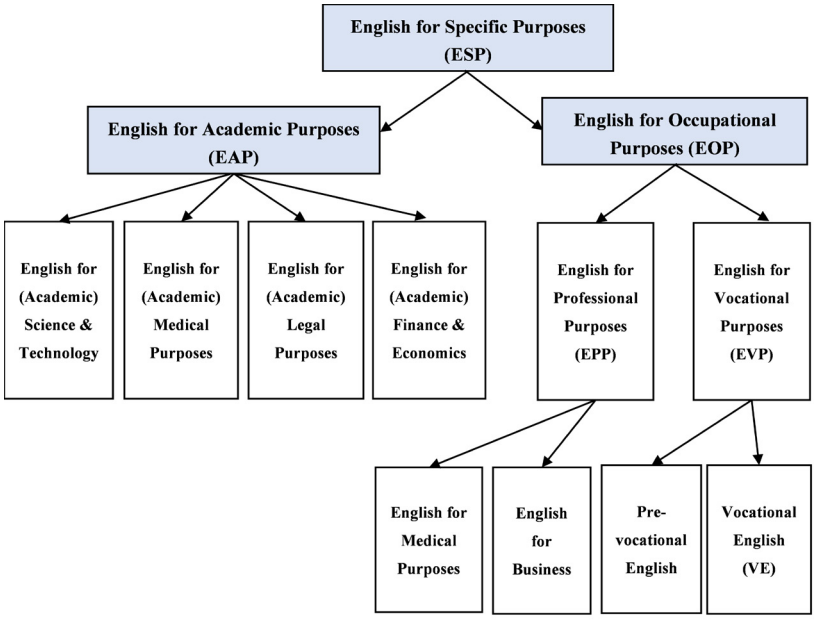
Firstly, according to Hwang and Tsai (2011), reviewing literature published over the past decade serves as an effective approach for identifying patterns and trends in the educational field. Building upon the study of Johns (2012), the time span for this research was set from 1st January, 2011 to 25th September, 2023, excluding publications in other years. Secondly, in terms of language, the publications in the English language were included, excluding papers in other languages. In addition, the filters of document types "Article" or "Review" were applied to refine the results (Chen, 2017; Jia & Harji, 2023). As for topic, all branches of ESP were used in the data retrieval.

Following the modified PRISMA diagram (Wang et al., 2023) and Chen's methodology (2017) in data retrieving for Citespace, the search query input the topic term "English for Specific Purpose*" in the WOSCC database. This query yielded 750 records as SET #1. Similarly, two more queries were conducted using the terms "English for Academic Purpose*" and "English for Occupational Purpose*", along with sub-branches of EAP and EOP, maintaining the same document type and language restrictions. These queries resulted in 997 and 415 records for SET #2 and SET #3, respectively. Subsequently, SET #1 to SET #3 were combined using "OR", and the amalgamation of these search results led to the retrieval of the final 2075 records (Figure 3).

In the first round of screening, only 1981 publications in the field of linguistics or education were included. All other publications beyond these two fields were excluded, because they were different to the subject of this review (Wang et al., 2023). All the 1981 records were downloaded and saved in "Abstract and full record (including cited references)", and input into Citespace for data analysis. In the second

Figure 2

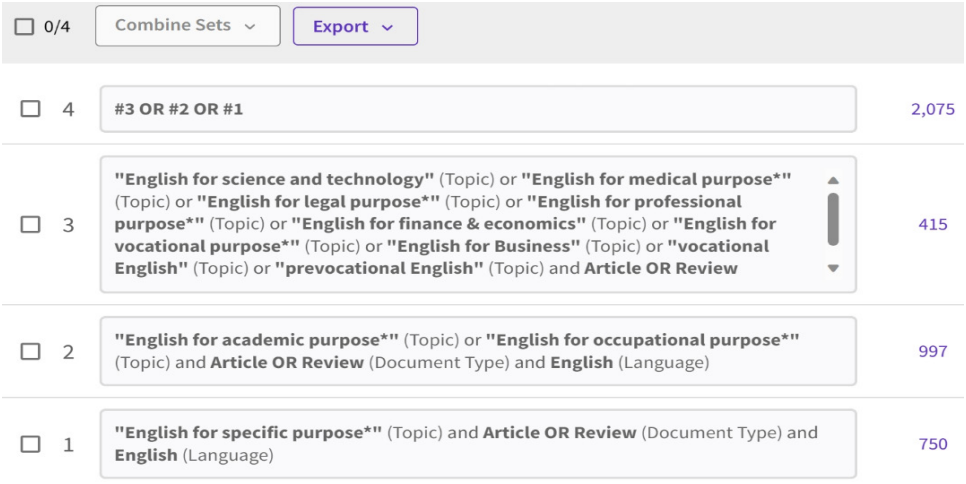
ESP Classification



Note. From *"Developments in English for Specific Purposes: A Multi-Disciplinary Approach,"* by T. Dudley-Evans & M. J. St John, 1998, Cambridge University Press. Copyright 1998 by Cambridge University Press.

Figure 3

Topic Search Queries



round of screening, the de-duplication function in Citespace screened and excluded again 324 conference proceedings, book chapter, editorial materials, meeting abstract, book review, and duplicate publications. Only articles (1608) and reviews (49) were retained after removing the above-mentioned document types and duplicated literature (Diao et al., 2022). The dataset of 1657 articles or reviews was retained for BA. The modified PRISMA flow diagram is shown in Figure **4**.

Data Analysis

Using Citespace 6.1. R6 for BA, this study used four techniques, including author and document co-citation analysis, co-word analysis, and cluster analysis (Chen, 2017).

First, the study examined the current state of ESP research by analyzing the distribution of publications, focusing on temporal patterns. Additionally, author co-citation analysis was employed to understand collaboration among researchers, specifically concentrating on co-country networks (Zhong et al., 2019). Second, the co-word analysis, specifically keyword clusters, was conducted to identify prominent thematic clusters. This was followed by an in-depth review of related articles to categorize the clusters and pinpoint major themes. Third, timezone view of keyword co-occurrence network (a visualization of co-word analysis) was generated to analyze the knowledge evolution of ESP research. Last, burst detection analysis was performed on keywords (co-word analysis) and cited references (document co-citation analysis) to identify emerging trends. A thorough review of articles citing burst keywords and references was undertaken to gain deeper insights into these trends.

RESULTS

This section presents the results of knowledge networks from the perspectives of current status, major themes, knowledge evolution, and emerging trends of ESP research.

Publication Distribution

The publication and geographic distribution of research in a specific field can facilitate understanding the research status of a specific field (Diao et al., 2022). Figure **5** shows the annual publications of ESP literature based on the dataset of 1657 records.

The solid line in Figure **5** shows the numbers of annual publications. Evidently, the ESP research is undergoing a steady growth trend in publications (Diao et al., 2022). The annual number of articles during this phase exceeded 60, and kept rising exponentially to 150 or so in the last three years (2020-2022). While 2023 saw a decline in the number of publications, research within this field remains ongoing.

Co-Country Network

To explore the geographic distribution of ESP research during the period from 2011 to 2023, co-country network was generated in Figure **6**.

As shown in Figure **6**, Citespace identified 95 countries and regions, indicating a growing global interest in ESP research among international scholars over the last decade. The most productive countries in ESP research include People's Republic of China (215 articles), USA (165 articles), England (127 articles), Spain (113 articles), Canada (100 articles), Australia (91), Iran (89), Ukraine (77), Taiwan (65), and Malaysia (64).

Figure 4

Modified PRISMA Flow Diagram

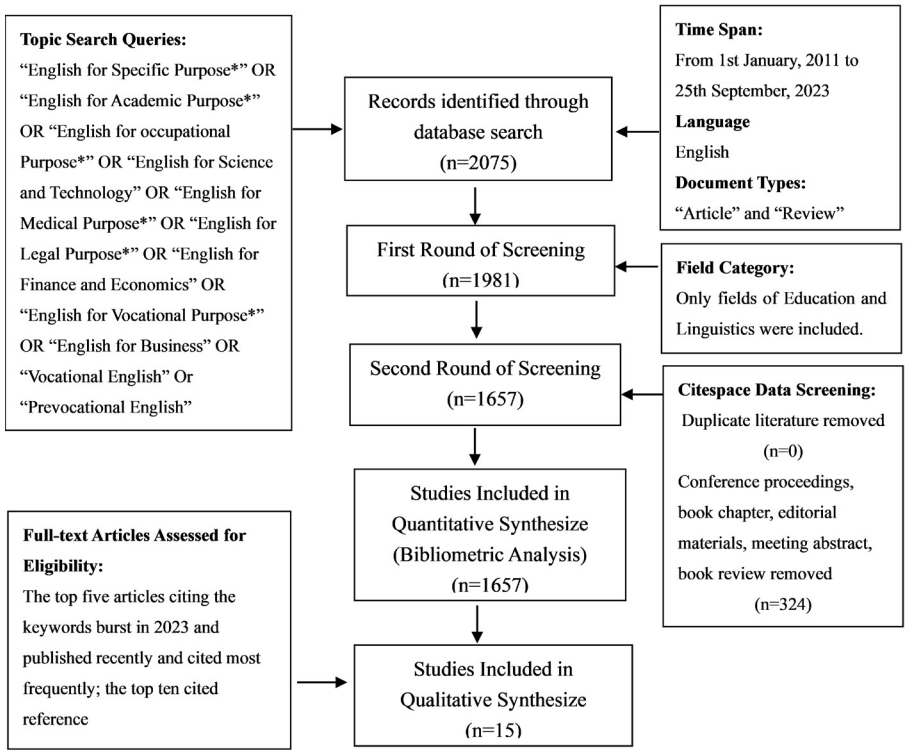
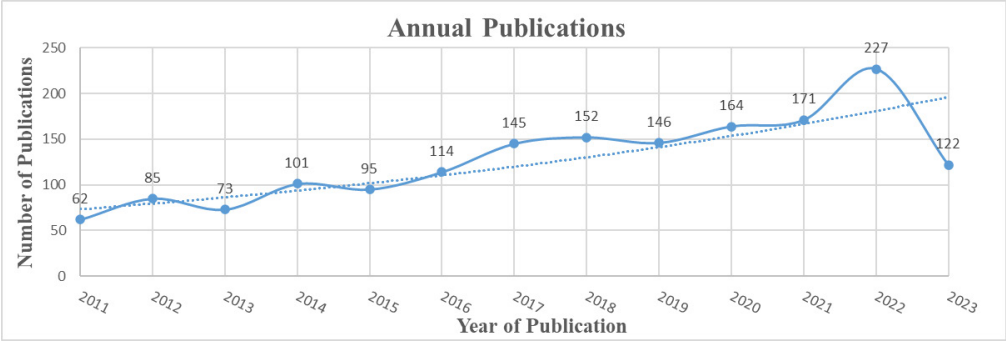


Figure 5

Annual Publications on ESP (2011-2023)



Keyword Clustering

Keywords represent research content, and those appearing frequently in literature pervasive topics within a specific field during a particular period (Wang et al., 2020). Chen (2017) highlighted that keywords analysis in pinpointing research trends and hotspots. To gain a better understanding of the research focus of ESP, keywords with frequency higher than 20 are presented in Table 1. Frequency refers to the number of times of citation in literature. Keywords with high frequency in the literature reflect the research hotspots in a specific field (Jia & Feng, 2022; Wang et al., 2020). High-frequency keywords of ESP research include "English for academic purposes", "students", "higher education", "academic writing", "needs analysis", "genre analysis", etc.

To understand the hidden structure of these keywords, cluster analysis and systematic review were conducted. The keyword-based clusters classify significant terms and contexts within a research field, uncover patterns and connections in the vast literature, and help identify underlying rules governing research themes (Diao et al., 2022). The clustering function generated 18 clusters. The cluster network displayed high modularity (0.7734) and a mean silhouette score of 0.9084, indicating significant homogeneity among cluster members and a highly rational clustering arrangement (Guo et al., 2022). By examining keywords in clusters

Figure 6

Co-Country Network of ESP Research (2011-2023)

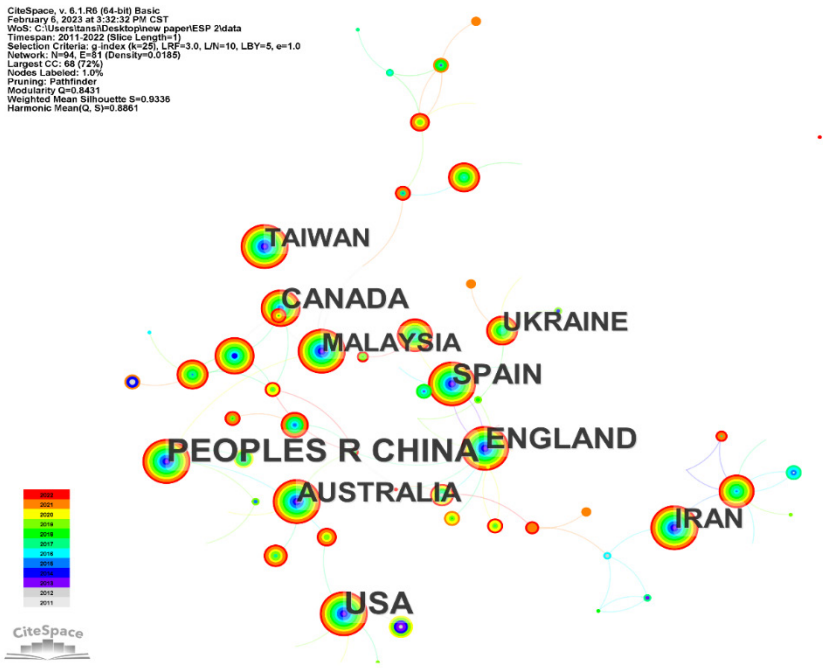


Table 1

High-Frequency Keywords of ESP Research (2011-2023)

No.	Keywords	Frequency	No.	Keywords	Frequency
1	English for specific purposes	248	16	applied linguistics	38
2	English for academic purposes	237	17	instruction	37
3	English	229	18	lexical bundles	35
4	language	201	19	corpus linguistics	34
5	students	131	20	teachers	33
6	higher education	109	21	performance	31
7	education	76	22	research articles	31
8	academic writing	75	23	needs	27
9	needs analysis	60	24	corpus	26
10	learners	59	25	proficiency	24
11	genre analysis	48	26	attitudes	24
12	knowledge	44	27	motivation	24
13	literacy	37	28	strategy	22
14	perceptions	35	29	genre	22
15	discourse	35	30	blended learning	21

and labeling them with the most indicative keyword, Figure 7 illustrates 13 clusters, each containing over 20 papers.

However, the cluster analysis in ESP research revealed overlapping cluster labels. Therefore, systematic review of keywords and abstracts in the "Node Details" were conducted, which provided citation information, and offered more insights into articles using the same keywords and those within the clusters (Jia & Harji, 2023). For instance, Cluster #0 "professional development" and Cluster #3 "teacher education" were combined due to their overlapping content. The same combination was applied to Cluster #1 "genre analysis" and Cluster #9 "discourse analysis". Cluster #5 "EAP"

was removed because it is a branch of ESP, which is the topic of this paper and does not represent the research themes.

Ultimately, four major themes emerged, including "linguistic inquiries", "teaching pedagogy", "student learning", and "teacher development". These four themes, along with their respective representative keywords, are presented in Figure 8.

Timezone View

The development of a research field can be identified through the changes of highly cited keywords over time (Diao et al., 2022). Timezone view of keyword co-occurrence network displays the evolution and development of keywords, and visualizes the research frontiers and the research trends of a certain field in different periods (Jiang et al., 2022). It is through the timezone view that we can see the time span and knowledge evolution of ESP research, as presented in Figure 9.

The figure shows the keywords that have appeared most frequently each year. In 2011, ESP scholars conducted research from a wide variety of research perspectives, such as academic writing, NA, genre analysis, lexical bundle, corpus linguistics, academic literacy, BE, etc. These perspectives represent all the above-mentioned themes except the theme "teacher". In 2012 and 2013, in addition to the conventional topics of academic English, academic vocabulary, discourse, need, new topics of blended learning and teacher began to gain wider attention.

The years 2014 and 2015 witnessed some new research interest in identity, students' English proficiency, ME, and case study. The conventional topics of word list, discourse, language need continued to remain the research focus. From 2016 to 2017, more focus had been on the themes of "stu-

Figure 7

Keyword Clusters of ESP Research (2011-2023)

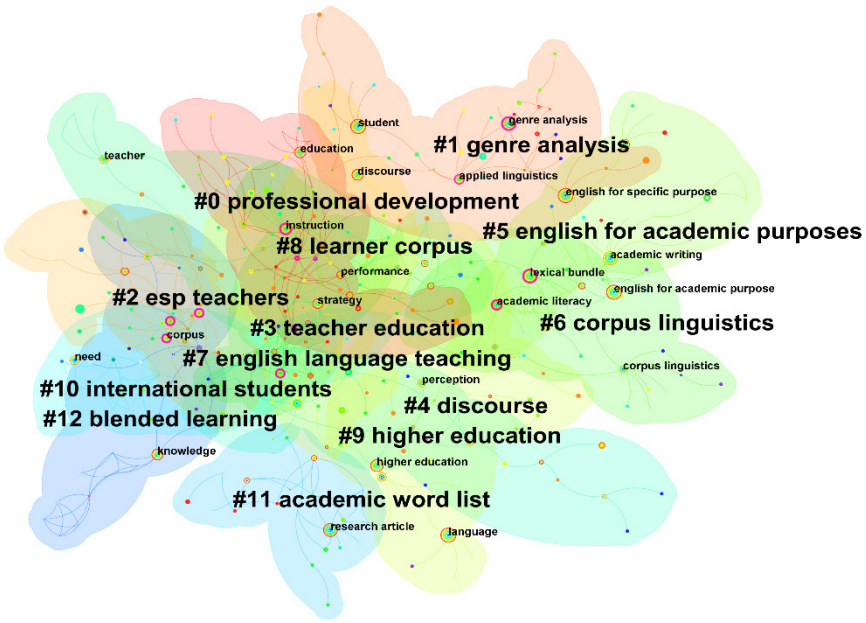


Figure 8

Major Themes of ESP Research (2011-2023)

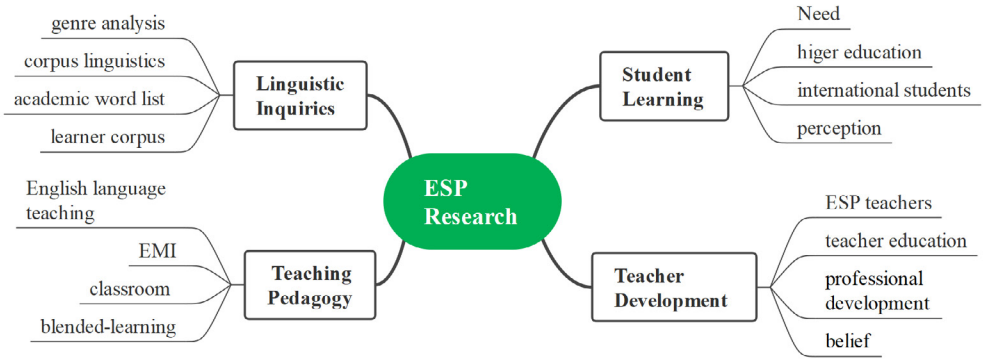
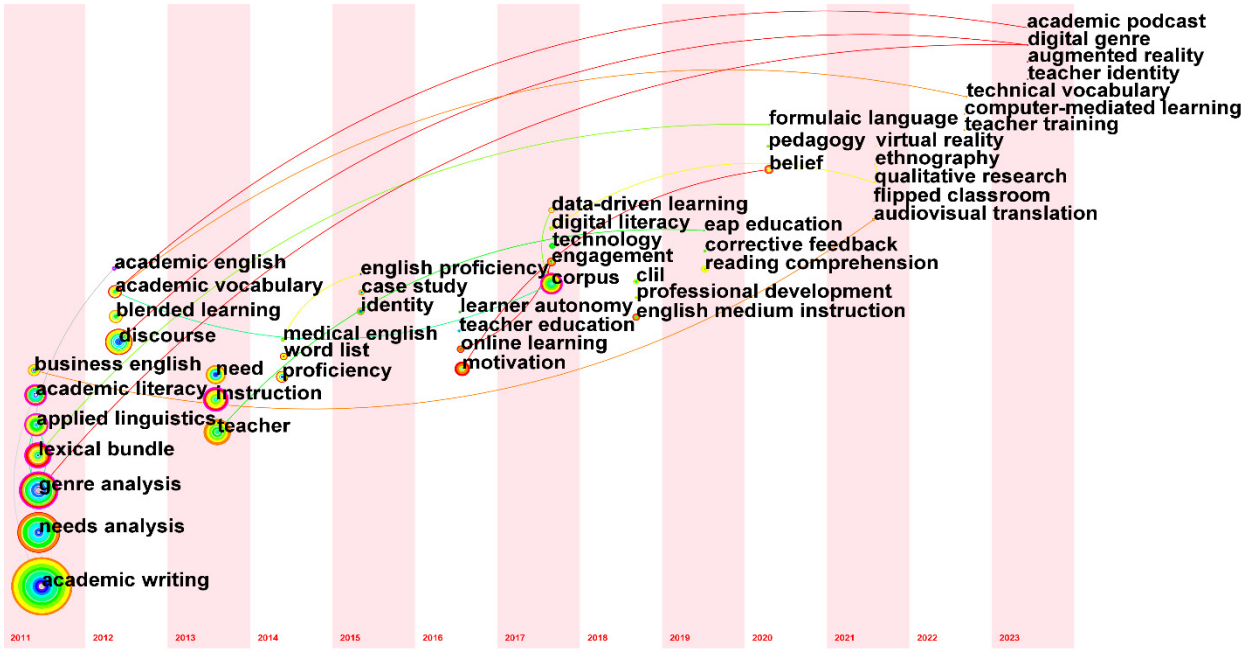


Figure 9

Knowledge Evolution of ESP Research (2011-2023)



dent" and "teacher", including keywords of "learner autonomy", "teacher education", "motivation", "engagement", etc. Some new trends related to technology, such as "online learning", "technology", "digital literacy", and "data-driven learning", emerged as well. Between 2018 and 2020, research topics centred around the themes "pedagogy" and "genre", which can be seen from keywords "content and language integrated learning (CLIL)", "English medium instruction (EMI)", "genre knowledge", "formulaic language", etc.

Finally, 2021 and 2023 saw the emergence of more technological terms including "virtual reality (VR)", "flipped classroom", "academic podcast", "digital genre", and "computer-mediated learning". "Teacher identity" and "teacher training" also received some attention.

Burst Detection

Analyzing research trends in a specific discipline provides valuable insights into future research directions. Burst detection in Citespace detects sudden increases in keywords or citations and unveils the research trends (Diao et al., 2022; Wang et al., 2020). Both the burst keywords and cited articles serve as crucial indicators of emerging trends in research (Zhou et al., 2019). Figure **10** displays the top 10 burst keywords. "Year" denotes the first occurrence of keywords, while "Strength" represents the burst strength value. "Begin" and "End" indicate the start and end points of the burst period. The bar visualizes the duration of burst.

As presented, the keywords with highest burst strengths were "research articles", "EAP", and "corpus linguistics", indicating a heated academic attention on EAP, especially research articles, and corpus analysis. However, current emerging patterns can be identified through the burst emergence and duration of keywords "teachers", "belief", and "curriculum", which started to burst in 2019, and continue to be influential till 2023. To facilitate the understanding of the trends, the major articles citing these three keywords that were published most recently and cited extensively, were critically examined (Chen, 2017). The major articles are presented in Table 2.

The keywords "teachers" and "belief" pertain to teachers' teaching methods and perceptions regarding their roles, competence, and challenges in ESP or EAP teaching. Bahrami et al. (2019) explored the current status of Iranian EAP teachers' involvement in both reading and conducting research, their EAP teaching careers, research engagement, and personal and institutional barriers in their academic commitment. Afshar and Ranjbar (2021) explored the level of assessment literacy among EAP teachers, including their beliefs towards assessment, their assessment methods in classrooms, and the underlying reasons for any discrepancies between their beliefs and practices. Tai and Tang (2021) explored teachers' teaching methods in mitigating students' adverse reactions to English-Medium Instruction (EMI) course, and examined whether students could alleviate their anxiety in by employing diverse learning strategies and minimizing their avoidance of English-Medium Instruction (EMI)

Figure 10

Keyword Burst of ESP Research (2011-2023)

Table 2

Major Citing Articles of Detected Keywords

Article	Authors
Exploring research-informed practice in English for academic purposes: A narrative study	Bahrami et al. (2019)
EAP teachers' assessment literacy: From theory to practice	Afshar and Ranjbar (2021)
The role of graduate students' learning strategies in reducing their English medium instruction avoidance: The mediation effect of language anxiety	Tai and Tang (2021)
Challenges and constraints in the design of an ESP course in Cambodia: implications for higher education institutions	Petraki and Khat (2022)
Needs analysis-based design principles for constructing a context-aware English learning system	Chiu et al. (2023)

course. These studies focused on ESP teacher development or EMI.

The keyword "curriculum" is related to ESP course development. Petraki and Khat (2022) examined the perspectives of academic and industry stakeholders regarding challenges and necessary measures for developing an ESP course in the Science, Technology, Engineering, and Mathematics (STEM) field at a Cambodian tertiary institution. Chiu et al. (2023) conducted a needs analysis for developing a system, materials, and activities for context-aware ubiquitous language learning and English for Fitness Purposes (EFP) in a fitness center as part of a design-based research project. Their foci of interest were on using needs analysis to develop ESP courses in different disciplines or social context.

To attain a profound comprehension of emerging trends, the burst detection of citations was conducted subsequently. Table 3 show the top ten cited reference.

As shown in Table 3, the books of Lillis and Curry (2010), and Nesi and Garders (2012), have the highest strengths. Lillis and Curry (2010) targeted at postgraduates and professionals in English language teaching, applied linguistics, etc., discusses how the dominance of English affects the practices of multilingual researchers in their academic writing and evaluation for publication. Nesi and Garders (2012) introduced the concepts of genre and genre family in the context of tertiary education, and the norms and conventions of the academic writing across different disciplines. Both works focus on academic writing. Research from 2012 to 2019 witnessed special academic attention paid to genre knowledge (Tardy, 2009), vocabulary learning (Nation, 2013), academic vocabulary (Gardner & Davies, 2014), and academic literacy (Wingate, 2015).

The recent emerging trends can be seen from the studies by Campion (2016), Macaro et al. (2018), Coxhead and Demecheleer (2018), and Ding and Bruce (2017), whose citation bursts last till 2022 or 2023. Campion (2016) explored teachers' views on their transition from General English to EAP. Macaro et al. (2018) conducted a systematic review of EMI in Higher Education. Based on corpus studies and methodology, Coxhead and Demecheleer (2018) investigated the technical vocabulary in the pluming field. In their book, Ding and Bruce (2017) explore previously overlooked aspects in EAP teachers' development, including the socio-economic, academic, and employment environments in which EAP prac-

Table 3

Citation Burst Detection of ESP Research (2011-2023)

References	Authors	Strength	Burst Duration
Academic writing in global context	Lillis and Curry (2010)	5.15	2011-2014
Building genre knowledge	Tardy (2009)	4.59	2012-2013
Genres across the disciplines: Student writing in higher education	Nesi and Garders (2012)	5.13	2014-2017
Learning vocabulary in another language	Nation (2013)	4.35	2015-2016
A new academic vocabulary list	Gardener and Davies (2014)	4.56	2016-2019
Academic literacy and student diversity: the case for inclusive practice	Wingate (2015)	3.99	2017-2019
"The learning never ends": Exploring teachers views on the transition from General English to EAP	Campion (2016)	4.77	2017-2022
A Systematic Review of English Medium Instruction in Higher Education	Macaro et al. (2018)	4.53	2020-2022
The English for academic purposes practitioner	Ding and Bruce (2017)	5.26	2019-2023
Investigating the technical vocabulary of plumbing	Coxhead and Demecheleer (2018)	4.87	2021-2023

titioners operate. By examining these contexts, they gain insights into the roles, expectations, and limitations faced by practitioners. Their research implies that ESP teacher development, EMI, and academic vocabulary, corpus-base teaching will be increasingly explored in the future.

DISCUSSION

The quantitative and objective review demonstrate the liveliness and diversity of ESP research from 2011 to 2023.

Current Research Status

As seen in Figures **5** and **6**, the field of ESP is continuously evolving and far from reaching saturation. As Johns predicted (2012), the period spanning from 2011 to 2013 witnessed a significant surge in international authorship, implying the global academic interest in ESP research during this timeframe.

Major Themes

As presented in Figure **8**, four major themes were found from the clustering analysis and systematic review. The first theme "linguistic inquiries" refers to linguistic research to describe how English is utilized in academic, study, and professional environments within the local community. It mainly includes genre analysis and corpus analysis. Ren and Lu (2021) compared narratives in corporate annual reports from American and Chinese companies, and suggested to incorporate the language variations in the genre into Business English teaching in China. Chan (2021) identified the challenges faced by English major graduates in Hong Kong to workplace communication in their unfamiliar fields, and their strategies to overcome these challenges. Otto (2021) employed a corpus-based strategy to pinpoint 18 frequently used terms in civil engineering writing.

The second theme, "teaching pedagogy", encompasses the practical aspects of teaching, such as teaching approach, innovations, and the evaluation of teaching and materials. Classroom-based research has gained more prominence during this period. Belmekki and Sekkal (2018) integrated process-genre approach into BE writing. They investigated how process-genre approach affected ESP students' writing achievement in terms of request letters as well as their attitude towards this approach, and found that process-genre approach had a significantly positive effect on students' writing achievement, including organization, vocabulary, grammar, and mechanics. Numerous studies have explored other teaching approaches, including CLIL (Wozniak, 2019; Yang, 2018), task-based and project-based learning (Hafner et al., 2017), and corpus-based teaching (Otto, 2021).

The third theme "student learning", concentrates on different learning variables and NA. ESP research mainly focuses on ESP learning of students in higher education (Liontou, 2021; Rose et al., 2020). Rose et al. (2020) examined the relationship between course performance, English proficiency, motivation, and academic language skills of students of a business program at a Japanese university. Liontou (2021) investigated Finnish and Chinese students' perceptions of language assessment for an ESP course in some university programs offered in Finland and China. NA is a core characteristic of ESP and EAP courses and an integral part of curriculum development (Chan, 2018). Since NA is a prerequisite for any ESP course, students' needs have been looked into

from different perspectives, such as learning motivation (Thepseenu, 2020), teaching approach (Jafari & Alemi, 2020), and teaching materials or textbooks (Guo & He, 2020).

The fourth theme, "teacher development", specifically pertains to the professional training and growth of ESP teachers. Ding and Bruce (2017) examined the role and development of EAP practitioners, and evaluated their professional and academic identities. Bahrami et al. (2019) highlighted the challenges faced by EAP teachers from diverse disciplinary backgrounds in conducting research practice due to personal and institutional barriers in Irian. Afshar and Ranjbar (2021) examined the assessment literacy of EAP teachers. However, research on EAP instructors still remains limited (Du et al., 2022), which was further discussed in emerging trends.

Knowledge Evolution

As presented in Figure 9, there are four features to the research evolution of ESP from 2011 to 2023. First, the linguistic inquiries, such as "genre", "corpus", and "discourse", is still an enduring research focus during this period. These topics, as the central concepts in ESP research ever since 1960s (Johns, 2012), are overlapping with each other, and have yet to be fully explored. As one persuasive topic of ESP research, linguistic inquiry could provide insights in linguistic practices in academic, professional or workplace contexts (Basturkmen, 2021). Le and Miller (2023) concentrated on the vocabulary required by medical students, while Bi (2020) focused on the vocabulary necessary for Chinese undergraduate students majoring in computer science to comprehend their specialized textbooks.

Second, varied pedagogies have been attempted, such as CLIL (Wozniak, 2019; Yang, 2018), task-based and project-based learning (Hafner et al., 2017), corpus-based teaching (Otto, 2021), etc. As a relatively new topic compared with genre and corpus, CLIL approach, which integrates language skills and content knowledge, has been widely adopted and researched internationally, Yang (2018) examined the satisfaction of students in a polytechnic university for their CLIL teaching materials, and the factors affecting their satisfaction. Wozniak (2019) analyzed the feedback of pre-service content teachers in French secondary schools after a period of training of teaching in the way of CLIL. Several difficulties were identified and suggestion were offered for improvement. Like all language education, classroom-based research will continue to gain prominence in ESP research.

Third, varied research methodologies have been adopted, which can be seen from the emergence of keywords including "case study", "qualitative case study", etc. Combining both qualitative and quantitative approaches, more case and ethnographic studies of learners in different learning and cultural contexts have been conducted to enhance ESP teaching and learning (Liu et al., 2021; Mortenson, 2022; Šelmić, 2021). Liu et al. (2021) focused on two groups of students in an EAP reading course at a Chinese university, and identified their reading strategies for academic texts. Mortenson (2022) integrated social justice-oriented content into an EAP writing course at a university on the East coast of America, and identified the barriers to the emergent bilingual students learning this course.

Fourth, technology has been implemented more frequently in ESP teaching and research recently, which can be implied from keywords such as "online learning" (Mihai et al., 2022), "blended learning" (Simonova, 2019), "flipped classroom" (Hsiao et al., 2021), "Virtual reality" (Chen et al., 2021; Hsu et al., 2022), etc. Technology has long played an important role in ESP instruction, and can be applied in different types of ESP courses. Chen et al. (2021) integrated VR technology into an ESP course which enabled students with multimodal information, and pointed out the importance of engaging students in immersive environments. Hsu et al. (2022) developed a VR platform for an ESP course and examined the relationship between self-efficacy and behavioural intention, and recommended that ESP teachers promote students' self-efficacy when integrating VR technology in their courses. With the inception of each new technology, new challenges and opportunities will be posed for both ESP learners and practitioners in their learning and teaching process, indicating the continuing predominance of technology in ESP research.

Emerging Trends

Keywords burst and citation burst results show that EAP, ESP teacher development, NA in under-researched settings, as well as EMI and Corpus-based teaching pedagogies, are new ESP research frontiers.

Firstly, the prevalence of EAP is revealed by a significant increase in published EAP research, especially in Asian countries (Hyland & Jiang, 2021). There is a notable academic emphasis on vocabulary learning (Nation, 2013), academic vocabulary (Gardner & Davies, 2014), academic literacy (Wingate, 2015), and lexical bundle (Liu & Chen, 2020).

As another prevalent topic, ESP teacher development is continuously drawing enormous academic attention. Tao and Gao (2018), and Mahendra (2020), both examined the challenges faced by ESP teachers in different social context when constructing their professional identity. Even more research was conducted to explore the EAP teachers' teaching beliefs, literacy, development of (Afshar & Ranjbar, 2021; Bahrami et al., 2019; Du et al., 2022). This corroborates Basturkmens' findings (2022), who emphasized that ESP teachers need to acquire new knowledge and redefine their professional identities. However, they face limited access to adequate support for their professional development and are often

marginalized in their roles. Therefore, the research into their development is still in need.

Furthermore, NA has been a prominent focus in ESP literature to date. The recent research demonstrates a robust representation of this research trend. This finding was consistent with the conclusions made by Basturkmen (2022) and Dou et al. (2023), that NA remains a core and enduring characteristic in this field. The recent studies apper to focus on the language needs of stakeholders coming from various cultural and linguistic backgrounds in under-research setting, such as STEM course (Petraki & Khat, 2022), EFP course (Chiu et al., 2023), and plumbing industry (Coxhead & Demecheleer, 2018).

Lastly, as discussed by Johns (2012), and Hyland and Jiang (2021), teaching practices were a focal point of the academic interest. Studies in recent years highlight the interest in implementing EMI or corpus-based teaching (Basturkmen, 2022). Macaro et al. (2018) conducted a systematic review of EMI in Higher Education. Some researchers explored the practices of EMI course and its impact on teaching in universities in different countries such as Indonesia (Simbolon, 2021), and Taiwan (Tai & Tang; 2021). Other researchers have utilized corpus analysis to create linguistic descriptions or word lists, incorporating them into their teaching methods. They drew on corpus-based methods and aimed to facilitate both ESP teaching and learning (Coxhead & Demecheleer, 2018; Liu & Chen, 2020; Otto, 2021).

CONCLUSION

This study employed bibliometric analysis via Citespace to provide a systematic, quantitative and objective overview of international ESP research during the period from 2011 to 2023. A dataset of 1657 bibliographic records within this timespan were analyzed. The analysis involved generating and examining various graphs and networks.

The publication distribution and co-country network were analyzed to understand the current research status. Evidently, ESP research has been experiencing steady growth in the past decade and is not yet saturated. Moreover, this research exhibits extensive international authorship and collaboration.

To identify the major themes of ESP research, keywords were used to generate clusters. The clustering results and systematic review suggested that researchers in ESP have focused on the four themes, including linguistic inquiries, teaching pedagogy, student learning, and teacher development.

Through the timezone view of keyword co-occurrence network, the knowledge evolution shows the research hot topics are highly dynamic and overlapping within the years from 2011 to 2023. Four features were identified, including the predominance of pedagogy, genre and corpus analysis, varied pedagogies, varied research methodologies, and more technology-assisted teaching.

Lastly, the results of keyword and citation burst detection, as well as systematic review of citing articles of burst keywords were examined to explore the emerging trends in ESP. The recent studies exhibit growing research interests in EAP, ESP teacher development, different learning needs in under-researched settings, as well as EMI and Corpus-based teaching pedagogies.

Admittedly, this study has several limitations. Firstly, this study used Citespace as the analysis tool, but different tools might yield divergent outcomes. Future research endeavors could explore alternative tools such as BibExcel, VOSviewer, or the Science of Science Tool to present the findings. Secondly, the most productive authors, institutions, and journals were not examined. Furthermore, only English-language articles sourced from the WOSCC database were analyzed, excluding regional journals in different languages. Researchers in ESP could explore localized contexts, or contrast local and international trends in the future.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS CONTRIBUTIONS

Sining Tan: Conceptualization, software validation, visualization, and writing-original draft, writing- reviewing and editing.

Madhubala Bava Harji: Conceptualization, writing- reviewing and editing.

Xiaogang Hu: Software validation, visualization, and writing- reviewing and editing.

REFERENCES

Afshar, H. S., & Ranjbar, N. (2021). EAP teachers' assessment literacy: From theory to practice. *Studies in Educational Evaluation*, *70*, 101042. https://doi.org/10.1016/j.stueduc.20 21.101042

Bahrami, V., Hosseini, M., & Atai, M. R. (2019). Exploring research-informed practice in English for academic purposes: A narrative study. *English for Specific Purposes*, *54*, 152-165. https://doi.org/ 10.1016/j.esp.2019.02.003

Basturkmen, H. (2021). ESP research directions: Enduring and emerging lines of inquiry. *Language Teaching Research Quarterly*, *23*, 5-11. https://doi.org/10.32038/ltrq.2021.23.02

Basturkmen, H. (2022). Current trends in ESP research in the Asia Pacific region. *World Englishes*, *41*(4), 512-522. https://doi.org/10.1111/weng.12601

Belcher, D. D. (2006). English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly*, *40*(1), 133-156. https://doi.org/ 10.2307/40264514

Belmekki, A., & Sekkal, F. (2018). The effect of process-genre approach on ESL students' achievement in writing business letters. *European Journal of Research and Reflection in Educational Sciences*, *6*(2), 10-14. http://dx.doi.org/10.21070/jees.v4i2.2598

Bi, J. (2020). How large a vocabulary do Chinese computer science undergraduates need to read English-medium specialist textbooks? *English for Specific Purposes*, *58*, 77-89. https://doi.org/10.1016/j.esp.2020.01.001

Campion, G. C. (2016). 'The learning never ends': Exploring teachers' views on the transition from General English to EAP. *Journal of English for Academic Purposes*, *23*, 59-70. http://dx. doi.org/10.1016/j.jeap.2016.06.003

Chan, C. S. (2018). Proposing and illustrating a research-informed approach to curriculum development for specific topics in business English. *English for Specific Purposes*, *52*, 27-46. https://doi.org/10.1016/j.esp.2018.07.001

Chan, C. S. (2021). University graduates' transition into the workplace: How they learn to use English for work and cope with language-related challenges. *System*, *100*, 102530. https://doi.org/10.1016/j.system.2021.102530

Chen, C. (2006). "CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature." *Journal of the American Society of Information Science and Technology*, *57*(3), 359–377. https://doi.org/10.1002/asi.20317

Chen, C. (2013). *Mapping scientific frontiers: The quest for knowledge visualization*. Springer Science & Business Media.

Chen, C. (2017). Science mapping: a systematic review of the literature. *Journal of Data and Information Science*, *2*(2), 1-40. https://doi.org/10.1002/asi.20317

Chen, C. H., Hung, H. T., & Yeh, H. C. (2021). Virtual reality in problem-based learning contexts: Effects on the problem-solving performance, vocabulary acquisition and motivation of English language learners. *Journal of Computer Assisted Learning*, *37*(3), 851-860. http://dx.doi.org/ 10.1111/jcal.12 528

Chen, X., & Liu, Y. (2020). Visualization analysis of high-speed railway research based on CiteSpace. *Transport Policy*, *85*, 1-17. https://doi.org/10.1016/j.tranpol.2019.10.004

Cheng, A. (2011). ESP classroom research: Basic consideration and future research questions. In D. Belcher, A. M. Johns, & B. Paltridge (Eds.), *New directions in English for specific purposes* (pp. 44–72). University of Michigan Press.

Chiu, W. Y., Liu, G. Z., Barrett, N. E., Liaw, M. L., Hwang, G. J., & Lin, C. C. (2023). Needs analysis-based design principles for constructing a context-aware English learning system. *Computer Assisted Language Learning*, *36*(1-2), 176-204. https://doi.org/10.1080/ 09588221.2021.1905666

Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes, 51, 84-97.*

Diao, B. C., Adnan, N., Harji, M. B., & Ravindran, L. (2022). Evolution and hotspots of peer instruction: a visualized analysis using CiteSpace. *Education and Information Technologies*, *28*(2),1-18. https://doi.org/10.1007/s10639-022-11218-x

Ding, A., & Bruce, I. (2017). *The English for academic purposes practitioner*. Palgrave Macmillan.

Dou, A. Q., Chan, S. H., & Win, M. T. (2023). Changing visions in ESP development and teaching: Past, present, and future vistas. *Frontiers in Psychology*, *14*, 1140659. https://doi.org/10.3389/fpsyg.2023.1140659

Du, J., Li, W., & Li, Q. (2022). Who Are We, What Can We Do, and What Do We Think? Review on EAP Teacher Development. *Chinese Journal of Applied Linguistics*, *45*(4), 532-550. http://dx. doi.org/10.1515/CJAL-2022-0403

Dudley-Evans, T. & St John, M. J. (1998). *Developments in English for Specific Purposes: A Multi - Disciplinary Approach*. Cambridge University Press.

Fu, R. D., Yang, D. G., Jin, C. F., Cai, T. Y., Wu, S. Q., and Wang T. (2021). Research progress and review on spatial differentiation of urban housing price: a bibliometric analysis using CiteSpace. *Journal of University of Chinese Academy of Sciences*, *38*(6), 782-790. http://dx.doi. org/10.7523/j.issn.209 5-6134.2021.06.008

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305-327. https://doi.org/10.1093/applin/amt015

Guo, Y., Xu, Z. Y. R., Cai, M. T., Gong, W. X., & Shen, C. H. (2022). Epilepsy with suicide: A bibliometrics study and visualization analysis via CiteSpace. *Frontiers in Neurology*, *12*, 823474. https://doi.org/10.3389/fneur.2021.823474

Guo, G., & He, B. (2020). Evaluation of business English writing course-book based on the needs analysis model. *International Linguistics Research*, *3*(3), 11-21. https://doi.org/10.30560 /ilr.v3n3p11

Hafner, C. A., Miller, L., & Ng, C. K. F. (2017). Creativity and digital literacies in English for specific purposes. *Essential Competencies for English-medium University Teaching*, 111-123. http://dx.doi. org/10.1007/978-3-319-40956-6_8

Hsiao, I. C. V., Hung, S. T. A., & Huang, H. T. D. (2021). The flipped classroom approach in an English for specific purposes (ESP) course: A quasi-experimental study on learners' self-efficacy, study process, and learning performances. *Journal of Research on Technology in Education*, *55*(3), 1-20. http://dx.doi.org/10.1080/15391523.2021.1976329

Hsu, C. C., Chen, Y. L., Lin, C. Y., & Lien, W. C. (2022). Cognitive development, self-efficacy, and wearable technology use in a virtual reality language learning environment: A structural equation modeling analysis. *Current Psychology*, *41*(3), 1-15. http://dx.doi.org/10.1007/s12144-021-02252-y

Hu, F. (2019). Visualization analysis of research of English for specific purposes in China based on citespace (2000-2017). *Theory and Practice in Language Studies*, *9*(9), 1183-1190. http://dx.doi.org/10.17507/tpls.0909.15

Huebner, T. (2015). Academic writing for graduate students: Essential Tasks and skills. J. M. Swales & Ch. B. Feak. 2012. Ann Arbor, MI: University of Michigan Press (418 pages). *Journal of Humanities*, *18*(2), 113-115. http://dx.doi.org/10.1163/26659077-01802006

Hutchinson, T., & Waters, A. (1987). *English for Specific Purposes*. Cambridge University Press.

Hwang, G. J., & Tsai, C. C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, *42*(4), E65–E70. https://doi.org/10.1111/j.1467-8535.2011.01183.x

Hyland, K. (2012). ESP and writing. *The handbook of English for specific purposes* (pp. 95-113). https://doi.org/10.1002/9781118339855.ch5

Hyland, K., & Jiang, F. K. (2021). A bibliometric study of EAP research: Who is doing what, where and when? *Journal of English for Academic Purposes*, *49*, 100929. https://doi.org/10.1016/j.jeap.2020.100929

Jafari, P. S., & Alemi, M. (2020). Engineering students' motivation to learn technical English in ESP courses: Investigating Iranian teachers' and students' perceptions. *RELC Journal*, *51*(2), 212-226. https://doi.org/10.1177/0033688218811371

Jia, S., & Harji, M. B. (2022). Systematic *review of mobile-assisted task-based learning based* on WoS (2013-2022). *Journal of Information Technology Education: Research*, *21*, 571-595. https://doi.org/10.28945/5034

Jia, S. J., & Harji, M. B. (2023). Themes, knowledge evolution, and emerging trends in task-based teaching and learning: A scientometric analysis in CiteSpace. *Education and Information Technologies*, *28*, 1-20. http://dx.doi.org/10.1007/s10639-023-11586-y

Jia, X. L., & Feng, Y. (2022). A knowledge graph analysis of cultural heritage research in China in past 30 years. *Journal of North-West Minority University of Philosophy, Society and Science, 6*, 124–135.

Jiang, W., Aishan, T., Halik, Ü., Wei, Z., & Wumaier, M. (2022). A Bibliometric and visualized analysis of research progress and trends on decay and cavity trees in forest ecosystem over 20 Years: An application of the CiteSpace Software. *Forests*, *13*(9), 1437. https://doi.org/ 10.3390/f13091437

Johns, A. M. (2012). The history of English for specific purposes research. *The handbook of English for specific purposes*, *107*, 5-30. https://doi.org/10.1002/9781118339855.ch1

Karimnia, A., & Khodashenas, M. R. (2018). Medical students' English language learning needs and perceptions. *Sustainable Multilingualism*, *13*, 168-195. https://doi.org/10.2478/sm-2018-0016

Le, C. N. N., & Miller, J. (2023). A core meaning-based analysis of English semi-technical vocabulary in the medical field. *English for Specific Purposes*, *70*, 252-266. https://doi.org/10.1016/j.esp.2023.01.006

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for academic purposes*, *22*, 42-53. http://dx.doi.org/10.1016/ j.jeap.2016.01.008

Lillis, T. M., & Curry, M. J. (2010). *Academic writing in global context*. Routledge.

Liontou, M. (2021). Conceptions of assessment as an integral part of language learning: A case study of Finnish and Chinese university students. *Languages*, *6*(4), 202. https://doi.org/10.3390/lang uages6040202

Liu, C. Y., & Chen, H. J. H. (2020). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes*, *58*, 122-137. https://doi.org/10.1016/j.esp.2019.12.003

Liu, X., Gu, M. M., & Jin, T. (2021). Strategy use in collaborative academic reading: Understanding how undergraduate students co-construct comprehension of academic texts. *Language Teaching Research*, *0*(0). https://doi.org/10.1177/13621688211025688

Liu, Y., & Hu, G. (2021). Mapping the field of English for specific purposes (1980–2018): A co-citation analysis. *English for Specific Purposes*, *61*, 97-116. https://doi.org/10.1016/j.esp. 2020.10.003

Macaro, E., Curle, S., & Pun, J. Jiangshan An, and Julie Dearden. 2018. *A Systematic Review of English Medium Instruction in Higher Education. Language Teaching*, *51*(1), 36-76. http://dx.doi.org/10.1017/S0261444817000350

Mahendra, A. W. (2020). Constructing identity: Experiences of Indonesian ESP teachers in a language institute. *English Language Teaching Educational Journal*, *3*(3), 229-240. https://doi.org/10.12928/eltej.v3i3.2560

Mihai, M., Albert, C. N., Mihai, V. C., & Dumitras, D. E. (2022). Emotional and social engagement in the English language classroom for higher education students in the COVID-19 online context. *Sustainability*, *14*(8), 4527. http://dx.doi.org/10.3390/su14084527

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, *70*(2), 151-167. https://doi.org/10.1080/00335638409383686

Morrow, P. R. (1989). Conjunct use in business news stories and academic journal articles: A comparative study. *English for Specific Purposes*, *8*(3), 239-254. https://doi.org/10.1016/0889-4906(89)90015-X

Mortenson, L. (2022). Integrating social justice-oriented content into English for Academic Purposes (EAP) instruction: A case study. *English for Specific Purposes*, *65*, 1-14. https://doio.or/10. 1016/j.esp.2021.08.002

Nation. I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education.* Cambridge University Press.

Otto, P. (2021). Choosing specialized vocabulary to teach with data-driven learning: An example from civil engineering. *English for Specific Purposes*, *61*, 32-46. https://doi.org/10.1016/j.esp.2020.08. 003

Pan, J. F., Li, Z. F., & Xu, D. X. (2022). Constructing a diversified curriculum for college English in medical schools based on student needs analysis. *Journal of Higher Education*, *1*, 83-86+91. https://doi.org/10.19980/j.CN23-1593/G4.2022.01.020

Petraki, E., & Khat, K. (2022). Challenges and constraints in the design of an ESP course in Cambodia: implications for higher education institutions. *Asia Pacific Journal of Education*, *42*(2), 260-275. https://doi.org/10.1080/02188791.2020.1798738

Pratiwi, Y. E., Rukmini, W., & Faridi, A. (2017). The linguistic problems of students' competence in writing Business English letter. *Language Circle, 11*(2), 117-126. https://doi.org/10.15294/lc. v11i2.9584

Ren, C., & Lu, X. (2021). A multi-dimensional analysis of the management's discussion and analysis narratives in Chinese and American corporate annual reports. *English for Specific Purposes*, *62*, 84-99. https://doi.org/10.1016/j.esp.2020.12.004

Rose, H., Curle, S., Aizawa, I., & Thompson, G. (2020). What drives success in English medium taught courses? The interplay between language proficiency, academic skills, and motivation. *Studies in Higher Education*, *45*(11), 2149-2161. https://doi.org/10.1080/03075079.2019.1590690

Sabe, M., Chen, C., Perez, N., Solmi, M., Mucci, A., Galderisi, S., ... & Kaiser, S. (2022). Thirty years of research on negative symptoms of schizophrenia: A scientometric analysis of hotspots, bursts, and research trends. *Neuroscience & Biobehavioral Reviews*, *144*, 104979. https://doi.org/10.1016/j. neubiorev.2022.104979

Šelmić, N. (2021). Case Study Approach to Teaching and Learning English for Medical Purposes. *Journal of Teaching English for Specific and Academic Purposes*, *9*(3), 333-337. https://doi.org/10.22190/JTESAP2103333S

Si, J. (2020). An analysis of business English coursebooks from an ELF perspective. *ELT Journal*, *74*(2), 156-165. https://doi.org/10.1093/elt/ccz049

Simonova, I. (2019). Blended approach to learning and practicing English grammar with technical and foreign language university students: comparative study. *Journal of Computing in Higher Education*, *31*(2), 249-272. http://dx.doi.org/10.1007/s12528-019-09219-w

Simbolon, N. E. (2021). English Medium Instruction (EMI) practice: Higher education internationalization in Indonesia. *Journal of Language, Education, and Humanities*, *8*(2), 72-83. https://doi.org/10.22373/ej.v8i2.8961

Stubbs, M., & Halbe, D. (2012). *Corpus linguistics: Overview*. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1377–1379). Blackwell.

Sun, L., & Asmawi, A. (2022). Research Progress and Trend of Business English Writing Instruction in China (2002-2021): A Bibliometric Analysis. *World Journal of English Language, 12*(6), 240-254. https://doi.org/10.5430/wjel.v12n6p241

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential Tasks and Skills* (3rd ed.). University of Michigan Press.

Tai, T. E., & Tang, C. W. (2021). The role of graduate students' learning strategies in reducing their English medium instruction avoidance: The mediation effect of language anxiety. *Asia Pacific Journal of Education*, *41*(2), 368-384. https://doi.org/10.1080/02188791.2020.177793 7

Tao, J. T., & Gao, X. A. (2018). Identity constructions of ESP teachers in a Chinese university. *English for specific Purposes*, *49*, 1-13. https://doi.org/10.1016/j.esp.2017.09.003

Tardy, C. M. (2009). *Building genre knowledge*. Parlor Press.

Thepseenu, B. (2020). Needs analysis for ESP course development: Thai civil engineering students' perspectives. *Asian Journal of Education and Training, 6*(3), 433-442. https://doi.org/ 10.20448/journal. 522.2020.63.433.442

Tu, S. J., Jin, C., Chen, B. T., Xu, A. Y., Luo, C., & Wang, X. (2022). Study on the fusion of sports and medicine in China from 2012 to 2021: A bibliometric analysis via CiteSpace. *Frontiers in Public Health*, *10*, 3322. https://doi.org/10.3389/fpubh.2022.939557

Vesna, P. P. (2022). Classification of English for Specific Purposes and overview of the language skills in ESP context. *Beder Journal of Educational Sciences*, *26*, 117-132. http://eprints.uklo.edu.mk/ id/eprint/7143

Wang, G., Li, S., Zhang, Z., Hou, Y., & Shin, C. (2023). A visual knowledge map analysis of cross-border agri-food supply chain research based on CiteSpace. *Sustainability*, *15*(14), 10763. https://doi.org/10.3390/su151410763

Wang, Z., Ma, D., Pang, R., Xie, F., Zhang, J., & Sun, D. (2020). Research progress and development trend of social media big data (smbd): Knowledge mapping analysis based on Citespace. *ISPRS International Journal of Geo-Information*, *9*(11), 632. http://dx.doi.org/10.3390/ijgi9110632

Wingate, U. (2015). *Academic literacy and student diversity: the case for inclusive practice.* Multilingual Matters.

Wozniak, S. (2019). English for Academic Purposes and the enhancement of French trainee content teachers' professional skills. *Journal of Teaching English for Specific and Academic Purposes*, *6*(3), 341-352. http://dx.doi.org/10.22190/JTESAP18033 41W

Yang, W. (2018). Evaluating contextualized content and language integrated learning materials at tertiary level. *Latin American Journal of Content & Language Integrated Learning, 11*(2), 236-274. https://doi.org/10.5294/laclil.2018.11.2.4

Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. *Automation in Construction*, *101*, 17-31. https://doi.org/10.1016/j.autcon.2018.12.013

Zhou, W., Chen, J., & Huang, Y. (2019). Co-Citation analysis and burst detection on financial bubbles with scientometrics approach. *Economic research-Ekonomska istraživanja*, *32*(1), 2310-2328. https://doi.org/10.1080/1331677X.2019.1645716

https://doi.org/10.17323/jle.2023.12426

The Effectiveness Data-Driven Vocabulary Learning: Hands-on Concordancing through a Pedagogical Corpus

Sibel Tosun [1], **Hatice Sofu** [2]

[1] Firat University, Elazig, Turkey

[2] Cukurova University, Adana, Turkey

ABSTRACT

Introduction: Although extensive studies have been carried out on the effectiveness of corpora on teaching vocabulary, the exploration of whether learners can benefit from a pedagogical corpus, particularly regarding hands-on engagement by lower-level learners, has received little attention.

Purpose: To address this gap in the literature, this study sets out to explore the effectiveness of Data-Driven Learning (DDL) in enhancing the vocabulary acquisition of EFL students at a state university in Turkey through a pedagogical corpus.

Method: The quasi-experimental study employed a mixed-method research design, in which both quantitative and qualitative data were gathered through vocabulary tests, student questionnaires, and semi-structured interviews. Fifty-eight low-level students with an average age of 19 served as participants. The experimental group made use of hands-on concordancing while the control group received conventional course book-based instruction to learn the target words.

Results: The results indicate that pedagogical corpora have significant potential in facilitating vocabulary learning of low-level learners. The vocabulary tests revealed that the students who practiced with DDL outperformed the students who received traditional vocabulary instruction in both the post-test and the delayed post-test. The findings from student questionnaires, and semi-structured interviews also denoted that the participants held positive attitudes towards using concordancing to expand their vocabulary and grow aware of some aspects of words such as part of speech information, different meanings and usages, lexico-grammatical structures, and collocations.

Conclusion: The present study provides useful implications for collection and use of a pedagogical corpus for classroom use.

KEYWORDS

corpus-based instruction, data-driven learning (DDL), concordance lines, vocabulary learning, pedagogical corpus

Citation: Tosun S., & Sofu H. (2023). The Effectiveness Data-Driven Vocabulary Learning: Hands-on Concordancing through a Pedagogical Corpus. *Journal of Language and Education, 9*(3), 176-190. https://doi.org/10.17323/jle.2023.12426

Correspondence:
Sibel Tosun,
sibeltosun@outlook.com

Received: May 18, 2021
Accepted: September 15, 2023
Published: September 30, 2023

INTRODUCTION

Considering the impact corpora have had on language pedagogy, they are acknowledged to have "revolutionized" language teaching in various areas (Conrad, 2000, p. 549) both directly and indirectly (Römer, 2008). Dictionary making (e.g., Gouws, 2021; Hunston, 2002; O'Keefe et al., 2007), textbook and material development (e.g., Boulton, 2012; Friginal & Roberts, 2022; McEnery & Xiao, 2011), design of syllabi and testing materials (e.g., Boulton 2009; Hunston, 2002) including their validation and standardization (McEnery & Xiao, 2011) are some areas where corpora have indirectly contributed to language teaching. A resurgence of interest has also been witnessed in incorporating corpora into language teaching directly (Römer, 2008) through DDL, an approach to language learning introduced by Johns (1991).

A myriad of studies focused on the impact of the DDL and corpus use in L2 learning. In particular, recent meta-analyses (Boulton & Cobb 2017; Cobb & Boulton, 2015; Lee et al., 2019), have put forward the positive outcomes reported by corpus studies. With regard to teaching language skills, some investigated the role of corpora and concordancing in writing (Gilmore, 2009; Huang, 2014), and some in grammar (Boulton, 2009; Girgin, 2011; Vannestal and Lindquist, 2007). Although a considerable number of studies exist in literature to date, the studies on the effectiveness of corpus use in vocabulary learning based on students' performance (Boulton, 2008, 2012; Chan & Liou, 2005; El-Esery, 2015; Gilmore, 2009; Golabi, 2022; Kazaz, 2015; Koosha & Jafarpour, 2006; Lee & Lin, 2019) have focused on the learning of intermediate/advanced level students. Since existing corpora mostly appeal to high-level learners (Flowerdew, 2012, Meunier, 2011), the research on lower-level students' vocabulary learning remained relatively few. It is possible that researchers have tended to work with higher level students because native corpora have a high vocabulary load (Balunda, 2009, Sinha, 2021), and are not considered learner-centered, and culturally and contextually appropriate for all (Meunier, 2011). This could be one of the reasons as to why corpora use has not been normalized in language teaching and learning, and research-practice gap still exists (Chambers, 2019). To bridge this gap, the creation of a contextually relevant pedagogical corpus (as suggested by Braun, 2005) emerges as a potential solution, particularly in addressing challenges associated with the application of general corpora, especially in terms of learner-corpus interaction complexities (Kavanagh, 2021). Especially promising for learners at lower proficiency levels, who primarily engage with classroom language and struggle to comprehend less common vocabulary within native corpora, a pedagogical corpus offers the potential to amplify benefits.

Central to this study is the exploration of whether learners can benefit from a pedagogical corpus, a research that has received limited attention, particularly regarding hands-on engagement by lower-level learners with such a resource. Therefore, the novelty of this research lies in its attempt to employ a pedagogical corpus that comprises textbooks and teaching materials for vocabulary instruction. With this objective, the study seeks to investigate the efficacy of Data-Driven Learning (DDL) in enhancing students' vocabulary acquisition through the utilization of a pedagogical corpus.

LITERATURE REVIEW

Data-Driven Learning

Tim Johns's concept of DDL (1991) is a method of language learning that involves the exploration of regularities and patterns in language samples. Johns's earlier attempts to incorporate corpora into language teaching, as well as his seminal publications (1986, 1988, and 1991), and his website has become the backbone of subsequent pedagogical applications of corpora and DDL. One reason for the attention DDL received at the time was that it brought a new perspective to language teaching and learning. For example, it differs from traditional teaching methods in terms of the nature of instruction and the roles of students and teachers. It provides learners with the opportunity to examine corpora and recognize language patterns (Boulton, 2012; Hunston, 2002; O'Keeffe et al., 2007; Friginal & Roberts, 2022) enhances their ability to make generalizations (Gilquin, 2021; Johns, 1991), and thereby fostering inductive learning (O'Keeffe et al., 2007). In this regard, it also has the potential to enhance autonomous learning (Barabadi & Khajavi, 2017; Binkai, 2012).

As a prominent "computing tool for the data-driven approach" (Johns, 1991, p. 2), concordancers can analyze and organize massive amounts of texts in a very short time and present "potential patterns" by reducing all language data to a simple list in alphabetical order (Scott & Tribble, 2006, p. 5). Concordance lines and the keyword-in-context (KWIC) format facilitate the process of identifying patterns such as "lexical features" and "phraseology" of words (Sripicharn, 2003, p. 204) not only quantitatively but "beyond the frequency of the words" (O'Keeffe et al., 2007, p. 2).

Vocabulary Teaching Through DDL

While teaching English in schools, irrespective of whether it is presented under integrated skills courses or separately, traditional methods are commonly used to teach vocabulary such as providing synonyms, antonyms, L1 translations and definitions (Balcı & Çakır, 2011). Traditional vocabulary learning is usually characterized as an inactive process for learners as they usually obtain word lists and try to memorize them (Chen, 2004). However, memorizing does not necessarily mean that acquisition takes place (Nation, 2001). Understanding the form, meaning and usage of words (Nation, 2001) and knowledge of surrounding vocabulary are also crucial for acquisition (Harmer, 1993). In this respect, DDL is a potent alternative to traditional methods as it guides learners to observe and analyse data, look for patterns implicitly, and become autonomous researchers.

DDL is considered a viable approach to vocabulary learning on various grounds. First, it serves as an opportunity to get exposed to words in different contexts and forms (Barabadi & Khajavi, 2017; Wu et al, 2010), which in turn facilitates vocabulary expansion (Nation, 2001). Moreover, it increases both vocabulary breadth and depth as vocabulary items are presented in their collocative environment (Chen, 2004), and in their most common forms and patterns, which provides a means of studying collocates (McEnery & Xiao, 2011; Varley, 2009). On top of that, the above-specified vocabulary gains are not temporary as research (El-Esery, 2015, Karras, 2015) suggests that DDL yields long-term benefits and better vocabulary retention compared to traditional vocabulary

learning. DDL research (Binkai, 2012; Varley, 2009) revolved around immediate student attitudes and perceptions, while fewer studies focused on learners' performance (Barabadi & Khajavi, 2017; El-Esery, 2015, Kazaz, 2015). Furthermore, conducting empirical research on DDL offers limited results due to their tendency to focus on "specific, immediate learning outcomes", which makes it difficult to examine long-term achievement (Boulton, 2010, p.536). Therefore, conducting more empirical research on the long-term effects of DDL seems to be focal for our increased understanding of vocabulary retention through DDL practices.

Although its merits have been amply acknowledged in the existing literature, corpus resources and tools have not gained widespread adoption among language teachers and learners in the language classroom (Pérez-Paredes, 2022). Corpus use in language pedagogy is not without its criticism. Conversing views on the suitability of corpora for low-level students exist. While acknowledging that corpora such as the BNC appeal to advanced-level users, Aston[1] asserts that learners do not need to make sense of all data in corpus samples. On the issue, Flowerdew (2012) maintains that mostly higher-level learners can make sense of corpora while Lee & Liou (2003) found that low-level learners can benefit from corpora more compared to high-level learners. Conceding the difficulty of corpus samples, Chen (2004) suggests preparing concordance sheets to avoid dealing with irrelevant data. Toriida (2016) further recommends that teachers compile their corpus from textbooks, readers and journal articles for learners.

Despite being authentic in nature, the authenticity of native corpora has been questioned for learners from another culture, as well as their effectiveness in language learning (see, e.g., Widdowson, 2003). Drawing attention to the pedagogically relevant corpora, Braun (2005) postulates that authenticity can be improved to some extend if a corpus is pedagogically relevant in terms of content, language and culture. Textbook corpora and pedagogical corpora emerge as two important candidates at this point. As an extended version of textbook corpora, Meunier and Gouverneur (2009) define pedagogic corpora as "large enough and representative sample of the language, spoken and written, a learner has been or is likely to be exposed to via teaching material, either in the classroom or during self-study activities" (p.186). The terms *pedagogic corpora* and *pedagogical corpora* were used later in works of other scholars (see, e.g., Bennet, 2010; Chambers 2019), and such corpora have previously been created and exploited for textbook analysis (Bergström et al., 2023; Chen & Yuhua, 2023; Meunier, & Gouverneur, 2009; Sun & Dang, 2020), and writing genre analysis (Melissourgou & Frantzi, 2019). However, raising awareness of learners of all the occurrences of a word in different contexts (Huston, 2002) is an underrecognised exploitation of pedagogic corpora. Thus, taking the aforementioned issues into account, this study explored the effectiveness of DDL on vocabulary learning of tertiary-level EFL students by utilizing a pedagogical corpus. Two research questions guided the study:

(1) Is there a statistically significant improvement in vocabulary knowledge of target words for the participants who practiced through DDL when compared to those who received traditional vocabulary instruction through textbook materials?

(2) What are the attitudes of the DDL group towards learning vocabulary through DDL?

METHOD

Setting & Participants

The present study was undertaken at the School of Foreign Languages, at a state university in Turkey. At this institution, learners who cannot meet the language requirements to study in their respective departments receive general English language skills instruction. Through a one-year academic program, language skills are presented in an integrated way through "Main Course" and other supporting skill courses such as Listening & Speaking, and Reading & Writing. However, there is no special course devoted to vocabulary instruction.

Fifty-eight Turkish students (M = 44, F = 14) whose ages ranged between 18 and 22 took part in the study. The participants were placed in elementary classes based on the results of the placement test administrated at the beginning of the year and were receiving pre-intermediate level instruction during the study.

In this quasi-experimental study with a comparison group pre-test/post-test design, the participants were divided into experimental (N = 29) and control groups (N = 29) based on convenience sampling; that is, four classes with almost the same number of students were selected. Two intact classes were assigned to the experimental group, and two other intact classes formed the control group. To control some confounding variables such as course delivery, course content and duration, two instructors made sure that both groups followed exactly the same curriculum, courses and materials as planned by the administration and always updated each other for specific classroom practices.

Corpus Compilation

Before the intervention, a pedagogical corpus was compiled by the researchers with reading texts learners can

[1] Aston, G. (1998). *Learning English with the British National Corpus*. [Paper presentation]. 6th Jornada de Corpus, UPF, Barcelona. https://www.sslmit.unibo.it/~guy/barc.htm

make sense of while doing their analysis. For this reason, advanced-level texts were excluded. Samples of reading texts from elementary, pre-intermediate and intermediate level course-books and listening scripts from websites (see appendix A for the list of the sources) were included in the data set to compile a sample of both written and spoken language. The corpus contained a total of 358,972 tokens and 14,402 word types. While 27% of the corpus consisted of spoken language scripts, 63% consisted of written texts.

Classroom Materials

For the participants in the experimental group, researchers created five paper-based concordance handouts, as lower-level learners require guidance and support from their teachers to navigate the data they encounter during vocabulary searches (Granath, 2009). The handouts were developed following Nation's (2001) categorization of aspects of word knowledge: form, meaning and use. The questions in the handouts were constructed to raise participants' awareness of word forms, prefixes and suffixes, form-meaning relationships, collocations, and word usage (see appendix B for a sample concordance handout).

The aim was to introduce five target vocabulary items in each session. The target words were selected from the scheduled units in the Reading & Writing coursebook, with which the participants were being instructed. An adapted version of the *Vocabulary Knowledge Scale* (Schmitt & Zimmerman, 2002) was administrated to select less familiar vocabulary items among 36 pre-selected target words. The participants rated their own vocabulary knowledge with a four-level scale. 25 words which held the highest-rated "I do not know the word" option were selected as final target words.

For the exploration of the pedagogical corpus, the participants used the computer software "Antconc 3.4.4". The researchers downloaded the software for students' use before the intervention.

Data Collection Tools

Quantitative data was collected through a vocabulary test, which was administrated at three time points as pre-test, post-test and delayed post-test. Since an experimental design with a control group is adopted for the study, the existence of a control group reduces the test effect, in that the differences in the groups are not attributed to the retesting effect, but more likely to the intervention (Cook & Campbell, 1979). The vocabulary test, which was worth 100 points, comprised four parts: a) seven multiple-choice questions b) seven paragraph gap-filling questions c) six questions that require filling in concordance lines with target words d) five sentence-matching questions. Two academics holding M.A. in ELT examined the test in terms of content and face validity. They particularly checked whether the test includes all target words, and whether it measures different aspects of word knowledge specified in the objectives. They also evaluated the test with relation to level-appropriateness, clarity and formatting. Later, a pilot study was conducted with 20 students who were on the same level as the target population. The Cronbach Alpha was calculated as .94, which indicated high internal consistency.

A small questionnaire was also developed to explore students' perceptions of DDL after the intervention. It comprised 10 items in the format of a 5-point Likert Scale. While items 1 and 2 were created by the researchers, the remaining items were adapted from Boulton (2010), Girgin (2011), and Jablonkai & Čebron (2017). The reliability of the questionnaire was checked after data collection since it was only targeted at the research sample, and it was found to have internal consistency, r = .720.

Semi-structured focus group interviews were also conducted with seven volunteers from the experimental group to further investigate the efficacy of data-driven vocabulary learning, and explore opinions and experiences. Interviews were conducted after the intervention, and in the participants' mother tongue.

The Data Collection Procedure and Data Analysis

The study was conducted within three months in Spring Semester, 2016. One week prior to the pre-test, the Vocabulary Knowledge Scale was administered, and the final decision was made on the target words to be introduced. For the pre-test, which was given to both groups on the same day, the participants were given 30 minutes. Following that, the experimental group was given an introductory session on how to use the concordance program, open corpus files, and search for the target words.

During five-week treatment, the participants in the experimental group received 12 sessions that took 70-80 minutes each. In each session, they received a concordance handout, and were asked to analyse the corpus data following the "Identify - Classify - Generalise" procedure specified by Johns (1991, p. 5). Within this period, which took about 45-50 minutes of the session, the students were asked to identify relevant concordance lines, classify the patterns, and make generalizations to answer the questions in their handout. In the meantime, the instructing researcher observed the class and answered student questions. In the last 20-25 minutes of the session, a whole class discussion was held to elicit correct responses, alternative sentences or phrases identified from the concordance lines, and to uncover what they had learned. On the other hand, the control group covered their coursebook according to the schedule, and were introduced to the target vocabulary items through explicit teaching of teacher explanations and coursebook activities.

Upon completion of the treatment, the immediate post-test was administered to both groups in the same week. The experimental group was given the attitude questionnaire immediately after the post-test. The following week, seven students from the experimental group were interviewed. Five weeks after the immediate post-test, the participants were also given the delayed post-test.

For data analysis, the quantitative data collected through the vocabulary tests and the questionnaire were analysed using the Statistical Package for the Social Science (SPSS). On the other hand, the qualitative data, which were collected through interviews, were analysed using Creswell's (2012) qualitative content analysis scheme. The data were first transcribed, and the codes were created based on the recurring statements. Later, the themes that emerged from the codes were labelled.

RESULTS

The Effectiveness of DDL Based on the Vocabulary Tests

In an attempt to answer the first research question, which aims to find out if the participants show any improvement in their vocabulary knowledge of the target words, a descriptive analysis was first computed for the pre-test, post-test, and delayed post-test of the control and the experimental group (DDL group).

Table 1 illustrates that the control group (M = 40.76, SD = 21.1) scored 3 points higher than the experimental group (M = 37.8, SD = 18) in the pre-test. The mean scores for the immediate post-test indicate that although both groups showed improvements, the participants in the DDL group (M = 68.96, SD = 19.02) outperformed the control group (M = 53.85, SD = 20.29). Regarding the scores of delayed post-test, which was administrated five weeks after the post-test, the scores of both groups decreased compared to their post-test scores. However, the mean score for the control group (M = 48.23, SD = 22.47) dropped slightly more than the experimental group (M = 64.67, SD = 23.05).

Before comparing results for any significance, the baseline scores of participants in each group were first compared. The data was normally distributed and homogenous, so independent samples t-test was performed. The results presented in Table 2 indicate that although mean pre-test score of the control group was 2.93 points higher than experimental group's mean, these means did not differ significantly between the experimental group (M = 37.8, SD = 18) and the control group (M = 40.76, SD = 21.1) ([t [51] = -.54, p = 0.591]). Therefore, it can be concluded that the groups were not significantly different prior to intervention.

In order to explore the impact of concordance training on the vocabulary knowledge of target vocabulary items in three time intervals, 2x3 Mixed ANOVA was also performed. First, the assumptions of homogeneity of variances were tested and satisfied based on Levene's F test for the pre-test (F(1, 51) = 1.32, p = .256), post-test scores (F(1, 51) = 1.12, p = .30) and delayed post-test scores (F(1, 51) = 0.65, p = .80). Mauchly's test indicated that the assumption of sphericity has not been met, χ 2 (2) = 7.94, p = .020. Hence, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity (ε = .91).

Table 1

Descriptive Statistics for the Test scores of Control and Experimental Groups

Tests	Groups	*N*	*M*	*SD*
Pre-test	Experimental	27	37.85	18.06
	Control	26	40.77	21.13
Post-test	Experimental	27	68.96	19.02
	Control	26	53.85	20.29
Delayed post-test	Experimental	27	64.67	23.05
	Control	26	48.23	22.47

Table 2

Independent Samples T-Test for Pre-Test Scores of the Control and Experimental Group

	Levene's Test for Equality of Variances		T-test for Equality of Means				
	F	Sig.	t	df	p.	Mean Dif.	Std. Error Dif.
Pre-test	*1.32*	*.256*	*-.54*	*51*	*.591*	*-2.93*	*5.39*

Table 3

Within Subject Effects for the Control and the DDL group

Source		Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared
time	Sphericity Assumed	14654.92	2	7327.46	51.16	.000	.51
	Huynh-Feldt	14654.92	1.83	7994.22	51.16	.000	.51
time*group	Sphericity Assumed	3390.92	2	1695.46	11.84	.000	.19
	Huynh-Feldt	3390.92	1.83	1849.74	11.84	.000	.19

Table 4

Pairwise Comparisons of Test Scores across the Control and the DDL Group

Group	(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
experimental	1	2	-31.11	3.57	.000	-38.28	-23.94
		3	-26.81	3.63	.000	-34.10	-19.53
	2	3	4.30	2.53	.096	-.792	9.38
control	1	2	-13.08	3.64	.001	-20.38	-5.76
		3	-7.46	3.70	.049	-14.89	-.034
	2	3	5.61	2.58	.034	.430	10.80

As shown in Table 3, the main effect of time (pre-test, post-test and delayed post-test) on vocabulary test scores was significant, Huynh-Feldt, $F(1,91.66) = 51,16$, $p = <.001$, with a large effect (Eta-squared = .51), indicating that there were changes over time in vocabulary scores across the whole sample. This effect was qualified with a significant time and group (experimental/control) interaction effect, Huynh-Feldt, $F(1.83, 91.66) = 11.84$, $p = <.001$, eta squared = .19 revealing that the changes in vocabulary scores of participants over time are not equivalent across the two groups. This indicates that the groups are changing in different ways.

Post-hoc pairwise comparisons were also performed using paired-samples t-test to determine the levels of significance.

Follow-up pairwise comparisons presented in Table 4 indicated that the vocabulary scores of the treatment group significantly increased from time 1 to time 2 ($p = <.001$, cohen's $d = 1.68$), and time 1 to time 3 ($p = <.001$, $d = 1.29$) with a large effect size. However, no statistically significant difference was found between time 2 and time 3 ($p = .096$, $d = .20$), indicating that although the vocabulary scores have decreased from time 2 to 3 in the experimental group, it was not statistically significant. As for the control group, there was a statistically significant difference between time 1 and 2 ($p = .001$, $d = .63$) with a medium effect size, and time 1 and time 3 ($p = .049$, $d = .38$) with a small effect size. On the other hand, there was a significant decrease in the vocabulary scores from time 2 to 3 ($p = .034$, $d = .21$), which showed a rather poor retention rate in the vocabulary scores. These findings indicate that the experimental group showed a significantly greater improvement in the post test compared to the control group. Although both groups obtained lower scores in the delayed post-test than the post-test, the decrease in the scores is significant only in the control group, which implies that there is an overall higher improvement in the DDL group.

Learner Attitudes towards DDL

The effectiveness of DDL was further explored with an attitude questionnaire and focus group interviews. The findings regarding the attitude questionnaire are presented in Table 5.

As indicated in Table 5, overall, learners seemed to have a positive attitude towards DDL and believed that it improved their English. The mean score for item 4 was the highest ($M = 4.29$, $SD = .66$), showing that a big portion of students thought that studying concordance lines had a facilitating role in increasing knowledge of word usage. The lowest mean score, on the other hand, belonged to item 6 ($M = 2.07$, $SD = .91$), which indicated learners' disagreement on the difficulty of concordance lines. The participants found

DDL practices enjoyable (*M* = 3.96, *SD* = .70), wanted to continue studying with concordance lines (*M* = 4.03, *SD* = .75), and suggested that instructors utilize concordance to teach vocabulary (*M* = 4.11, *SD* = .93).

The findings from the interviews align with the results of vocabulary tests and the attitude questionnaire. The emerging themes and codes were listed in Table 6. The frequency of the codes expressed by participants is presented in parentheses.

The interview analysis revealed three themes that reflect the participants' experiences with data-driven learning. Overall, all the participants mentioned that they had fun while studying through concordance lines emphasizing its dissimilarity to other vocabulary learning methods they know. Some stated that they enjoyed studying vocabulary since it was computer-based and had a colourful display. One student remarks:

> *I had fun learning vocabulary with concordance lines because in class, we generally write or read, but here we did something different with computers and it was not boring at all.*

Regarding the strengths of data-driven learning, which emerged as the second theme, all the participants noted that they had the chance to analyse words in different sentences, and understand how words were used in those sentences, as seen in the extracts below:

> *There were only one or two example sentences in textbooks, but here I have seen a lot of sentences. I have also learned other forms of words.*

> *With the help of concordance lines, I have learned how words are presented in different forms such as verbs, nouns adjectives, etc., and I have also learned which prepositions are used after certain words because they were easy to see with concordance lines.*

The statements on how concordance lines helped students learn more words than they intended are worth attention. In particular, participants mentioned gains in word forms, collocations and lexico-grammatical structures. Some also pointed out the permanency of the knowledge acquired through data-driven learning. The following quotations show in what ways participants benefited from data-driven learning:

> *Concordance lines helped me learn vocabulary because we learned how the words we come across in the classroom are used in sentences, and their position in those sentences... while doing an activity, you learn a lot, for example, the prepositions, and the words with different meanings.*

> *Words in coursebooks are restricted, but here we can learn more words with concordance lines.*

> *Colourful words helped me a lot to identify which words are nouns, adjectives, verbs, etc. I sometimes remember the words with their colours. It became easier for me to remember them.*

The last theme that was drawn from student interviews was the concerns related to DDL practice. Two participants stated that doing research seemed complicated at first, yet they immediately added that they figured it out in a short time. Four participants were also cautious about using only concordance lines for vocabulary learning and suggested learning through concordance lines along with coursebook instruction. One participant touched upon this with the following extract:

> *I think that learning through concordance lines is beneficial, but it would be more effective when used with activities in our coursebook. I think they support each other.*

Based on the findings, we may conclude that data-driven learning had a positive impact on the vocabulary learning of the learners especially in the areas such as vocabulary expansion, word usage, word formation and collocations.

Table 5

Mean Scores for All the Items of the Attitude Questionnaire

	Items	*M*	*SD*
I think....			
1.	studying vocabulary through concordance lines is enjoyable.	3.96	.70
2.	studying vocabulary through concordance lines helps improve my English.	4.25	.59
3.	using concordance lines improved my English writing ability	3.74	.71
4.	using concordance lines is helpful for learning the usage of vocabulary.	4.29	.66
5.	using concordance lines in the learning of English vocabulary increased my confidence in learning English vocabulary.	3.29	.72
6.	learning vocabulary through concordance lines is more difficult than learning vocabulary through a coursebook.	2.07	.91
7.	learning vocabulary through concordance lines is more boring than learning vocabulary through a coursebook.	2.88	1.21
8.	I prefer using concordance lines in learning of English vocabulary to using a coursebook in learning of English vocabulary.	3.51	1.12
9.	I recommend that teachers should use concordance lines so as to teach vocabulary in EFL classes.	4.11	.93
10.	I would like to do more concordance activities in class.	4.03	.75

Table 6

Students' Perceptions on the Effectiveness of Data-Driven Learning

Themes	Codes
Affective factors	Fun (x7)
	Attractive due to being different (x6)
	Colourful layout (x5)
	Enjoyable since computer-based (x4)
Strengths	Multiple sample sentences (x7)
	Word forms (x7)
	Word usage (x7)
	Collocations + lexico-grammatical structures (x6)
	Learning beyond intended (x4)
	Permanent (x4)
Concerns	Analysing seemed a bit complicated at first (x2)
	A combination of coursebook and concordance can be more effective (x4)

DISCUSSION

The present study sought to explore the efficiency of the data-driven approach on EFL students' vocabulary learning. The results indicated that although both groups demonstrated learning of the target words, the DDL group showed a greater improvement and performed significantly better than the coursebook group on the post-test. The observed improvements in both groups were no surprise since the target vocabulary items were unknown to the participants before the study, and after the instruction period, it was expected that both groups would show improvement to some extent. However, the gain was higher by the DDL group, which indicated that data-driven practice, enriched with corpus and concordance was more efficient in enhancing students' vocabulary knowledge. The results echo previous empirical studies (see, for example, Binkai, 2012; El- Esery, 2015; Koosha & Jafarpour, 2006; Karras, 2015 and Kazaz, 2015), which reported that corpus-based learning yielded better results than other methods of vocabulary instruction.

The results regarding the delayed post-test revealed that the DDL group had a higher retention rate although both groups showed a decrease in the test scores, which indicated that DDL was more beneficial for the learners than the conventional coursebook-based instruction in the long run as well. The reasoning behind better retention could be drawn from the interviews in that some learners linked remembering words for a longer time to various examples, concordance display and colourful layout, supporting Schmitt (2008), who highlighted that exposure coupled with attention directed to lexical items results in better attention.

As regards to the learners' perceptions, the analysis of the questionnaires and interviews reflected positive attitudes towards DDL. This supports previous studies (Alsehibany & Abdelhalim, 2023; Chan & Liou, 2005; Gilmore, 2009; Kaur & Hegelheimer, 2005; Kazaz, 2015; Oktavianti et al., 2022; Sripicharn, 2003; Sun & Wang, 2003; Yoon & Hirvela, 2004; Youssef, 2021; Varley, 2009), which explored students' perceptions of DDL. The deeper analysis appears to have revealed that the learners showed enthusiasm towards using the concordance program and found DDL fun mostly because it was computer-based. Given that the learners are highly dependent on technology, and spend a great deal of their time on computers, it seems reasonable that they opt for learning English through technology and hands-on experience. This finding is also consistent with Chao[2], who pointed out the connection between the positive attitude and the technology-assisted nature of DDL. The results highlight the importance of integrating computer-based corpus analysis in classrooms rather than paper-based concordance as learners show enthusiasm towards DDL practices especially because they are computer-aided.

Another significant finding to emerge from the analysis is that the learners do not find studying vocabulary through DDL more difficult than coursebooks. This is contrary to the previous study by Vannestal & Lindquist (2007), who reported negative feelings of especially weak students towards corpus use. The difficulty of practicing DDL is usually associated with the process of using a concordance program, its display, and the language level of the concordance output (Chatpunnarangsee, 2015). However, in the present study, learners received a demo session, and studied through a

[2] Chao, P. (2010). *A study of collocation learning of junior high students in Taiwan via concordance.* [Paper presentation]. International conference on English teaching (2010), Kaohsiung, Taiwan. http://www2.kuas.edu.tw/edu/afl/20100430Final/Word/2010comp_EPCA.pdf.

pedagogical corpus compiled through texts of elementary and intermediate materials, which eased the burden on the shoulder of the learners. Employing a specialized corpus likely increased the learners' ability to make sense of the data and reduced the need for teacher supervision and monitoring (Aston, 2001). Hence, no feedback was received concerning the challenges of corpus analysis after the introductory session. This finding has significant practical implications for the integration process of DDL. Accordingly, guidance is needed at the beginning stages to help students familiarize themselves with both the concordance program and the concordance output. Classroom handouts play a crucial role in guiding students in the process of dealing with language if teachers have a predetermined aim in mind, such as introducing certain vocabulary items, grammar subject, or lexico-grammatical structures. At this point, there are implications for material developers as well. Since developing such materials is time-consuming for teachers, they need to design more DDL activities and corpus-based materials to motivate teachers in utilizing those ready-made resources without much effort.

In company with general contentedness, the students also perceived concordance-based practice and activities helpful for several reasons. First, they reported significant gains in word forms and word usage. Interview results also revealed that DDL helped students become familiar with lexico-grammatical patterns and collocations which were not intended to be introduced in the first place probably because DDL stands out in revealing grammatical and lexical structures by making patterns more visible (McEnery & Xiao, 2011). This supports Varley's (2009) findings regarding the positive effect of DDL in improving awareness of collocational expressions and lexico-grammatical patterns.

Another notable finding of the study was on the preferences of learners on which type of instruction they would like to receive for vocabulary learning. Although all students expressed positive views about DDL, they all recommended it as a supplementary study rather than a substitute for textbook instruction. The reason behind this preference could be the learning habits of the learners, who received conventional instruction during their primary school years. Naturally, learners feel safe when the information is presented to them, but feel challenged when they are required to use their cognitive skills while dealing with corpus output (O'Sullivan, 2007). Following the students' suggestions on integrating DDL into conventional teaching might create an effective learning environment especially for learners with different learning styles. In line with students' preferences, Meunier (2002) also suggests using corpora as a complementary method as not all types of exercises are compatible with corpus practice.

The present study also points to the need for teacher training. Although the significance and the effect of corpora on language teaching have been widely recognized, teaching programs on the use and evaluation of corpus materials are undervalued (McCarthy, 2008). Indicating that teachers have been "consumers" of corpus materials, McCarthy (2008) maintains that consumer teachers can turn into "active" users of corpus-based materials when corpora are integrated into teacher education; teachers are trained on the evaluation and use of corpus materials; corpora are built when necessary; teachers become "central stakeholders" or "lobbyists", and gain recognition in language education research (p. 565). At a more practical level, Granath (2009) points out the importance of training teachers on using corpora at the university level in courses such as "syntax, written proficiency and translation", which will get them to adopt the habit of consulting corpus rather than just resorting to dictionaries and grammar books (p. 47). For this to happen, it is important that teachers are equipped with corpus literacy. Ma et al. (2021) highlight that while corpus literacy functions as an educational resource, it does not automatically equip educators with the pedagogical skills needed to effectively integrate corpora into classroom teaching. Their research underscores the necessity for teachers to acquire corpus-based language pedagogy (CBLP), which involves integrating corpus linguistics technology into language pedagogy to enhance language instruction.

CONCLUSION

Our study delved into the effects of corpus and data-driven practices on students' vocabulary learning, contributing to the evolving literature on integrating corpora into EFL instruction. Notably, our findings demonstrated the superiority of Data-Driven Learning (DDL) in expanding and retaining vocabulary when compared to traditional coursebook-based instruction. Learner perceptions also yielded positive attitudes toward DDL and concordancing, emphasizing the tangible benefits of these approaches.

This study underscores the significant potential of pedagogical corpora, particularly those sourced from textbooks and instructional materials, in facilitating vocabulary learning, especially for low-level learners. These corpora offer direct access without the complexity of advanced sentence structures, compensating for the limitations of traditional coursebooks in presenting words with diverse meanings and forms, yet incorporating corpora into the classroom necessitates careful planning, teacher guidance, and instructional material support.

Our findings have broader implications for language education and invite educators to consider the potential benefits of corpus-driven practices. As we look ahead, the exploration of different corpus types and their roles in supporting low-level learners remains an important question, offering opportunities for future studies to continue advancing vocabulary instruction in diverse contexts. In future research, we encourage a comprehensive exploration of DDL's impact

on productive vocabulary skills and a systematic analysis of various corpus types. We believe that such investigations will further enhance our understanding of vocabulary acquisition.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Sibel Tosun: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Validation; Writing – original draft; Writing – review & editing.

Hatice Sofu: Conceptualization; Investigation; Methodology; Resources; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

REFERENCES

Alsehibany, R. A., & Abdelhalim, S. M. (2023). Enhancing academic writing vocabulary use through direct corpus consultation: Saudi English majors' perceptions and experiences. *Arab World English Journal*, *14*(1) 166-182. https://dx.doi.org/10.24093/awej/vol14no1.11

Aston, G. (2001). Learning with corpora: An overview. In G. Aston (Ed.), *Learning with corpora* (pp. 7–45). Athelstan.

Balcı, Ö., & Çakır, A. (2011). Teaching vocabulary through collocations in EFL Classes: The case of Turkey. *International Journal of Research Studies in Language Learning, 1*(1), 21-32. https://doi.org/10.5861/ijrsll.2012.v1i1.31

Balunda, S.A. (2009). *Teaching academic vocabulary with corpora: Student perceptions of data-driven learning* [Master's thesis]. Indiana University. https://scholarworks.iupui.edu/handle/1805/2049

Barabadi, E., & Khajavi, Y. (2017). The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary. *Cogent Education*, *4*(1), 1283876. https://doi.org/10.1080/2331186X.2017.1283876

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press. https://doi.org/10.3998/mpub.371534

Bergström, D., Norberg, C., & Nordlund, M. (2023). Do textbooks support incidental vocabulary learning? –a corpus-based study of Swedish intermediate EFL materials. *Education Inquiry*, 1-19. https://doi.org/10.1080/20004508.2022.2163050

Binkai, J. (2012). An empirical study on corpus-driven English vocabulary learning in China. *English Language Teaching*, *5*(4), 131–137. https://doi.org/10.5539/elt.v5n4p131

Boulton, A. (2008). Looking for empirical evidence of data-driven learning at lower levels. In B. Lewandowska-Tomaszczyk (Ed.), *Corpus linguistics, computer tools, and applications: State of the art* (pp. 581-598). Peter Lang.

Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL, 21*(1), 37-54. https://doi.org/10.1017/S0958344009000068

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

Boulton, A. (2012) Language awareness and medium-term benefits of corpus consultation. In A. Gimeno Sanz (Ed.), *New trends in CALL – working together* (pp. 39–46). Macmillan.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348-393. https://doi.org/10.1111/lang.12224

Braun, S. (2005) From pedagogically relevant corpora to authentic language learning contents. *ReCALL 17*(1), 47-64. https://doi.org/10.1017/s0958344005000510

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, *52*(4), 460-475. https://doi.org/10.1017/S0261444819000089

Chan, T., & Liou, H. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning, 18(3),* 231-250. https://doi.org/10.1080/09588220500185769

Chatpunnarangsee, K. (2015). Corpora, concordancing and collocations. *Journal of English Studies, 10*, 102-136.

Chen, W. A. N. G., & Yuhua, L. I. U. (2023). Measuring vocabulary use in Chinese tertiary textbooks: Potentials for incidental vocabulary learning. *Asian Journal of English Language Teaching, 32*(1), 131-154.

Chen, Y.H. (2004). The use of corpora in the vocabulary classroom. *The Internet TESL Journal 10*(9). http://iteslj.org/Techniques/Chen-Corpora.html

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478–497). Cambridge University Press. https://doi.org/10.1017/cbo9781139764377.027

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly, 34*(3), 548–60. https://doi.org/10.2307/3587743

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and analysis issues for field settings*. Houghton Mifflin.

Creswell, J. W. (2012). *Educational Research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.

El-Esery, A. M. (2015). Corpus concordance for enhancing EFL learners' vocabulary retention. *Journal of Education and Training*, *3*(1), 56-64. https://doi.org/10.5296/jet.v3i1.8202

Flowerdew, L. (2012). *Corpora and language education.* Palgrave Macmillan. https://doi.org/10.1057/9780230355569

Friginal, E., & Roberts, J. (2022). Corpora for materials design. In *The Routledge Handbook of corpora and English language teaching and learning* (pp. 131-146). Routledge. https://doi.org/10.4324/9781003002901-11

Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal, 63*(4), 363-372. https://doi.org/10.1093/elt/ccn056

Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics, 31*(2), 229-247. https://doi.org/10.1111/ijal.12317

Girgin, U. (2011). *Corpus-based actıvıtıes at lower levels of EFL proficiency: The effectiveness of usıng concordance lınes on grammar learnıng* [Master's thesis]. Bilkent University. http://repository.bilkent.edu.tr/handle/11693/15191

Golabi, N. (2022). The effect of concordancing based teaching on student's vocabulary retention. *Journal of Research in Techno-based Language Education, 2*(1), 25-32.

Gouws, R. (2021). Expanding the use of corpora in the lexicographic process of online dictionaries. In M. Piosik, J. Taborek, & M. Woźnicka (Eds.), *Korpora in der Lexikographie und Phraseologie: Stand und Perspektiven* (pp. 1-20). De Gruyter. https://doi.org/10.1515/9783110716955-001

Granath, S. (2009). Who benefits from learning how to use corpora? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 47–65). John Benjamins. https://doi.org/10.1075/scl.33.07gra

Harmer, J. (1993). *The practise of English language teaching*. Longman.

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL*, *26*(02), 163-183. https://doi.org/10.1017/s0958344014000020

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press. https://doi.org/10.1017/cbo9781139524773

Jablonkai, R. R., & Čebron, N. (2017). Corpora as tools for self-driven learning: A corpus-based ESP course. In N. A. Alias & J. E. Luaran (Eds.), *Student-Driven learning strategies for the 21st century classroom* (pp. 274-298). IGI Global. https://doi.org/10.4018/978-1-5225-1689-7.ch018

Johns, T. (1991). Should you be persuaded two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom concordancing (ELR Journal 4)* (pp. 1-16). Centre for English Language Studies.

Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL, 28*(2), 166-186. https://doi.org/10.1017/S0958344015000154

Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, *18*(4), 287–310. https://doi.org/10.1080/09588220500280412

Kavanagh, B. (2021). Norwegian in-service teachers' perspectives on language corpora in teaching English. *Nordic Journal of Language Teaching and Learning, 9*(2), 90-106. https://doi.org/10.46364/njltl.v9i2.933

Kazaz, I. (2015). *Corpus-aided language pedagogy: The use of concordance lines in vocabulary instruction* [Master's thesis]. Bilkent University. http://repository.bilkent.edu.tr/handle/11693/16944

Koosha, M., & Jafarpour, A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal Quarterly, 8*(4), 192-209.

Lee, C. Y., & Liou, H. C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching and Learning, 27*(3), 35-56.

Lee, H., Warschauer, M. & Lee, J. H. (2019) The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721-753. https://doi.org/10.1093/applin/amy012

Lee, P., & Lin, H. (2019). The effect of the inductive and deductive data-driven learning (DDL) on vocabulary acquisition and retention. *System*, *81*, 14-25. https://doi.org/10.1016/j.system.2018.12.011.

Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. Computer *Assisted Language Learning, 35*(9), 2731-2760. https://doi.org/10.1080/09588221.2021.1895225

McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching, 41*(4), 563-574. https://doi.org/10.1017/s0261444808005247

McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (vol. 2, pp. 364-380). Routledge. https://doi.org/10.4324/9780203836507

Melissourgou, M. N. & Frantzi, K. T. (2019). Stories, second language writers, and genre awareness. In V. Kourtis-Kazoulis, T. Aravossitas, E. Skourtou, & P.P. Trifonas (Eds.), *Interdisciplinary Research Approaches to Multilingual Education* (pp. 117-129). Routledge

Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In Granger, S., Hung, J. & Petch-Tyson, S. (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 119-141). John Benjamins. https://doi.org/10.1075/lllt.6.10meu

Meunier, F. (2011). Corpus linguistics and second/foreign language learning: Exploring multiple paths. *Revista Brasileira de Linguística Aplicada*, *11*(2), 459-477. https://doi.org/10.1590/S1984-63982011000200008

Meunier, F., & Gouverneur, C. (2009). New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material. In K. Aijmer (Ed.), *Studies in corpus linguistics* (vol. 33, pp. 179–201). John Benjamins. https://doi.org/10.1075/scl.33.16meu

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to classroom. language use and language teaching*. Cambridge University Press. https://doi.org/10.1017/cbo9780511497650.009

O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL, 19*(3), 269-286. https://doi.org/10.1017/s095834400700033x

Oktavianti, I. N., Triyoga, A., & Prayogi, I. (2022). Corpus for language teaching: Student' perceptions and difficulties*. PROJECT (Professional Journal of English Education), 5*(2), 441-455. https://doi.org/10.22460/project.v5i2.p441-455

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1-2), 36-61. https://doi.org/10.1080/09588221.2019.1667832

Römer, U. (2008). Corpora and language teaching. In A. Ludeling & M. Kyto (Eds.), *Corpus linguistics: An international handbook* (vol. 1, pp. 112-130). Mouton de Gruyter.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language teaching research*, *12*(3), 329-363. https://doi.org/10.1177%2F1362168808089921

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145-171. https://doi.org/10.2307/3588328

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education.* Benjamins. https://doi.org/10.1075/scl.22

Sinha, T. S. (2021). EFL learners' perception of and attitude to corpus as a vocabulary learning tool. *The Reading Matrix: An International Online Journal, 21*(2), 106-119.

Sripicharn, P. (2003). Evaluating classroom concordancing: The use of concordance-based materials by a group of Thai students. *Thammasat Review, 8*(1), 203–236.

Sun, Y. C., & Wang, L. Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, *16*(1), 83-94. https://doi/10.1076/call.16.1.83.15528

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System, 93*, 102279. https://doi.org/10.1016/j.system.2020.102279

Toriida, M. C. (2016). Steps for creating a specialized corpus and developing an annotated frequency-based vocabulary list. *TESL Canada Journal*, *34*(1), 87-105. https://doi.org/10.18806/tesl.v34i1.1257

Vannestal, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course*. ReCALL, 19*(3), 329-350. https://doi.org/10.1017/S0958344007000638

Varley, S. (2009). I'll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning, 22*(2), 133-152. https://doi.org/10.1080/09588220902778294

Widdowson, H.G. (2003). Expert beyond experience: Notes on the appropriate use of theory in practice. In. Newby, D. (Ed.), *Mediating between theory and practice in the context of different learning cultures and languages* (pp. 23–30). Council of Europe.

Wu, S., Witten, I.H. & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL, 22(1),* 83-102. https://doi.org/10.1017/s0958344009990218

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, *13*(4), 257-283. https://doi.org/10.1016/j.jslw.2004.06.002

Youssef, A. F. F. A. (2021). Learners' Attitudes Towards Utilizing Corpus-based Approach in the Saudi EFL *Context. College of Arts Journal, 5*(58), 5-36. https://doi.org/10.21608/jfabsu.2021.159373

APPENDIX A

Sources used for the corpus

Acevedo, A., Lethaby, C., & Harmer, J. (2007). *Just listening and speaking: For class or self study*. Marshall Cavendish Education.

BBC (2007). *BBC learning English: Talk about English webcast: Archive.* http://www.bbc.co.uk/worldservice/learningenglish/webcast/tae_insight_archive.shtml

Bernard, L. (2002). *Absolute English*. http://absolutenglish-972.pagesperso-orange.fr/notes/carnival/raw/_starthere.htm

Blanchard, K. & Bonesteel, L. (*2010). From reading to writing 2*. Pearson Education.

British Council (n.d). *Learn English teens: listening*. http://learnenglishteens.britishcouncil.org/skills/listening-skills-practice

Dummett, P., Hughes, J., & Stephenson, H. (2015). *Life: Student book 2*. Cengage Learning.

Evans, V. & Dooley, J. (2000). *Reading & writing targets 2*. Express Publishing.

Evans, V. & Dooley, J. (2000). *Reading & writing targets 3*. Express Publishing.

Greet, J. (Ed.). (2010). *Timesaver intermediate listening.* Mary Glasgow Magazines.

Hill, L. A. (1998). *Elementary stories for reproduction*. Oxford University Press.

Hill, L. A. (1998). *Intermediate stories for reproduction*. Oxford University Press.

Kandiller, B., & Velioğlu, A. (2012). *Reader at work 1*. Ankara: METU Press.

MacIntyre, T. (2009*). Reading explorer 2*. Boston: Heinle Cengage Learning.

Mc Carthy, M., Mc Carten, J., & Sandiford, H. (2005*). Touchstone level 2, student's book.* Cambridge University Press.

Ostrowska, S. (2014). *Unlock 2: Reading & writing skills*. Cambridge University Press.

Oxenden, C., Latham-Koenig, C., & Seligson, P. (2009). *New English file: Pre-intermediate student's book*. Oxford University Press.

Oxenden, C., Latham-Koenig, C., & Seligson, P. (2004). *New English file: Intermediate Student's book*. Oxford University Press.

Publishing, S. (2002). *Reading comprehension skills and strategies level 3*. Saddleback Educational Publishing.

Richards, J. C., & Sandy, C. (2008). *Passages* (2nd edition). Cambridge University press.

Scott, R. (1987). *Reading: Elementary*. Oxford University Press

Türkan, C. C., & Türkan, N. (2006). *More to read 1*. METU Press.

APPENDIX B

Sample Concordance Handout

Analyse the words/phrases given and answer the following questions.

1. Search for join*

What is the part of speech of the word "join"? _____

What preposition(s) directly follows "*join*"? _____

List 6 phrases that collocate with *the word "join"* as an object of the verb. (e.g.: join a racing school) _____

2. Search for appl*

Identify 2 different parts of speech in the concordance lines (e.g.: noun, verb, adjective, adverb, etc.) and write 2 example concordance lines below. _____

Find a concordance line in which "apply" is used in an infinitive construction. Write it below. _____

What verbs precede the infinitive form? _____

What preposition(s) directly follows *apply*? _____

List the words or phrases after the preposition(s) you wrote above. _____

3. Search for look* after

What is the part of speech of "look after"? _____

Find a concordance line in which "look after" is used as a main verb. Write it below. _____

Find a concordance line in which "look after" is used in an infinitive construction. Write it below. _____

List 6 phrases that collocate with *"look after"* as the object of a verb. (e.g.: look after my family) _____

What do the sentences have in common? What do people normally 'look after'? _____

Looking at the concordance lines, write a sentence using "look after". _____

https://doi.org/10.17323/jle.2023.16211

What Can Count As Critical Academic Literacy Education?

Filippos Tentolouris

University of Thessaly, Volos, Greece

ABSTRACT

Background and Purpose: The issue of what can count as "critical" in relation to academic literacy education has not been discussed in detail in relative review studies. Therefore, this opinion article aims to contribute to this issue by exploring the question whether a field of academic literacy education can be underpinned.

Approach: First, I revisit some models of academic literacy education (rhetorical models of critical consciousness, models of critical language awareness, genre-based models, multiliteracies, ethnographic-based academic literacies) which have been considered as "critical" in taxonomies of these review studies. Then, I compare these models showing their similarities and differences regarding what is "critical" and how it is situated within academic literacy education.

Conclusion: Finally, I argue that since there are contrasting conceptualisations among these models in relation to what is "critical" and how it can be associated with academic literacy education, critical academic literacy education can count as a relativist and not a unified field.

KEYWORDS

academic literacy, critical language awareness, critical pedagogy, multiliteracies

INTRODUCTION

The fundamental premise of critical literacy education is the dialectical relation between literacy education and social dimensions, in other words a type of literacy education with a political orientation to "the cultural, ideological, and sociolinguistic content" (Luke, 2012, p. 5). Many theorists (e.g., Freebody, 2017; Luke, 2012; Luke & Freebody, 1997; Pennycook, 2001; Yoon & Sharif, 2015), in their attempt to define and describe this kind of literacy, have suggested various relative taxonomies with models of critical literacy education.

However, since what differentiates a "critical" from a "non-critical" model is not always discussed explicitly within these taxonomies, the encompassed models vary from one taxonomy to another, e.g., ethnographic-based models are included in Freebody's taxonomy but not in Yoon's & Sharif's. Furthermore, the various taxonomies do not specify what can count as critical academic literacy education, because they cluster various models without associating them with a particular level of education (e.g., primary or higher education). In other words, they do not discuss if there are certain issues of critical literacy at stake at different levels of education and if a specific approach could be more suitable for a specific level or not.

Although many studies (e.g., Blummer, 2016; Clarence & McKenna, 2017; Ivanič & Clark, 1997; Yu, 2021) have presented how critical academic literacy education can be implemented from the perspective of a particular model, e.g., a linguistic (Ivanič & Clark, 1997) or an ethnographic one (Clarence & McKenna, 2017), the issue of whether a critical academic literacy education can be delineated as a relatively stable (sub)field of research and teaching has not been discussed. For example, in some recent review papers on academic literacy education (e.g., Fu & Wang, 2022; Li, 2022), this field is depicted as a continuous shift among different trends (e.g., from the sociocultural to the new literacies studies) without a distinction between "critical" and "non-critical" ones. On the one hand, such fluidity could celebrate difference and openness. On the other, it could impede the development of the field as a distinct field de-

Citation: Tentolouris F. (2023). What Can Count As Critical Academic Literacy Education? *Journal of Language and Education, 9*(3), 191-197. https://doi.org/10.17323/jle.2023.16211

Correspondence:
Filippos Tentolouris,
ftentolouris@uth.gr

Received: October 31, 2022
Accepted: September 15, 2023
Published: September 30, 2023

marcating what should be considered as critical both in the context of empirical research investigating applications of critical literacy education but also in the context of educational practice informed by critical literacy education.

The goal of this paper is, then, to revisit and compare some models of critical academic literacy education which have been discussed in general terms in some of the above taxonomies. This can contribute to the international ongoing discussion (Fu & Wang, 2022; Wang & Derakhshan, 2023; Wang et al., 2022) regarding how the field of academic literacy education can be mapped out and how stimulating contexts of learning and teaching academic literacy can be organised.

REVISITING SOME MODELS OF THE CRITICAL ACADEMIC LITERACY EDUCATION

Rhetorical Models Focusing on "Critical Consciousness"

Although Paulo Freire did not coin the concept of "critical literacy" and did not use it in his initial works (e.g., Freire, 1972), he is considered by some theorists as its main founder (Luke, 2012, p.5). His model of literacy education (a pedagogy of raising "critical consciousness") has been described as "emancipatory" from a Marxist perspective (Gee, 1990). In particular, Freire (1972) argued that language, apart from being a medium of communication, is a means of political action, since oppressed people can gradually discover that they can describe reality from their own perspective through literacy.

After Freire, theorists from the field of Critical Pedagogy (see for a review, McArthur, 2010) criticised the reproductive function of school literacy, which is supposed to be ideologically neutral, and proposed a "critical literacy" (Kretovic, 1985). This kind of literacy can help students not only reflect on what counts as a valid and reliable source of knowledge but also examine what is present and what is silenced in it (e.g., in textbooks of history).

The idea of critical literacy and how it signals implicit power relations and ideological positions was recontextualised in academic literacy education by some theorists within the field of New Rhetoric in North America in the end of the 80's and the beginning of the 90's. For example, Chase (1988) and Bizzell (1992), echoing Freire's view that oppressed people should understand the political dimension of literacy, argued that students are politically oppressed individuals who should be helped so as to attain a "critical consciousness" in their attempt to master the academic discourse. They should realise that in their discourse communities textual conventions are not monolithic but they have ideological perspectives, e.g., science should be written through an impersonal style because it is considered to be a systematic and "objective" form of knowledge.

This new rhetoric perspective is not based on a specific linguistic theory (e.g., a functional grammar) but on the contrary it places emphasis on explicating textual features by focusing on "unpacking complex relations between text and context" (Freedman & Medway, 1994, p. 8). For example, Chase (1988) was among the first rhetoricians who discussed the ideological perspective of academic writing by focusing on case studies. Following the critical pedagogist Henry Giroux who argued that students follow three different paths in the learning process within a dominant culture (accommodation, opposition and resistance), he suggested that these paths can also help us understand how students deal with academic writing conventions.

The first path refers to how students "accommodate" themselves within discourse communities without "risking", e.g., a student effectively structures an essay following a standard conventional way -introduction, chapters which represent aspects of knowledge-general conclusions- without questioning any source or aspect of this knowledge. The second path is related to how writing conventions are questioned by some students because they find them too abstract to express themselves. They decide to write their texts in their own unsystematic way even if they take the risk to fail. Finally, the third path refers to how students decide to resist to what is considered to be a "typical" and an "appropriate" textual representation of knowledge. But, unlike the students above, they adopt an explicit perspective of using them in their own and alternative way. For example, a student decides to write history by incorporating strategies from creative writing because he/she believes that this manner reflects the subjective and interpretative perspective of history and he/she wants to construct a corresponding identity as a historian.

Therefore, students' critical academic writing cannot be developed when university writing programmes have as a single goal to help students deal with courses writing demands by equipping them with standard and decontextualised language features (Bizzell, 1992, pp. 129-152). If disciplinary life is essentially rhetorical requiring from its members to find their own "voice" and participate in its dialogues about the status of knowledge and its useful social outcomes, students should then be academically socialised in such way; to realise the political aspect of writing and to make practice of it.

Linguistic Models Focusing on "critical Language Awareness"

In the late 80's and in the beginning of 90's a group of linguists at Lancaster University (Clark et al., 1991; Fairclough,

1989) produced an approach of critical discourse analysis which was based on Halliday's systemic functional grammar. According to this grammar (Halliday, 1985), language is considered to be a semiotic resource of meaning making which serves social functions, e.g., people through modality (e.g., modal verbs, adverbs) commit themselves to truth and obligation to different degrees. Critical discourse analysis has an explicit Marxist perspective, e.g., society is stratified into dominant and non-dominant classes, in which language is considered as a medium of reproducing or challenging power relations and ideological positions, projected as "real", "given" or "stable" in human communication.

This group made a step further by transforming this linguistic approach into a pedagogy for all grades of education which was called "critical language awareness" (Clark et al., 1991; Fairclough, 1992). Regarding academic writing (Britton & Austin, 2022; Hankerson, 2023; Ivanič & Clark, 1997), the group focused on helping students from different backgrounds (e.g., overseas students), disciplines (e.g., science) and learning needs (e.g., writing essays or dissertations) perceive academic writing from a critical linguistic perspective.

According to this perspective (Ivanič & Clark, 1997, pp. 63-67), writing is not a set of skills, which can be passed up to learners, but a total of social practices in relation to various academic texts such as essays, assignments, articles and reports. Since these practices are shaped within specific contexts and social relations, students should reflect on their sense of the power relationships involved, and of their status in order to decide how they want to position themselves as writers: "a crucial aspect of critical language awareness is to empower students by providing them with the opportunities to discover and critically examine the conventions of the academic discourse community and to enable them to emancipate themselves by developing alternative to the dominant conventions" (Clark, 1992, p. 137).

For example, one of the important issues in academic writing is how writers organise the responsibility for the propositions mentioned in their texts. The degree to which a text appears to be "authoritative" or "believeable" depends on the language features which in Halliday's functional grammar are associated with modality as described above. According to Ivanič & Clark (1997, pp. 171-72), such features of modality are the following: (a) modal adverbs (e.g., "perhaps"), (b) evaluative adjectives (e.g., "tentative"), (c) modal auxiliaries (e.g., "may"), (d) lexical verbs (e.g., "think"), (e) context-dependent nouns (e.g., "factor", "variable"), (f) use of active or passive voice (e.g., "It is widely believed…") (g) appeals to authority (e.g., Fairclough argues that..") and (h) lexical choice (e.g., "collateral damage vs civilian deaths").

By reflecting on the epistemological consequences which emerge from selecting some of these features with regard to the status of knowledge and the construction of their identity, students can develop their awareness of writing as a social practice. For example, they can challenge the use of an abstract and "objective" academic discourse which is usually highly valued in the most industrialised and educated societies (Ivanič & Clark, 1997, pp. 145-146).

Linguistic Models Focusing on the Empowering Function of Genres

In the mid 80's another group in Australia, which was also influenced by Halliday's systemic functional grammar, worked on a common agenda on how models of genre-based literacy education could be developed (see for different models, Cope & Kalantzis, 1993). Although a Marxist perspective or an alternative sociological one was not explicitly adopted, a common feature of these models is the belief that an overt instruction of genres -recurrent textual patterns of real social life (Rose, 2015)- can be empowering for students. In particular, if they learn how to fit the appropriate linguistic forms into various communicative situations (e.g., to use the conventions of descriptive and argumentative genres), it will facilitate them in having an easier and more active access and participation in social life.

At the academic educational context, this required that students should acquire the writing conventions which are dominant in their discipline (Swain, 2009; Tribble & Wingate, 2013; Yu, 2021) in order to be capable of reproducing them without problems. To give an example (Swain, 2009), one of the main issues in writing an academic essay with evaluative propositions, is how students can orchestrate these propositions and produce arguments which are simultaneously personal (signaling writers' beliefs) and collective (signaling shared beliefs within a community). If coherent argumentation and how it is built within essays is a significant feature for the success or the failure of these essays, students can be helped to deal with it through an explicit teaching on how to "engage" themselves with propositions, arguments and value judgements.

Specifically, they can learn that "engagement" is organised into systems that are divided further into smaller subsystems. For example, the system of "dialogic contraction" is divided into "disclaim" and "proclaim" (White, 1997) and the latter into "concur" (e.g., the phrase "of course"), "pronounce" (e.g., the phrase "It is clear") and "endorse" (e.g., the phrase "the paper shows"). Therefore, through mastering the features of engagement students will be able to present and evaluate shared beliefs within a community from their own perspective, signaling a step further in their academic socialisation.

Linguistic Models Focusing on "Critical Framing"

Another model of literacy education, multiliteracies, was suggested in the mid 90's by the New London Group (1996).

The fundamental premise of this model is that contemporary communication incorporates resources of meaning which were not dominant in the past (e.g., images, sound), "traveling" through technology (e.g., social media) from global to local contexts. This model is based on both, Halliday's Functional Grammatical Perspective as well as on its extension to the understanding of other semiotic systems such as images and music. For example, Kress & van Leeuwen (1996) suggested that the representation of reality, what Halliday (1985) called the "ideational" function of language, can be accomplished in images through narrative and conceptual representations (the former as sequences of actions while the latter as taxonomies and systems of ideas).

This group, as the previous one, does not adopt an explicit Marxist agenda or an alternative one but its proposed four stage model of literacy education contains a stage of "critical framing" (1. "situated practice", 2. "overt instruction", 3. "critical framing" and 4. "transformed practice"). In this stage students -using features of a metalanguage of text design which were taught in the previous stage- are called to analyze texts by associating text functions with participants' interests and power in the process of communication. Although this seems to be to some degree similar with what is suggested in CLA, it is not clear if this type of analysis aims to raise a kind of consciousness toward a challenge of the existing conventions or aims to help students gain access in the new complex communicative landscape, in a corresponding empowering way as described in the previous section (Pennycook, 2001, pp. 99-100).

According to academic models of "multiliteracies" (Blummer, 2016; Jones, 2009; Newman, 2002), contemporary academic learning environments are inevitably multimodal and hypertextual. Students are asked not only to comprehend but also to produce oral (e.g., power point presentations) and written texts (e.g., essays) which are combined with visual categories (e.g., figures, charters, tables, equations, graphs). For example, Jones (2009) discussed in detail how technical (e.g., scientific terms hyperlinked with internet glossaries or databases of images) and abstract information (e.g., graphs showing measured scientific phenomena) constitute important visual learning materials in an undergraduate first year science course. If new information and communication technologies are used in today's universities shifting from print- to multimedia-based learning materials with an important role of images, then academic literacy shifts correspondingly to academic literacies (e.g., visual literacy, verbal literacy).

Therefore, specific practices should be designed by academic institutions so as to develop new multimedia competencies in students (Blummer, 2016). For example, through supportive courses students can learn how to design academic genres by analysing and reflecting on specific texts in order to take decisions not only on how to maximise their effectiveness, i.e., to decide which information will be written and which information can be depicted through images, but also on how to construct their identities as designers.

Ethnographic-Based Models Focusing on a Heuristic Critical Dimension

The issue of academic literacy has also been approached by ethnographic-based models which have emerged within the field of New Literacy Studies (Clarence & McKenna, 2017; Hasrati & Tavakoli, 2019; Zhou et al., 2020). The starting point of these models is not a specific linguistic perspective (e.g., a functional one as described above) which predetermines to some degree the meaning of literacy (see for details, Street, 2000). On the contrary, literacy is seen as a set of practices, namely patterns of using literacy which are shaped within specific sociocultural events. This entails that meaning which emerges within practices, cannot be isolated from the events and, therefore, it is difficult to make generalisations in advance regarding the functions fulfilled by literacy.

For example, "writing a note" seems to be an easy task which is learnt once and for all but when this task is associated with specific contexts -writing a note to a family member compared with a note to colleague- it becomes more complex. The former cannot be considered in advance as simpler at the level of style and grammar compared with the latter; parents might write relatively complex notes by giving detailed instructions to their children while two colleagues may use a simple codified language in their communication (e.g., using elliptical sentences) in order to save time.

Concerning academic writing these models aim to pinpoint that academic writing is a situated, complex and nuanced issue which is associated with students' epistemological practices of using literacy (academic literacies). Since these practices are framed within specific institutions and affected by asymmetrical power relations between academic staff and students, an effective academic writing presupposes not only an appropriate use of academic language but also a way to "fit" this language to institutional context which many times is an implicit issue.

Despite the fact that ethnographic-based models are not labelled "critical", concepts which are associated with the critical dimension of literacy such as "ideology" or "power" are used in them in a heuristic way. Lea & Street (1998, 2000), for instance, showed through ethnographic techniques (e.g., interviews with tutors and students, instructions for writing assignments) that a failure of fitting content to context might be the outcome of the tension between the epistemological and power relationships between tutors and students. They might have contrasting expectations and interpretations concerning how academic knowledge should be approached which might not be articulated explicitly, e.g., which ideas should be taken for granted and not included in an essay or an epistemological stance which is not con-

sidered as compatible with a specific way of approaching a topic). Additionally, tutors might not be interested in exploring why some students "fail" or in helping them "improve" their texts, because they might consider their role as assessors and not as facilitators. In other words, the success of academic texts lies on the epistemological and power relationships among all those who are involved in writing and evaluating knowledge represented in writing and not in writing per se.

If the boundaries between what is "good" and "poor" academic writing are not completely predetermined since the criteria of evaluation vary from discipline to discipline and from tutor to tutor (even at the same department), academic writing can be conflicting and contested. Therefore, it is important for students to understand through specialised workshops and materials on academic writing (e.g., Lea & Street, 2006) that there is no an academic literacy as a neutral skill which can be packaged up and passed to students.

COMPARING THE REVISITED MODELS

Comparing the above models of academic literacy education and reflecting on how all or some of them have been associated to critical literacy education in the relevant taxonomies, we might end up with the following three interrelated points. First, not all models were labelled "critical" by their initiators which raises an important question on whether this label is a prerequisite for someone to categorise it within a critical academic literacy education. For example, Yoon & Sharif (2015) do not encompass in their taxonomy ethnographic-based models of literacy education, believing probably that despite their social perspective, they do not have an explicit "critical" dimension or a transformative element. On the contrary, Pennycook (2001) and Freebody (2017) include these models by arguing that ethnographic-based models foreground a social and empowering dimension of literacy which "fits" a critical literacy education. Therefore, genre-based literacy models, multiliteracies and ethnographic-based models might be included or excluded from taxonomies or reviews of critical academic literacy education depending on theorists' perspectives on whether such model of literacy education should foreground a critical perspective or not.

Second, not all models of literacy education adopt an agenda of resistance and consciousness, inspired by the Marxist theory. Rhetorical models which focus on critical consciousness and models of critical language awareness call for pedagogies which can help students realise the political aspect of academic writing. According to these models, writing is not just a medium of presenting how knowledge is negotiated but it is also a medium through which tutors and students construct identities, as writers and readers, as well as regimes of truth by depicting reality from various, and many times, contrasting perspectives. On the contrary, the other models (genre-based models, multiliteracies and ethnographic-based models) do not promote an agenda of resistance and consciousness by raising explicit questions of the political aspect of academic writing. Instead, they seem to focus on how students can be empowered through a kind of reflexive academic socialisation within the academic discourse communities. Specifically, if students realise the complex textual, multimodal and contextual aspects of academic literacy, they will be able not only to participate more effectively in their communities, but also to reflexively position themselves in them as well as to explain possible tensions and failures in their assignments.

Third, the revisited models were produced within different epistemological orientations. Models of critical language awareness, genre-based literacy models and multiliteracies were based in Halliday's systemic functional grammar. They place emphasis on linguistic and multimodal structures, for example how an argumentative genre can be thematically organised or how images contain narrative, conceptual or mixed representations as well as how students can build specific identities within them through language features of modality. In contrast, rhetorical and ethnographic-based models are not committed to a specific grammar or a linguistic perspective. The former aim at showing how students might struggle or not in how to position themselves concerning the representation of academic knowledge which is reflected in the thematic and stylistic organisation of their texts. The latter focus on how textual meaning is interdependent with literacy practices which are shaped within a specific academic context signaling explicitly or implicitly epistemological and power relations among students and teachers.

The above comparison denotes that there are no tight criteria for what really counts as a critical literacy model of education, since what is "critical" and how it is linked with academic literacy education are issues which are conceptualised explicitly or implicitly in different ways. Fu and Wang (2022) have argued that it is important for teachers to perceive academic literacy education as a field with continuous shifts among different trends (e.g., from a socio-cognitive to a sociocultural one). Given that contemporary academic literacy is complex (e.g., consider the development of global digital practices), such a view presupposes a kind of sharp delineation of this field which seems to be impossible.

However, by accepting a contrary conceptualisation of critical academic literacy education as "a historical work in progress" with "no correct or universal models" (Luke, 2012, p. 9), we slip explicitly into "relativism", an admission that critical academic literacy education cannot be delineated as a relatively stable field. Although this might be celebrated in the academic debate, it cannot offer secure criteria regarding which model is critical and which one is not. The delineation of such criteria could inform research aiming to evaluate the application of critical literacy education models,

everyday teaching practices by those wishing to apply critical literacy education and also higher education training in critical literacy education.

CONCLUSION

In this opinion article, I have argued that given the contrasting conceptualisations among different models, critical academic literacy education should be considered a relativist instead of a unified field. Yet, the opinion presented in this article offers a brief, limited and potential picture of how critical academic literacy education can be depicted. Such opinion can be used as a starting point for reflection and as a stimulus for further systematic research. Future research could investigate the application of critical models in practice and collect relevant previous reviews and descriptions of such models which we might not be aware of. Elaborating further on what can count as critical academic literacy education and understanding its complexity within the global/interactional context could fruitfully inform critical academic literacy education practice. In sum, we should rather perceive the field as a type of historical work in progress by making steps not only forward but also backwards.

ACKNOWLEDGEMENTS

I am deeply indebted to the anonymous reviewers for the Journal of Language and Education whose feedback helped me clarify important points of this article as well as to my colleague and collaborator Professor Eleftheria Tseliou who inspired me to write this article.

DECLARATION OF COMPETING INTEREST

None declared.

REFERENCES

Barton, D., Ivanič, R., Appleby, Y., Hodge, R., & Tusting, K. (2007). *Literacy, lives and learning.* Routledge. http://dx.doi.org/10.4324/9780203608302

Bizzell, P. (1992). *Academic discourse and critical consciousness.* University of Pittsburgh Press. http://dx.doi.org/10.2307/j.ctt7zwb7k.9

Blummer, B. (2015). *Some visual literacy initiative in academic institutions: A literature review from 1999 to the present. Journal of Visual Literacy,* 34(1), 1-34. https://doi.org/10.1080/23796529.2015.11674721

Britton, E. R., & Austin, T.Y. (2022). Critical and dominant language learner ideologies: A case study of two Chinese writers' experiences with a critical language writing pedagogy. *TESOL QUARTERLY, 56*(2), 629-655. https://doi.org/10.1002/tesq.3086

Chase, G. (1988). Accommodation, resistance and the politics of student writing. *College Composition and Communication, 39*(1), 13-22. https://www.jstor.org/stable/357812

Clarence, S., & McKenna, S. (2017). Developing academic literacies through understanding the nature of disciplinary knowledge. *London Review of Education, 15*(1), 38-48. https://doi.org/10.18546/LRE.15.1.04

Clark, R. (1992). Principles and Practice of CLA in the Classroom. In N. Fairclough (Ed.), *Critical Language Awareness* (pp. 117-140). Longman. http://dx.doi.org/10.4324/9781315845661

Clark, R., Fairclough, N., Ivanič, R., & Martin-Jones, M. (1991). Critical language awareness. Part II: Towards critical alternatives. *Language and Education, 5*(1), 41–54. https://doi.org/10.1080/09500789109541298

Cohen, J. R. (1998). *Communication criticism: Developing your critical powers*. Sage.

Cope, B., & M. Kalantzis (Eds.). (1993). *The powers of literacy. A genre approach to teaching writing*. The Falmer Press.

Fairclough, N. (1989). *Language and power*. Longman. http://dx.doi.org/10.4324/9781315838250

Fairclough, N. (Ed.). (1992). *Critical language awareness. Longman*. http://dx.doi.org/10.4324/9781315845661

Freebody, P. (2017). Critical-Literacy education: "The Supremely Educational Event". In B. V. Street & S. May (Eds.), *Literacies and language education (Encyclopedia of language and education)* (pp. 95-107). Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-02252-9

Freedman, A., & Medway, P. (1994). Locating genre studies: Antecedents and prospects. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 2-18). Taylor & Francis.

Freire, P. (1972). *Pedagogy of the oppressed*. Penguin Books.

Fu, J., & Wang, Y. (2022). Inspecting EFL teachers' academic literacy development in multilingual contexts: A global vision. *Heliyon, 8*(12), 1-6. http://doi.org/10.1016/j.heliyon.2022.e12143.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. Arnold.

Hankerson, S. (2023). "The world has to stop discriminating against African American Language" (AAL): Exploring the language ideologies of AAL-speaking students in college writing. *Written Communication, 40*(2), 587–619. https://doi.org/10.1177/07410883221146484

Hasrati, M., & Tavakoli, P. (2019). The rise of non-dissertation track master's programmes: An academic literacies approach. *Innovations in Education and Teaching International, 56*(5), 639-651. https://doi.org/10.1080/14703297.2019.1570304

Gee, J. P. (1990). *Social linguistics and literacies: Ideologies in discourses*. The Falmer Press.

Jones, J. (2009). Multiliteracies for academic purposes: Multimodality in textbook and computer-based learning materials in science at university. In A. McCabe, M. O' Donnell, & R. Whittaker (Eds.), *Advances in Language and Education (pp. 103-121). Continuum.*

Kretovics, J. R. (1985). Critical literacy: Challenging the assumptions of mainstream educational theory. *The Journal of Education,* 167(2), 50-62. https://www.jstor.org/stable/42742092

Lea, M. R., & Brian V. S. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education, 23*(2), 157-172. https://doi.org/10.1080/03075079812331380364

Lea, M. R. & Street, B. V. (2000). Student writing and staff feedback in higher education: An academic literacy approach. In M. R. Lea & B. Stierer (Eds.), *Student writing in higher education: New context*s (pp. 32-46). Open University Press – Society for Research into Higher Eduction.

Lea, M. R. & Street, B. V. (2006). The "academic literacies" model: Theory and applications. *Theory into Practice, 45*(4), 368-377. https://doi.org/10.1207/s15430421tip4504_11

Li, D. (2022). A review of academic literacy research development: From 2002 to 2019. *Asian-Pacific Journal of Second and Foreign Language education, 7*(5). https://doi.org/10.1186/s40862-022-00130-z

Luke, A. (2012). Critical literacy: Foundational notes. *Theory Into Practice, 51*(1), 4-11. http://dx.doi.org/10.1080/00405841.2012.636324

Luke, A., & Freebody, P. (1997). Critical literacy and the question of normativity: An introduction. In S. Muspratt, A. Luke, & P. Freebody (Eds.), *Constructing critical literacies: Teaching and learning textual practice* (pp. 118). Allen and Unwin.

McArthur, J. (2010). Time to look anew: critical pedagogy and disciplines within higher education. *Studies in Higher Education, 35(*3), 301-315. https://doi.org/10.1080/03075070903062856

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60-92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Newman, M. (2002). *The designs of academic literacy: A multiliteracies examination of academic achievement.* Bergin & Garvey.

Pennycook, A. (2001). *Critical applied linguistics: A critical introduction.* Lawrence Erlbaum. http://dx.doi.org/10.4324/9781410600790

Rose, D. (2015). New developments in genre-based literacy pedagogy. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 227-242). Guilford.

Street, B. V. (2006). Literacy events and literacy practices: theory and practice in the New Literacy Studies. In M. J. Marily & K. Jones (Eds.), Multilingual Literacies: Reading and writing different worlds (pp. 17-35). John Benjamins.

Swain, E. (2009). Constructing an effective 'voice' in academic discussion writing: An appraisal theory perspective. In A. McCabe, M. O' Donnell, & R. Whittaker (Eds.), *Advances in language and education* (pp. 166-184). Continuum.

Tribble, C., & Wingate, U. (2013). From text to corpus: A genre-based approach to academic literacy instruction. *System, 41*(2), 307-321. https://doi.org/10.1016/j.system.2013.03.001

Wang, Y., & Derakhshan, A. (2023). Enhancing Chinese and Iranian EFL students' willingness to attend classes: The role of teacher confirmation and caring. *Porta Linguarum, 39*(1), 165-192. https://doi.org/10.30827/portalin.vi39.23625

Wang, Y., Derakhshan, A., & Azari Noughabi, M. (2022). The interplay of EFL teachers' immunity, work engagement, and psychological well-being: Evidence from four Asian countries. *Journal of Multilingual and Multicultural Development*. https://doi.org/10.1080/01434632.2022.2092625

White, P. R. (1997). Dialogue and intersubjectivity: Reinterpreting the semantics of modality and hedging. In M. Coulthard, J. Cotterill, & F. Rock (Eds.), *Working with dialogue* (pp. 67-80). Neimeyer.

Yoon, B., & Sharif, R. (2015). History of critical literacy through the various theoretical frameworks. In B. Yoon & R. Sharif (Eds.), *Critical literacy practice applications of critical theory in diverse setting*s (pp. 9-19). Springer.

Yu, S. (2021). Giving genre-based peer feedback in academic writing: sources of knowledge and skills, difficulties and challenges. *Assessment & Evaluation in Higher Education, 46*(1), 36-53. https://doi.org/10.1080/02602930802187316

Zhou, J., Zhao. K., & Dawson, P. (2020). How first-year students perceive and experience assessment of academic literacies. *Assessment & Evaluation in Higher Education, 45*(2), 266-278. https://doi.org/10.1080/02602938.2019.1637513

CONTENT

EDITORIAL

RESEARCH ARTICLES

OPINION PAPERS