ISSN 2411-7390

jle

JOURNAL OF LANGUAGE & EDUCATION

Volume 9 Issue 1, 2023

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

EDITOR-IN-CHIEF

EDITORIAL BOARD

EDITORIAL TEAM

CONTENT

EDITORIAL

RESEARCH ARTICLES

https://doi.org/10.17323/jle.2023.17001

Education 4.0: The Concept, Skills, and Research

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

[1] National Research University Higher School of Economics, Moscow, Russian Federation
[2] Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation
[3] Moscow State Institute of International Relations (MGIMO University), Moscow, Russian Federation

ABSTRACT

Introduction. With Industry 4.0 and Work 4.0 entering the world, modern education is undergoing transformations in terms of educational practices, skillsets and competencies, teaching and learning methodologies (including flipped classroom, blended learning, self-regulated learning, project-based learning, inquiry-based learning, student-centred pedagogy), digital tools used at all educational levels, as well as barriers and challenges. This string of changes is covered by the new buzzword "Education 4.0". It is not so far finally defined. There are various explanations of the concept. Most align with the Fourth Industrial Revolution and Industry 4.0. The JLE editorial aims to overview the emerging research field of Education 4.0 aligned with Industry 4.0, outlining the potential lines of research for JLE authors.

Industry 4.0. The transformation of production at large is beginning on the basis of a set of innovative technologies and completely brand-new processes. Their combination constitutes the underpinning of Industry 4.0.

Skillsets in Education 4.0. There are numerous views of the skills needed for graduates ready for Industry 4.0. One of the most replicated lists embraces the 10 skills offered by the World Economic Forum in 2016 with later updates. Most researchers outline numerous technical, communication, digital, and cognitive skills as a skillset of Industry 4.0.

Teaching and Learning in Education 4.0. The traditional pedagogy or face-to-face learning, still dominant worldwide, is going to combine with innovative approaches, including, e-learning, and blended learning as a mixture of e-learning ang face-to-face learning. In addition, all student-centered technologies add to the future pedagogical landscape: self-regulated learning, project-based learning, flipped classroom, etc.

Research on Education 4.0. The authors searched the Scopus for the documents related to "Education 4.0" to find that the total of 483 results unevenly distributed from 2010 to 2023, with a high of 137 in 2022. The analysis of the publications on Education 4.0 proves that the research field is developing fast, though publications authored by researchers from the developing countries prevail in the search results. At the same time, most of the selected publications came out in the Scopus-indexed low-quartile or discontinued journals. A disproportionately low number of articles published by the authors from the OECD countries depletes the quality of the research field.

Conclusion. The editorial overview of the concept of Education 4.0 may serve as a topical guidance for researchers at large and potential JLE authors focused on educational research. Further studies in the field may cover skillsets and competencies for Industry 4.0; teaching and learning approaches in Education 4.0; new educational frameworks and environments.

KEYWORDS

Industry 4.0, Fourth Industrial Revolution, Education 4.0, skills, competencies, blended learning, e-learning, face-to-face learning.

Citation: Tikhonova E., & Raitskaya L. (2023). Education 4.0: The concept, skills, and research. *Journal of Language and Education, 9*(1), 5-11. https://doi.org/10.17323/jle.2023.17001

Correspondence:
Elena Tikhonova,
etihonova@hse.ru

Received: March 20, 2023
Accepted: March 25, 2023
Published: March 31, 2023

INTRODUCTION

Innovation has been transforming the ways the economy functions throughout the mankind history. In the past, there were periods of great innovations that are marked as industrial revolutions: (1) The Industrial Revolution began when the steam engine was invented by

Thomas Newcomen in 1712, and consequently introduced in various industrial sectors (starting with a steam powered mass-producing print technology). From 1781 on, it brought about Industry 1.0. (Yıldız, 2019). The transition from hand production to machines and "from farming and feudal society to the new manufacturing process" (Xu et al., 2018) marked that period; (2) Industry 2.0. was based on electrical power and assembly lines built at factories and followed the Technological Revolution and the spread of railways and telegraph networks. The invention of the internal combustion engine ushered in a new era of rapid industrialization (Xu *e*t al., 2018). The beginning of the Second Industrial Revolution dates back to the 1860s-1870s. It lasted into the early 20th century till World War I; (3) The second half of the 20th century witnessed flourishing brand-new technologies related to the Internet and accompanying innovations called the Digital Revolution. They transformed the ways of production and were labelled as "Industry 3.0". The greatest achievements of the period were computers and robots; (4) The term "Industry 4.0" has been popularized by the World Economic Forum to describe the trends towards to technologies and processes, including cyber-physical systems, Internet of things, cloud computing, artificial intelligence, computer generated product design, three-dimensional (3D) printing. Industry 4.0 encapsulates a set of technological changes in industry at large. The concept is generally defined via an enumeration of the technologies (Fuchs, 2018). The Third and the Fourth Revolutions have merged while "the lines between physical, digital, and biological spheres" (Xu et al., 2018) are beginning to blur initiated by the amalgamation of Internet, information, and communication technologies (ICTs) and physical machinery (Kumar et al., 2019). Industry 4.0 is often described as the Industrial Internet, or Internet 4.0 (Echeberria, 2020).

Every period of industrial history requires the corresponding form of education. The first three industrial revolutions led to Education 1.0 (teacher-centered approach), Education 2.0 (peer assessment encouraged, high teacher important), and Education 3.0 (co-constructed, first student-centered) respectively (Miranda et al., 2021). At present, Education 4.0 is coming into existence, with new educational technologies based on a range of tools in various environments. It is to response to all the requirements the economy of Industry 4.0 will impose towards employees of the future.

This editorial review overviews the fledging field of Education 4.0, with the objective of outlining the JLE potential key lines of research for further publications. To this end, the JLE editorial dwells upon the background of Industry 4.0, fuelling transformations in Education 4.0 (Industry 4.0); the key aspects of the pedagogy in the Industry 4.0-driven environments (Education 4.0 in Focus); the challenges Industry 4.0 and Work 4.0 make in respect of graduates' employability (Skillsets in Education 4.0), the major pedagogical approaches applied in Education 4.0 and most efficient ways of teaching and learning in online and mixed environments (Teaching and Learning in Education 4.0), and an overview of the research on Education 4.0 to explore the emerging field (Research on Education 4.0).

Industry 4.0

The concept of Industry 4.0 was originally formed in Germany ahead of other countries. It lay a foundation for the European Union (EU) concept accepted in 2016. The Ministry of Education and Research of Germany highlighted that in Industry 4.0 "equipment, machines and single components continuously exchange information" along the value chain[3].

The European Union outlined the key characteristics of Industry 4.0, reduced to interoperability, virtualization, decentralization, real-time capability, service orientation, modularity[4]. Simply put, the concept of Industry 4.0 is an integration of cyber and physical worlds via introducing new technologies (Sony & Naik, 2020). The gist of transformations in Industry 4.0 embraces the following features:

(1) vertical networking in smart production;

(2) horizontal integration via global value chain networks;

(3) through-engineering across the entire value chain;

(4) acceleration of production through exponential technologies (Echeberria, 2020).

Learners within Education 4.0 must get prepared for Industry 4.0 and Work 4.0. The technologies related to Industry 4.0 shorten the life cycle of most professions. The concept ща Education 4.0 marks "the impacts of the systematic changes brought about by the Fourth Industrial Revolution" in education (Oliveira & Saraiva, 2023). It implies "self-regulated learning, critical thinking, collaborative and teamwork skills supported by digital expertise" (Thite et al., 2021).

Skillsets in Education 4.0

Industry 4.0 requires professional who are ready for change, capable of acquiring new knowledge, and improving their skills. Professional education within the concept must be adapted with new models for teaching and learning with a

[3] Bundesministerium für Bildung & Forschung (2013). *Zukunftsbild „Industrie 4.0"*[Federal Ministry for Education and Research]. Bundesministerium für Bildung und Forschung. S.6. https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/zukunftsbild-industrie-4-0.pdf?__blob=publicationFile&v=4

[4] Directorate-General for Internal Policies (2016). *Industry 4.0. Study for the ITRE Committee.* European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2016/570007/IPOL_STU(2016)570007_EN.pdf

focus on interdisciplinary skills (Kipper *et al*., 2021). At present, there is no unanimous skillset covering all potential demands of the future labour market. One of the widely replicated lists is the one offered by the World Economic Forum headlined "The 10 skills you need to thrive in the Fourth Industrial Revolution: Top skills in 2020" (World Economic Forum, 2016 and updates) that include:

(1) complex problem solving;

(2) critical thinking;

(3) creativity;

(4) people management;

(5) coordinating with others;

(6) emotional intelligence;

(7) judgement and decision making;

(8) service orientation;

(9) negotiation;

(10) cognitive flexibility[5].

Many authors highlight other skills, including self-management, flexibility, adaptability, communication skills, initiative, collaborative work, teamwork, interdisciplinary, problem solving, creativity, give and receive feedback, leadership, and other skills (Kipper et al, 2021; Chigbu et al., 2023; Karpenko et al., 2021; Teo et al., 2021).

Employable graduates must possess adequate skills, though some studies suggest that there is "a skill gap between university students and fresh graduates in the industry" (Thite et al., 2021). Education 4.0 aims to bridge such gaps by aligning professional education, especially in engineering, with real production. One of the prominent educational technologies of this kind is a Teaching Factory 4.0. Such factories are created to change over to Engineering Education 4.0, for instance, the SEPT Learning Factory, or a state-of-the-art facility simulating the factory of the future for educational and research aims (Elbestawi et al., 2018). Embedding most important graduate skills, including innovation, critical thinking, metacognition, teamwork, collaboration and communication skills, into Education 4.0, especially in online courses, will require great efforts. The academia at large is striving to incorporating and reflecting on the new technologies, concepts, and paradigms, though sometimes too gradually (Catal & Tekinerdogan, 2019). To address this challenge, algorithms, roadmaps, and other blueprints are put forward. For instance, Coşkun et al. (2019) put up a road map which includes three pillars addressing the curriculum development, laboratory concept, and student club activities. More research is necessary to meet the demands of educations institutions across all domains.

Teaching and Learning in Education 4.0

Education has been based on the classroom approach for all its history (Chigbu et al., 2023). It is still a dominant global technology[6]. The modern pedagogical models applied in teaching and learning are shifting to other approaches, including electronic-based (e-learning or internet-based learning), blended learning (a mixture of face-to-face and electronic-based learning; with instructional modalities – classroom, distance, and self-paced learning; instructional designs; and delivery media), self-regulated learning, project-based learning, revised face-to-face learning to provide for a combination of face-to-face, remote, and online interaction (Jones, & Sharma, 2021). Jones, & Sharma (2021) consider revised face-to-face learning as a set of problem-solving, study groups, presentations, formative assessments, and formative peer evaluations. Traditional lectures turn into interactive lectures that no longer look like a teacher's monologue.

These new pedagogies of Education 4.0 are being adapted unevenly on the global scale, with poor countries being far behind the frontrunners. The economies of the former are not ready for Industry 4.0. Their educational systems are less developed and under-equipped to promote novel educational initiatives (Chigbu *et al*., 2023).

Research on Education 4.0

The search query "Education 4.0" in the titles, abstracts or keywords in the Scopus database brought 483 results (as of March 5, 2023), distributed from 2010 to 2023. Though the first publications date to 2010-2013, the concept of Education 4.0 got a boost not earlier than 2019 after it had been adopted by the European Union (2016) and articulated by the World Economic Forum (2019). The trend towards more publications is definite, with a spike of 137 documents in 2023 (see Fig.1).

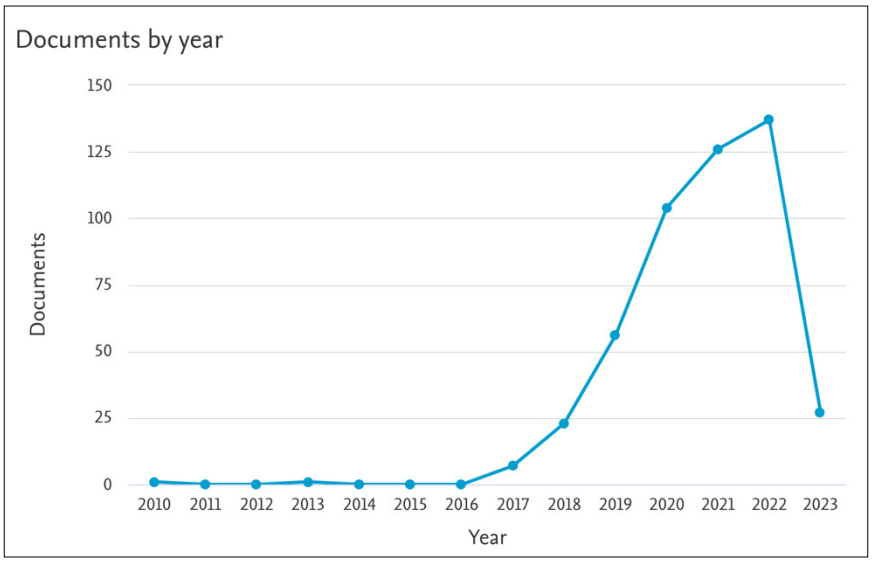
Though, the subject field of Education 4.0 is still getting its shape, there are several reviews on the topic. 16 in the search results, including 8 entirely relevant reviews in English (Ah-

[5] World Economic Forum (2016). The 10 skills you need to thrive in the Fourth Industrial Revolution: Top Skills in 2020. https://www.weforum.org/agenda/2016/01/the-10-skills-you-need-to-thrive-in-the-fourth-industrial-revolution (accessed March 12, 2023).

[6] Marinoni, G., Land, H. Van, & Jensen, T. (2020). *The impact of COVID-19 on higher education around the world: IAU Global Survey Report*. https://www.uniss.it/sites/default/files/news/iau_ covid19_and_he_survey_ report_final_may_2020.pdf.

Figure 1

Research on Education 4.0: Breakdown by Year



Note. Scopus Database as of March 5, 2023. Scopus database. Copyright 2023 by Elsevier.

Table 1

Basic Characteristics of the Search Results on Education 4.0

Characteristics	Number (and/ or Percentage)
Total Search Results	483
Articles	195 (40.4 %)
Review Papers	16 (3.3 %)
Conferences Papers	224 (46.4 %)
Book Chapters	31 (6.4 %)
Journals and Other Sources	205
Authors	318
Most Prolific Author – Ramírez-Montoya, M.S. (Tecnologico de Monterrey, Monterrey, Mexico)	15
One-Authored Documents	57
Top Cited Publication (Benešová & Tupa, 2017)	282
Country with the Highest Results – Malaysia	66
Most Productive Year 2022	137

mad et al., 2022; Butt et al., 2022; Chaka, 2022; Costan et al., 2021; Dao et al., 2023; González-Pérez, & Ramírez-Montoya, 2022; Moraes et al., 2022). Some of the reviews as well as other publications came out in the journals that do not belong to high quartiles or are brought out by the publishers with dubious reputation as a bibliometric review stated (Dao et al., 2023). The bibliometric review on Education 4.0 for 2017-2021 also found that "the research field is very fragmented" (Dao et al., 2023).

The ten most productive countries include Malaysia (n=66), Mexico (n=62), Indonesia (n=60), India (n=44), Germany (n=29), Spain (n=26), Romania (n=21), Brazil (n=19), United States (n=17), and Russian Federation (n=15). As few OECD countries contributed to the research on Education 4.0, the field "cannot develop strongly" (Dao et al., 2023). The contribution on part of many developing countries confirms the fact that their interest in Industry 4.0 and Education 4.0 is outpacing the one of technology-driven countries.

The most frequent keyword "Education 4.0" was included in 367 publications out of the total of 483 (75.98 %). The other most popular keywords were Industry 4.0 (151 publications), student (147), engineering education (145), technology (123), and e-learning (81). Other keywords in our search results covered artificial intelligence, augmented reality, internet of things and others. Almost all most frequent words were in

Figure 2

Word Cloud of the Top 250 Keywords

conformity with the findings in the review mentioned above (Dao et al., 2023).

The thematic spread in the selected publications is rather wide. The most cited articles from the reviewed list cover general features of Education 4.0 (Himmetoglu, Audug, & Bayrak, 2021; Mansor, Abdullah, & Rahman, 20200; requirements for education and qualification in Industry 4.0 (Benešová, & Tupa, 2017); digital transformation towards Education 4.0 (de S Oliveira, & de Souza, 2022; Jain, & Jain, 2022); engineering education (Mogos et al., 2018; Srivani et al., 2022; Polyakova, 2020); and country-related issues of Education 4.0 (Buasuwan, 2018; Anito, & Morales, 2019; Jain, & Jain, 2022). The analysis was subject to the articles only with numerous citations (over 7 as of April 1, 2023).

CONCLUSION

The research on Education 4.0 is on the rise while the Fourth Industrial Revolution is making some evident headway to Industry 4.0. The outline given above may navigate potential researchers through the field. The focus on this prospective field is certain to bring more publications in the near future. It would be reasonable for researchers to analyse national and regional experiences in promoting the concept, compare country-related experiences and the best practices in promoting the appropriate teaching and learning models in the face-to-face, online and mixed environments. Further research may embrace skillsets, both on a national and global scale; competencies of Industry 4.0, specific of various sectors of economy; teaching and learning approaches in Education 4.0; new educational frameworks and environments.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

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.

REFERENCES

Ahmad, I., Sharma, S., Singh, R., Gehlot, A., Priyadarshi, N., & Twala, B. (2022). MOOC 5.0: A roadmap to the future of learning. *Sustainability (Switzerland), 14*(18). https://doi.org/10.3390/su141811199

Anito, J. C., & Morales, M. P. E. (2019). The pedagogical model of Philippine steam education: Drawing implications for the reengineering of Philippine steam learning ecosystem. *Universal Journal of Educational Research, 7*(12), 2662-2669. https://doi.org/10.13189/ujer.2019.071213

Benešová, A., & Tupa, J. (2017). Requirements for education and qualification of people in industry 4.0. *Procedia Manufacturing, 11*, 2195-2202. https://doi.org/10.1016/j.promfg.2017.07.366

Buasuwan, P. (2018). Rethinking Thai higher education for Thailand 4.0. *Asian Education and Development Studies, 7*(2), 157-173. https://doi.org/10.1108/AEDS-07-2017-0072

Butt, R., Siddiqui, H., Soomro, R. A., & Asad, M. M. (2020). Integration of industrial revolution 4.0 and IOTs in academia: A state-of-the-art review on the concept of Education 4.0 in Pakistan. *Interactive Technology and Smart Education, 17*(4), 337-354. https://doi.org/10.1108/ITSE-02-2020-0022

Catal, C., & Tekinerdogan, B. (2019). Aligning education for the life sciences domain to support digitalization and Industry 4.0. *Procedia Computer Science*, *158*, 99–106. https://doi.org/10.1016/j.procs.2019.09.032

Chaka, C. (2022). Is Education 4.0 a sufficient innovative, and disruptive educational trend to promote sustainable open education for higher education institutions? A review of literature trends. *Frontiers in Education, 7.* https://doi.org/10.3389/feduc.2022.824976

Chigbu, B., Ngwevu, V., & Jojo, A. (2023). The effectiveness of innovative pedagogy in the industry 4.0: Educational ecosystem perspective. *Social Sciences & Humanities Open*, *7*(1), 100419. https://doi.org/10.1016/j.ssaho.2023.100419

Coşkun, S., Kayıkcı, Y., & Gençay, E. (2019). Adapting engineering education to industry 4.0 vision. *Technologies, 7*(1). https://doi.org/10.3390/technologies7010010

Costan, E., Gonzales, G., Gonzales, R., Enriquez, L., Costan, F., Suladay, D., Atibing, N., Aro, J., Evangelista, S., Maturan, F., Selerio, E., & Ocampo, L. (2021). Education 4.0 in developing economies: A systematic literature review of implementation barriers and future research agenda. *Sustainability (Switzerland), 13*(22). https://doi.org/10.3390/su132212763

Dao, L. T., Tran, T., Van Le, H., Nguyen, G. N., & Trinh, T. P. T. (2023). A bibliometric analysis of research on education 4.0 during the 2017–2021 period. *Education and Information Technologies, 28*(3), 2437-2453. https://doi.org/10.1007/s10639-022-11211-4

de S. Oliveira, K. K., & de Souza, R. A. C. (2022). Digital transformation towards education 4.0. *Informatics in Education, 21*(2), 283-309. https://doi.org/10.15388/infedu.2022.13

Echeberria, A. L. (2020). *A digital framework for Industry 4.0. managing strategy*. Palgrave Macmillan.

Elbestawi, M., Centea, D., Singh, I., & Wanyama, T. (2018). SEPT learning factory for industry 4.0 education and applied research. *Procedia Manufacturing, 23*, 249–254. https://doi.org/10.1016/j.promfg.2018.04.025

Fuchs, C. (2018). Industry 4.0: The digital german ideology. *Triple C, 16*(1), 280-289. https://doi.org/10.31269/vol16iss1pp280-289

González-Pérez, L. I., & Ramírez-Montoya, M. S. (2022). Components of Education 4.0 in 21st century skills frameworks: Systematic review. *Sustainability (Switzerland), 14*(3). https://doi.org/10.3390/su14031493

Himmetoglu, B., Aydug, D., & Bayrak, C. (2021). Education 4.0: Defining the teacher, the student, and the school manager aspects of the revolution. *Turkish Online Journal of Distance Education, 21*, 12-28. https://doi.org/10.17718/TOJDE.770896

Jain, V., & Jain, P. (2022). From industry 4.0 to education 4.0: Acceptance and use of videoconferencing applications in higher education of Oman. *Journal of Applied Research in Higher Education, 14*(3), 1079-1098. https://doi.org/10.1108/JARHE-10-2020-0378

Jones, K., & Sharma, R. (Eds.) (2021). *Higher education 4.0. The digital transformation of classroomlLectures to blended learning.* Springer.

Karpenko, A., Zasorina, H., & Karpenko, N. (2021). Developing workforce skills for Industry 4.0. In M. Auer & D. Centea (Eds.), *Visions and Concepts for Education 4.0. Proceedings of the 9th International Conference on Interactive Collaborative and Blended Learning (ICBL 2020)* (p. 56-64). Springer. http://doi.org/10.1007/978-3-030-67209-6_7

Kipper, L. M., Iepsen, S., Dal Forno, A. J., Frozza, R., Furstenau, L., Agnes, J., & Cossul, D. (2021). Scientific mapping to identify competencies required by industry 4.0. *Technology in Society, 64.* https://doi.org/10.1016/j.techsoc.2020.101454

Kumar, K., Zindani, D., & Davim, J. P. (2019). *Industry 4.0. Developments towards the Fourth Industrial Revolution*. Springer.

Mansor, N. A., Abdullah, N., & Rahman, H. A. (2020). Towards electronic learning features in education 4.0 environment: Literature study. *Indonesian Journal of Electrical Engineering and Computer Science, 19*(1), 442-450. https://doi.org/10.11591/ijeecs.v19.i1.pp442-450

Miranda, J., Navarrete, C., Noguez, J., Molina-Espinisa, J.-M., Ramirez-Montoya, M.-S., Navarro-Tuch, S. A., Bustamante-Bello, M.-R., Rosas-Fernandez, J.-B., Molina, A. (2021). The core components of education 4.0 in higher education: Three case studies in engineering education*. Computers and Electrical Engineering, 93*, 107278. http://doi.org/10.1016/j.compeleceng.2021.107278

Mogos, R., Bodea, C., Dascalu, M., Safonkina, O., Lazarou, E., Trifan, E., & Nemoianu, I. (2018). Technology enhanced learning for industry 4.0 engineering education. *Revue Roumaine Des Sciences Techniques Serie Electrotechnique Et Energetique, 63*(4), 429-435.

Moraes, E. B., Kipper, L. M., Hackenhaar Kellermann, A. C., Austria, L., Leivas, P., Moraes, J. A. R., & Witczak, M. (2022). Integration of Industry 4.0 technologies with Education 4.0: Advantages for improvements in learning. *Interactive Technology and Smart Education.* https://doi.org/10.1108/ITSE-11-2021-0201

Oliveira S., & Saraiva, M. (2023). Leader skills interpreted in the lens of education 4.0. *Procedia Computer Science*, *217*, 1296-1304. http://doi.org/10.1016/j.procs.2022.12.327

Polyakova, T. (2020). Engineering pedagogy: On the way to "Education 4.0". *International Journal of Engineering Pedagogy, 10*(4), 4-8. https://doi.org/10.3991/ijep.v10i4.15021

Sony, M., & Naik, S. (2020). Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model. *Technology in Society, 61.* https://doi.org/10.1016/j.techsoc.2020.101248

Srivani, V., Hariharasudan, A., Nawaz, N., & Ratajczak, S. (2022). Impact of education 4.0 among engineering students for learning English language. *PLoS ONE, 17*(2 February) https://doi.org/10.1371/journal.pone.0261717

Suyudi, Nugroho, B. S., El Widdah, M., Suryana, A. T., Ibrahim, T., Humaira, M. A., Nasrudin, M., Mubarok, M. S., Abadi, M. T., Adisti, A. R., Gadzalia, S. S., Muqtada, M. R., Purwanto, A., Fahlevi, M., & Sudargini, Y. (2020). Effect of leadership style toward Indonesian education performance in Education 4.0 era: A schematic literature review. *Systematic Reviews in Pharmacy, 11*(10), 371-378. https://doi.org/10.31838/srp.2020.10.60

Teo, T., Unwin, S., Scherer, R., & Gardiner, V. (2021). Initial teacher training for twenty-first century skills in the fourth industrial revolution (IR 4.0): A scoping review. *Computers and Education, 170*. https://doi.org/10.1016/j.compedu.2021.104223

Thite, S., Ravishankar, J., Ambikairajah, E., & Ortiz, F.M. (2021). Work in progress: "Embedding graduate skills in online courses". In M. Auer & D. Centea (Eds.), *Visions and Concepts for Education 4.0. Proceedings of the 9th International Conference on Interactive Collaborative and Blended Learning (ICBL 2020)* (p. 56-64). Springer. http://doi.org/10.1007/978-3-030-67209-6_11

Xu, M., David, J. M., & Kim, S. H. (2018). The fourth industrial revolution: Opportunities and challenges. *International Journal of Financial Research, 9*(2), 90-95. https://doi.org/10.5430/ijfr.v9n2p90

Yildiz, T. (2019). Examining the concept of industry 4.0 studies using text mining and scientific mapping method. *Procedia Computer Science, 158,* 498-507. https://doi.org/10.1016/j.procs.2019.09.081

https://doi.org/10.17323/jle.2023.11045

Automated Measures of Lexical Sophistication: Predicting Proficiency in an Integrated Academic Writing Task

Randy Appel [1], **Angel Arias** [2]

[1] Waseda University, Tokyo, Japan
[2] Carleton University, Ottawa, Canada

ABSTRACT

Background. Advances in automated analyses of written discourse have made available a wide range of indices that can be used to better understand linguistic features present in language users' discourse and the relationships these metrics hold with human raters' assessments of writing.

Purpose. The present study extends previous research in this area by using the TAALES 2.2 software application to automatically extract 484 single and multi-word metrics of lexical sophistication to examine their relationship with differences in assessed L2 English writing proficiency.

Methods. Using a graded corpus of timed, integrated essays from a major academic English language test, correlations and multiple regressions were used to identify specific metrics that best predict L2 English writing proficiency scores.

Results. The most parsimonious regression model yielded four-predictor variables, with total word count, orthographic neighborhood frequency, lexical decision time, and word naming response time accounting for 36% of total explained variance.

Conclusion. Results emphasize the importance of writing fluency (by way of total word count) in assessments of this kind. Thus, learners looking to improve writing proficiency may find benefit from writing activities aimed at increasing speed of production. Furthermore, despite a substantial amount of variance explained by the final regression model, findings suggest the need for a wider range of metrics that tap into additional aspects of writing proficiency.

KEYWORDS

second language writing, automated analysis, corpus, integrated writing, L2 English

INTRODUCTION

Over the past three decades, advances in computer aided corpus linguistics have enabled new approaches to discourse analysis that were previously either too time consuming or unreliable to be considered methodologically valid (e.g., Crossley, 2020). Using these relatively new tools, researchers can now apply standardized, and therefore replicable, assessments of a wide range of linguistic metrics to assess various important constructs (e.g., writing proficiency), thus providing important insights into the language produced by first (L1) and second (L2) language users (e.g., Casal & Lee, 2019; Kyle & Crossley, 2017). For instance, Lu (2011) implemented the Lexical Syntactic Complexity Analyzer to examine lexical characteristics of Chinese learners and found several measures of clause complexity and length to be strongly associated with writing ability.

The advent and development of automated text analysis have also inspired research that has examined single and multiword measures in both independent and source-based prompts. In this regard, multiword measures such as n-grams have shown to be important predictors of essay quality in independent tasks but less so in source-based tasks (Kyle & Crossley, 2016). Other text

Citation: Appel R., & Arias A. (2023). Automated measures of lexical sophistication: Predicting proficiency in an integrated academic writing task. *Journal of Language and Education, 9*(1), 12-25. https://doi.org/10.17323/jle.2023.11045

Correspondence:
Randy Appel,
r_appel@waseda.jp

Received: July 01, 2020
Accepted: September 20, 2022
Published: March 31, 2023

analysis tools such as Coh-metrix (Graesser et al., 2004) and TAALES (Kyle & Crossley, 2015) allow researchers to extract a wide range of linguistic measures that aid in the exploration of texts in unprecedented ways. This new world of computer aided linguistic analysis has led to increased interest in automated forms of assessment that could be used to decrease the burden placed on teachers and increase student independence by creating alternative paths for feedback and evaluation. However, to take advantage of the advances being made and get a more complete picture of how well these metrics can be used to model raters' assessment of L2 discourse, it is important to continue pushing the field of study forward by examining an ever-wider range of writing tasks and linguistic metrics.

In particular, as it relates to source-based, L2 English academic writing (i.e., academic tasks that require writers to integrate material from previously exposed texts and/or listening materials), a very narrow range of task types has thus far been examined, with a frequent focus on summary tasks from the Test of English as a Foreign Language (e.g., Guo, Crossley, & McNamara, 2013; Kim & Crossley, 2018; Kyle, 2017; Kyle & Crossley, 2016; Plakans *et al*., 2019). The TOEFL iBT is an important resource, yet the availability of this data set seems to have created an overreliance on this task and a related lack of research targeting alternative source-based writing. Thus, we lack a full understanding of the effectiveness of automatically extracted linguistic measures in highlighting the features most relevant to raters' assessments of L2 writing proficiency in source-based writing beyond this task.

The present study aimed to extend previous research into the automated analysis of source-based L2 English written discourse by using an unpublished corpus of timed, argumentative, L2 English academic essays collected from a former test of academic English proficiency commonly used to evaluate international students aiming to study at English-medium post-secondary institutions in Canada (Appel & Wood, 2016)[3]. To account for factors that may influence the language being produced, this study controls for essay type, topic, and testing conditions by using a single version (i.e., one writing prompt) of this standardized test. To better understand the underlying discoursal features relevant to human raters' assessments of academic English writing proficiency in this test, the TAALES 2.2 software suit was used to extract 484 single and multi-word candidate measures of lexical sophistication and subsequently examine their potential relationship with holistic scores of this source-based, L2 English writing proficiency test[4].

Empirical Analyses of L2 English Writing Proficiency

The link between linguistic features and writing proficiency dates back to at least the 1970s (Crossley, 2020), with empirical analyses largely taking hold in the 1990s (Wolfe-Quintero *et al*., 1998). The main goal in many of these studies has been to reveal the features most relevant to differences in assessed proficiency, and therefore which elements should receive greater focus during target language instruction. While early examinations of the lexical and grammatical features in L2 discourse indicated specific linguistic elements relevant to differences in assessed proficiency, these studies were often limited to the examination of a single or limited number of metrics due to the time consuming and laborious nature of the manual analysis. Therefore, knowledge of how various lexical factors work in combination were difficult to come by, and reliance on human coding created the potential problem of a lack of consistency in evaluations. However, due to recent advances in computer technology and associated forms of automated analyses, these problems can largely be avoided. In fact, we are now able to consistently and accurately extract a wide range of lexical measures from submitted texts without the need for human intervention.

Automated Analyses of L2 Writing

Automated computer assisted analyses have enhanced our collective understanding regarding the lexical, phrasal, cohesive, and syntactic features most relevant to evaluations of differences in linguistic ability (e.g., Bi & Jiang, 2020; Lei *et al*., 2023; Li *et al*., 2023). For example, Lu (2011) used the automated extraction of 14 syntactic complexity measures to analyze 3,554 independent, argumentative essays from the Written English Corpus of Chinese Learners (Wen, Wang, & Liang, 2005). Findings indicated that a combination of 7 discoursal features (*complex nominals per clause*, *complex nominals per T-unit*, *coordinate phrases per T-unit*, *coordinate phrases per clause*, *mean length of clause*, *mean length of sentence*, *mean length of T-unit*) were the best predictors of differences in perceived writing ability.

While the increased speed and accuracy of analysis provided by automated extraction has led to a greater dependence on these tools, their increasing acceptance and application in recent years is likely, at least somewhat, related to their free (e.g., Compleat Lexical Tutor, Coh-metrix, AntConc), or low cost (e.g., WordSmith Tools, LIWC) nature. For example, the 14 lexical measures used by Lu (2011) are now freely available through an online tool, the Web-based L2 Syntactical Complexity Analyzer. Thus, additional researchers are

[3] This study was based on a retired version of the Canadian Academic English Language (CAEL) assessment. The current version of this test maintains the core construct of academic language proficiency yet uses a different structure from the version in the current research.

[4] This study was supported by Paragon Testing Enterprises (now a subsidiary of Prometric). The analyses and conclusions presented in this paper are those of the authors and do not reflect the views or positions of Prometric.

able to make use of these measures to verify the findings of published studies using alternative corpora.

Coh-metrix, another popular and freely available web-based text analysis tool, allows users to submit discourse samples for automatic extraction of over 100 lexical metrics spanning 11 broad categories (*descriptive statistics*, *text easability*, *referential cohesion*, *latent semantic analysis*, *lexical diversity*, *connectives*, *situation model*, *syntactic complexity*, *syntactic pattern density*, *word information*, *readability*). Similarly, Compleat Lexical Tutor offers users a wide range of freely available linguistic tools that can be used to calculate, among others, metrics related to word frequency in reference to a range of previously compiled corpora. While each of these tools offer easily accessible methods to evaluate the language appearing in the corpora under investigation, the range of measures available has remained relatively limited. However, with the release of the Tool for the Automatic Analysis of Lexical Sophistication (TAALES), this is no longer the case.

TAALES

Originally released in 2015, TAALES (Kyle & Crossley, 2015) is a downloadable software application that allows users to batch analyze large collections of text for a wide range of lexical features. As a downloadable application, TAALES holds a distinct advantage over alternative tools (e.g., Coh-metrix, Web-based Syntactic Complexity Analyzer, Compleat Lexical Tutor) in that it can be used offline and does not require each text to be submitted individually, thus allowing for a faster and more efficient form of analysis. Although initially limited to 130 indices, with the release of version 2.2, TAALES now supports the extraction of 484 single and multi-word metrics, thereby making it one of the most comprehensive freely accessible linguistic analysis tools available. With a focus on lexical sophistication, these measures cover 10 broad categories (discussed in greater detail below and fully described at https://www.linguisticanalysistools.org).

Automated assessments of Source-based Writing

With regard to the writing tasks examined in previous automated analyses of L2 English proficiency, an important distinction must be made between independent and source-based (i.e., integrated) tasks. The former is a task type that relies purely on the writers' personal experience to contribute content to the composition. For example, personal narratives (e.g., *Describe a time you had to deal with conflict and how you resolved it*) as well as opinion pieces (e.g., *Do you think it is necessary for students to work part-time while attending university?*) would both fall under this category when composed without reference to outside sources.

These independent tasks are easy to administer since previous exposure to reading and listening materials is not required and the prompt can be provided to writers without any time-consuming preparation. Perhaps because of the relative ease associated with this type of data collection, independent writing has served as the primary focus in the majority of previous studies aiming to automatically extract relevant linguistic features that could be used to model human raters' assessment of L2 English discourse (Guo *et al*., 2013). Findings based on these analyses of independent writing include the identification of text length (e.g., Ferris, 1994) lexical diversity (Crossley & McNamara, 2012), and average length per word (e.g., Grant & Ginther, 2000) as important factors in assessments of L2 English proficiency.

Despite the wide-spread use of independent tasks in automated analyses of L2 English, this type of writing is a poor representation of the real-world, post-secondary assignments students will be required to complete, since 'it is impossible to assign academic writing tasks that don't require preliminary readings' (Johns, 1993, p. 277). Therefore, to better understand how more academically minded tasks are evaluated by human raters, recent years have seen an increasing focus on source-based writing. For instance, Guo et al. (2013) used Coh-metrix to analyze summary tasks from the TOEFL iBT public use data set in relation to holistically assigned proficiency scores. Multiple regressions were used to identify seven indices (*number of words per text, past participle verbs, word familiarity, verbs in 3rd person singular present, semantic similarity, verbs in base form, and word frequency*) that accounted for over 50% of the variance in assessed proficiency on a 5-point scale.

Several studies have also examined the influence of automated measures of text cohesion on writing development. Although these studies initially focused on L1 discourse, an increasing amount of research has adopted an L2 focus (e.g., Casal & Lee, 2019; Guo et al., 2013; Kim & Crossley, 2018). In terms of automated measures of discourse cohesion in relation to L2 English writing quality, Kim and Crossley (2018) found that lexical overlap between paragraphs, a global cohesion metric, was a significant predictor of writing quality for both independent and integrated TOEFL iBT tasks. Also making use of the TOEFL iBT, Guo et al. (2013) revealed that variance in writing quality on integrated tasks could be successfully predicted by local measures of discourse cohesion.

Similar studies to the above have also been conducted by Kyle et al. (2016), Plakans et al. (2019), and Kyle (2017) with a range of lexical factors identified and variance in proficiency scores accounted for. However, one potentially limiting factor in this body of research is that the majority of data comes from a single source: the TOEFL iBT. This source-based task is a quick, written exercise that involves a short lecture and text on the same topic, followed by a 20-minute allotment in which the test taker is asked to produce a summary of the previously exposed to materials. Therefore, although technically source-based, this task type removes any need to present a position and argue it in an effective manner using support from outside sources. Thus, this research approach has continued to neglect source-based ar-

gumentative writing, a task type that has been repeatedly identified as a key component of the undergraduate writing genre (e.g., Wingate, 2012).

Research Question

The present study aimed to extend previous research targeting automated measures of writing development in L2 English discourse. Of the three broad constructs commonly featured in research of this kind (i.e., lexis, syntax, and cohesion), we have chosen lexical sophistication as our focus, as this has been argued as presenting the richest measures of writing development (Crossley, 2020). With an emphasis on a single construct of writing quality (i.e., lexical sophistication), we also aim to use a wider range of measures than those commonly applied in this area. Furthermore, while our objective is to expand the number of metrics used to assess lexical sophistication in order to better assess their value in predicting writing quality, we also narrow our focus by targeting a tightly controlled corpus of L2 English writing that shares a common writing prompt, topic, and task, with all samples produced under similar writing conditions.

The composition of this corpus is also important since it serves as an alternative source of integrated, timed, L2 English academic writing designed to better represent the type of task students will find necessary when studying in English medium post-secondary institutions. As a result, a main goal of the present research is to better identify which specific metrics of lexical sophistication are most associated with differences in assessed writing proficiency in academic tasks of this nature. Put differently, the purpose of this study was to explore potential answers to the following research question: Which automatically extracted single and multi-word metrics of lexical sophistication can be used to predict holistic ratings of L2 English writing proficiency on a timed, source-based writing task?

METHOD

Corpus

The corpus analyzed in this study is composed of timed, written responses (n = 589) to a single argumentative essay prompt included as part of a four section (reading response, lecture response, oral language response, written response) timed, high-stakes language test. The final portion of this test, the written response, provides the data used in this study. This test aims to assess L2 English learners' ability to use academic English as needed in order to fully participate in English-medium post-secondary institutions. As the final component in a multi-section linguistic evaluation on a single topic, writers are encouraged to make use of previous components of the test (e.g., lecture and reading texts) to help strengthen the position presented in their essays.

This test used a nine-band grading system, with 10 denoting the lowest level of writing proficiency and 90 indicating the highest assessed level (Appendix A). To ensure grade consistency, each essay was assessed by a team of three trained raters using a collaborative, read-aloud protocol. If all three raters agree on the grade the essay should receive, the score is recorded, and the process begins again with the following essay. In cases of disagreement among any of the three raters, discussions are initiated until a consensus is reached. The collaborative read-aloud protocol used to evaluate these essays is viewed as decreasing the focus on surface features, thereby encouraging raters to be more attentive to more substantive aspects of each piece of writing, such as organization, cohesion, and coherence of the presented argument. Thus, differences in essay length may be an important factor in ratings for this timed task, as greater essay length may be seen as providing evaluators with more opportunities to assess the macro features of each text. This emphasis on macro features contrasts with an overt focus on grammatical inaccuracies that may not interfere with the overall structure, organization, and message of each piece of writing[5].

Measures of Lexical Sophistication

In total, 484 indices of lexical sophistication were selected as candidate measures and included in the present study. These measures spanned 10 broad categories, which are summarized below and were selected since "lexical sophistication tends to provide the richest metrics of text quality" (Crossley, 2020, p. 417). For additional details on each category and the individual indices contained within them, see Kyle and Crossley (2015).

Word Frequency

Word frequency scores in TAALES are sourced from metrics related to several major corpora, including the Thorndike-Lorge magazine corpus (Thorndike & Lorge, 1944), Kucera-Francis (Kucera & Francis, 1967), Brown Corpus (Kucera & Francis, 1967), British National Corpus (BNC[6], 2007), Corpus of Contemporary American English (COCA[7], 2008), and SUBTLEXus (Brysbaert & New, 2009). Vocabulary in the submitted texts that does not appear in any of the corresponding corpora is excluded from the final frequency

[5] All essay ratings were provided by Paragon. However, the current version of the CAEL assessment uses a rating method that differs from the one described in this manuscript.

[6] British National Corpus. (2007). British National Corpus, version 3 (BNC XML ed.).

[7] Davies, M. (2008). The Corpus of Contemporary American English: 520 million words, 1990–present. Available online at https://corpus.byu.edu/coca/

counts. Indices from this category include raw frequency, mean frequency, and logarithmic frequency, with options allowing for individual reports or measures related to content words, function words, or all words (i.e., combination of content and function words).

Word Range

Word range information for several of the aforementioned corpora (e.g., Brown, BNC, SUBTLEXus, COCA) is also provided. As with word frequency, these metrics come in raw and logarithmic forms, with the option to restrict scores to content words, function words, or all words.

Contextual Distinctiveness

These values are based on the results from previous research, such as the Edinburgh Associative Thesaurus (Kiss et al., 1973), McDonald's co-occurrence probability (McDonald & Shillcock, 2001), and Latent Semantic Analysis values (Landauer *et al*., 2007).

Psycholinguistic Word Information

Largely based on information from the MRC psycholinguistic database (Coltheart, 1981), a collection of human ratings on more than 26 psychological properties from over 150,000 words, TAALES includes measures of familiarity, imageability, concreteness, and meaningfulness.

Word Neighborhood

Using scores from the English Lexicon Project (Balota et al., 2007), word neighborhood scores are reported for orthographic, phonographic, and phonological neighbors. For example, orthographic neighbors are calculated as the number of words that can be formed by changing only one letter in the target vocabulary item (e.g., *past* and *last*).

Word Recognition

Metrics from this category are calculated based on the norms reported in the English Lexicon Project (Balota et al., 2007). In general, longer word recognition times can be viewed as an indication of increased processing difficulty.

Age of Exposure

Indices from this category are based on data from the TASA corpus (Dascalu et al., 2016). Low age of exposure scores reflect texts with a higher proportion of words from lower grade levels, while texts earning higher scores are indicative of texts with more words from higher grades.

Semantic Relationships

Values in this category include hypernymy and polysemy norms from WordNet (Fellbaum, 1998), which contain information on connections between nouns, verbs, adjectives, and adverbs. Texts with high hypernymy tend to be more specific, have more concrete terms, and have fewer abstract words. High polysemy indicates texts with more sense relationships.

N-Gram Indices

Indices from this category are largely rooted in the bigram and trigram frequencies for the written and spoken sections of the BNC corpus created by Crossley et al. (2012). The proportion of n-grams is a measure of the percentage of unique n-grams present in the text that are also found in the reference corpora. TAALES 2.2 also includes five measures of association strength derived from the COCA: Mutual Information (MI), MI^2, t-score, ΔP, and approximate collexeme score.

Academic Language

Measures from this category aim to identify the prevalence of academic lexis in each submitted text using the Academic Word List (AWL; Coxhead, 2000) and the Academic Formulas List (Simpson-Vlach & Ellis, 2010) as reference points.

Another category included in this study was text length because "it can be considered especially construct-relevant when it comes to writing in a foreign language" (Fleckenstein et al., 2020, p. 2). In argumentative essays, particularly the timed ones analyzed in this study, using a wider range of linguistic resources can result in longer texts, which in turn can be arguably considered as indicative of higher proficiency.

Analysis

The identification of lexical factors associated with holistic judgments of L2 English writing proficiency on this timed, integrated task, followed a three-step process. First, Pearson correlation coefficients were used to explore the associations between the 484 candidate variables and the target construct of assessed writing proficiency. Second, those lexical metrics with the strongest correlations were assessed for redundancy (i.e., highly correlated predictor variables), with the goal of eliminating redundant lexical variables. Third, after examining redundancy, the remaining lexical measures were included as predictor variables in a stepwise multiple regression to highlight which of the lexical features identified in step two accounted for significant proportions of variance in human raters' assessments of holistic writing ability[8].

[8] Wherever possible, the decision to retain a variable was based on the strength of correlation. In other words, in each pair of collinear variables, the variable holding the strongest relation with assessed proficiency score was retained.

Pearson Correlations

As a first step in assessing the predictive strength of the 484 TAALES measures in accounting for differences in L2 English writing proficiency, using SPSS version 25 (IBM SPSS Statistics, 1989 – 2017), initial Pearson correlations were run between the full range of measures and the score awarded to each essay. Before running any correlations, all data were analyzed to ensure relatively normal distributions, a lack of outliers, and the absence of missing data. Adhering to Plonsky and Oswald's (2014) localized recommendations for meaningful correlations in L2 research ($r \geq .250$) resulted in a list of variables ($k = 36$) associated with assessed L2 English writing proficiency. This initial group of 36 variables included various metric options within the same category (e.g., Brown Frequency All Words [AW] and Brown Frequency Content Word [CW]). Thus, to avoid statistical dependencies between options "all words", "content words" and "function words" within the same category, only the option sharing the strongest correlation with band score was considered for further analysis. Following previous research and recommendations (e.g., Appel *et al*., 2019; Hinkle *et al*., 2003), to account for the existence of multicollinearity among these significantly correlated predictor variables, in any case where two predictor variables held a minimum correlation of .700 with each other, only one was retained. This helped eliminate the possibility of including multiple overlapping measures in the final step of the analysis (i.e., multiple regression). Following the identification of significantly correlated predictor variables and the elimination of multicollinearity, the final step was to build a multiple regression model to better understand the value of the correlated variables in accounting for differences in L2 English writing proficiency.

Multiple Linear Regression

Given the primary focus of this study was to predict differences in assessed L2 English academic writing proficiency based on single and multi-word metrics, a stepwise multiple regression model was subsequently fit to the data. This type of regression is appropriate for prediction (Keith, 2019) and predictor variables are handled automatically by the software package based on statistical criteria and techniques (Plonsky & Ghanbar, 2018). For this study, the technique of forward selection was used, consisting of first selecting the predictor variable that shared the strongest correlation with the criterion variable and then selecting the next variable with the highest contribution to the model on an iterative basis.

As in other types of multiple linear regression analyses, statistical assumptions were examined and met prior to the application of this statistical procedure to make valid interpretations and inferences from the data. In this regard, the data were screened for influential data points, and after fitting the regression model, the residuals were used to verify their normality and homoscedasticity. The model's residuals were normally distributed, but the homogeneity of variance was slightly violated. However, the variance of Y for each value of X was generally constant and "conditional on the level of each of the predictor variables and on the overall $\hat{Y}$ from the final regression equation" (Howell, 2013, p. 536). Furthermore, the variance inflation factor (≤ 2, see Table 3) indicated no collinearity issues. Additionally, Cook's distances (≤ 0.040) did not raise any concerns regarding outlying measures, thus preserving the model's integrity.

RESULTS

Correlations between TAALES Variables and Assessed Writing Proficiency

As a first step in our results, we provide Pearson correlation coefficients between holistically assessed writing score and the candidate predictive measures of writing proficiency (Table 1). These initial correlations were used to assess the value of each measure as a predictor variable and the amount of collinearity between variables that could be identified. After accounting for multicollinearity (i.e., variables that were too closely related), six predictor variables remained (Table 2). These six predictor variables were subsequently included in the regression model. These variables cover four of the previously mentioned categories (word recognition, word neighborhood information, word concreteness, word frequency), as well as one additional category (sample length).

As can be seen in Table 2, word count, word naming response time, and lexical decision time held positive relationships with assessed writing proficiency. Conversely, orthographic neighborhood frequency, unigram familiarity, and HAL frequency of closest phonological neighbors maintained negative correlations. The strongest correlation with assessed proficiency was for word count with a correlation of .516.

Lexical Predictors of Writing Proficiency

The regression model yielded four predictor variables free of collinear issues (see variance inflation factor): *total word count*, *orthographic neighborhood frequency*, *lexical decision time*, and *word naming response time*, suggesting that the four-predictor model was statistically significant, $F(4, 584) = 84.668$, $p < .001$ and accounted for approximately 36% of the variance in writing proficiency scores ($R^2 = .367$, adjusted R^2 36.3%) for the integrated writing task. *Total word count* explained the largest portion of the variance associated with writing proficiency, thus exhibiting the strongest weight in the model ($\beta = .452$, $t = 13.428$, $p < .001$), followed by *word naming response time* ($\beta = .151$, $t = 3.352$, $p = .001$), *orthographic neighborhood frequency* ($\beta = -.125$, $t = -3.22$, $p = .001$) and *lexical decision time* ($\beta = .120$, $t = 2.908$, $p = .004$). Note, however, that the predictor *orthographic neighborhood frequency* was negatively associated with writing proficiency scores.

Table 1

Pearson Correlations between Writing Score and Potential Predictive Lexical Measures

Variables	Score	WC	WNRTz CW	WNRT CW	WNRT	WNRTz	LDTsd CW	LDT	LDT CW
WC	0.52								
WNRTz CW	0.37	0.16							
WNRT CW	0.35	0.14	0.98						
WNRT	0.34	0.14	0.94	0.95					
WNRTz	0.33	0.15	0.95	0.94	0.98				
LDTsd CW	0.34	0.19	0.61	0.60	0.58	0.58			
LDT	0.34	0.13	0.84	0.85	0.87	0.85	0.68		
LDT CW	0.34	0.14	0.88	0.88	0.84	0.84	0.71	0.95	
ALD CW	0.32	0.14	0.85	0.85	0.81	0.81	0.57	0.82	0.86
LDTz	0.33	0.11	0.82	0.83	0.86	0.84	0.61	0.98	0.92
MRC AW	-0.34	-0.18	-0.65	-0.64	-0.67	-0.65	-0.47	-0.64	-0.61
HAL CW	-0.34	-0.15	-0.83	-0.82	-0.80	-0.81	-0.57	-0.78	-0.81
HAL	-0.32	-0.14	-0.64	-0.63	-0.70	-0.69	-0.48	-0.66	-0.62
ONF CW	-0.32	-0.16	-0.54	-0.52	-0.53	-0.54	-0.34	-0.52	-0.53

Variables	ALD CW	LDTz	MRC AW	HAL CW	HAL
WC					
WNRTz CW					
WNRT CW					
WNRT					
WNRTz					
LDTsd CW					
LDT					
LDT CW					
ALD CW					
LDTz	0.81				
MRC AW	-0.58	-0.62			
HAL CW	-0.90	-0.76	0.62		
HAL	-0.70	-0.64	0.64	0.79	
ONF CW	-0.48	-0.51	0.40	0.62	0.55

Note: WC = Word Count; WNRTz CW = Word Name Response Time (z-score) CW; WNRT CW = Word Naming Response Time CW; WNRT = Word Naming Response Time; WNRTz = Word Naming Response Time (z-score); LDTsd CW = Lexical Decision Time (standard deviation) CW; LDT = Lexical Decision Time; LDT CW = Lexical Decision Time CW; ALD CW = Average Levenshtein Distance of closest phonological neighbors CW; LDTz = Lexical Decision Time (z-score); MRC AW = MRC Familiarity CW; HAL CW = Average Log HAL frequency of closest orthographic neighbors CW; HAL = Average Log HAL frequency of closest orthographic neighbors; ONF CW = Orthographic Neighborhood Frequency CW; AW = All words; CW = Content Words.

DISCUSSION

The current study explored the relationship between a wide range of automatically extracted measures of lexical sophistication and assessments of L2 English proficiency on a timed, integrated, academic writing task. While the initial correlations used to identify metrics associated with differences in perceived L2 English writing proficiency yielded a somewhat unwieldy list of variables, eliminating weak correlations and multi-collinearity helped reduce the number of candidate variables to a more manageable level. The final multiple regression yielded a four-factor model accounting for 36.3% of variance (adjusted R^2) in assessed proficiency. Each of these four metrics is discussed in detail below.

Lexical Correlates Accounting for Significant Variance in Assessed Proficiency Scores

Total Word Count

The importance of *total word count*, which held the strongest correlation with assessed proficiency (r = 0.516) and accounted for the largest portion of variance (26.5%) in test scores, supports findings from previous studies (e.g., Kamimura & Oi, 2001; Sasaki, 2000) regarding developmental trends in L2 English writing proficiency using independent writing tasks. Although each of these studies has suggested a link between length of sample and assessed proficiency in independent tasks, results from the current study extend these findings to integrated, argumentative academic writing tasks produced under testing conditions. Although text length was the main predictor of proficiency, further research should include other text measures to continue exploring the construct complexities of second language writing. More recently, Crossley (2020) decided not to include text length in a systematic review of linguistic features in writing quality, but acknowledged that text length is likely the strongest predictor of writing development quality. Similarly, other studies have found a positive relationship between text length and human ratings, while controlling for language proficiency (McNamara et al., 2015).

The relationship between volume of output and proficiency in this study likely relates to the importance of writing fluency (i.e., speed of production) in assessments of this kind (e.g., timed tasks produced under testing conditions). In timed tasks, lower-level writers may lack sufficient opportunities to plan and compose their essays, thereby leading to reduced output volume. Thus, the ability to quickly produce discourse within a limited time may be interpreted as an important factor that significantly contributes to rater assessments of proficiency in L2 English writing.

The link between longer essays and perceptions of higher proficiency may also relate to the fact that longer texts give assessors more material on which to base their evaluation by increasing writers' opportunities to display more diverse

Table 2

Statistically Significant Correlations after Accounting for Multicollinearity

Category	Variable	Correlation
Sample Length	Word Count	0.516
Word Recognition	Word Naming Response Time – Content Words (z-score)	0.368
	Lexical Decision Time – Content Words (standard deviation)	0.338
Word Neighbor Information	Orthographic Neighborhood Frequency – Content Words	-0.316
Word Concreteness	Unigram Familiarity (mean)	-0.341
Word Frequency	HAL Frequency of Closest phonological neighbors	-0.315

Table 3

Results of the Multiple Regression

Variable	R^2	$R^2(adj)$	β	*t*	*p*	VIF
Word Count	0.267	0.265	0.452	13.428	0.000	1.047
Word Naming Response Time (z-score)	0.346	0.344	0.151	3.352	0.001	1.884
Orthographic Neighborhood Frequency (content words)	0.358	0.355	-0.125	-3.22	0.001	1.383
Lexical Decision (content words [standard deviation])	0.367	0.363	0.120	2.908	0.004	1.582

and linguistically advanced structures. In contrast, short samples may not enable writers to produce as wide an array of linguistic elements. For example, low-level writers (band scores of 10-20) are described as using 'restricted language' (Appendix A); without a sufficient volume of writing, it may be challenging for writers at this level to demonstrate the ability to move beyond this limited range of lexical items.

In longer texts, however, raters are more likely to find language instances that match the associated descriptors, regardless of occasional slips or mistakes that may be present. In fact, essays receiving a score of 30 are listed as 'unable to develop ideas' while adept writers (scoring at band 70) are described as using 'the readings and lecture effectively to support the thesis'. Thus, while it is clear that volume of output has a clear relationship with perceptions of L2 ability on this task, this finding may be more broadly linked to the ability of writers to express their ideas using varied lexis and present their position using adequate support from outside sources.

Word Naming Response Time (Z-Score)

Word naming response time contributed approximately 8% additional variance accounted for in the final regression model. This variable is a psycholinguistic measure of the accuracy and time native English speakers require to read aloud words in a given text. To calculate figures for this variable, L1 English speakers were presented with individual word or non-word samples and asked to read each aloud, with their responses being recorded and aggregated (Balota et al., 2007).

Potential reasons for the importance of this measure in assessments of perceived proficiency include the fact that words with longer naming response times and lower accuracy on word naming correctness decisions are generally seen as more sophisticated and thus indicative of more advanced levels of linguistic ability. For instance, both Kim *et al*. (2018) and Kyle *et al*. (2017) found a positive association between word naming response times and holistically scored independent writing tasks. Furthermore, recent research on L2 English speech by Berger et al. (2017) found that more frequent use of words with longer response times was associated with higher perceived proficiency among L2 English users. Findings from the current study echo these findings, thus bringing further support to the notion that the more frequent use of less accurately named words with longer word naming response times is a sign of higher L2 English proficiency in both oral and written communication.

Orthographic Neighborhood Frequency - Content Words

As previously mentioned, this variable provides a summary score for the number of orthographic neighbors for individual words in each sample. When specifically related to content words, as applied here, this metric is likely a truer measure of the number of orthographic neighbors present in each text since function words (e.g., auxiliaries, prepositions, quantifiers, pronouns) are excluded. Importantly, as evidenced by the correlations and multiple regression, this metric holds a negative relationship with assessed proficiency.

This negative association can be explained by the fact that words with fewer orthographic neighbors are generally seen as an indication of more advanced vocabulary. Thus, texts with fewer orthographic neighbors are generally viewed as more linguistically advanced than those texts with a higher number of orthographic neighbors. The relationship between orthographic neighborhood frequency and perceptions of linguistic ability is supported by results from Kim et al. (2018) who also found a negative relationship between the number of orthographic neighbors and perceptions of L2 English proficiency on independent writing tasks. Therefore, as with previous findings in this study, these results allow us to extend the results of previous research targeting independent writing tasks to academic, integrated writing produced under testing conditions.

Lexical decision time – Content Words (Standard Deviation)

Similar to word naming response time, lexical decision time suggests the importance of moving beyond a reliance on the most easily recognized, and often earliest learned vocabulary, in order to incorporate higher level lexis that may be less quickly recognized by native speakers – thereby demonstrating greater vocabulary knowledge (e.g., Appel et al., 2019). With research suggesting that orthographic neighborhood frequency can help account for variance in word naming and lexical decision tasks (e.g., Adelman & Brown, 2007), it seems likely that the combination of *word naming response time*, *orthographic neighborhood frequency*, and *lexical decision time* all point to the same underlying characteristic: increased use of advanced vocabulary that has generally been acquired at a more advanced stage of linguistic development.

Implications

Methodological Implications

Findings suggest that it is necessary to continue expanding the range of variables used in studies aiming to identify underlying features associated with differences in L2 English academic writing ability, particularly in relation to lexical sophistication in integrated tasks. Although a comparatively large range of variables (484) were targeted here, a rather limited number remained after accounting for multicollinearity and weak correlations with the target construct.

Thus, it would seem as though, despite efforts to increase the range of metrics that can be applied in research of this kind, substantial overlap is limiting the usefulness of this widened scope. As a result, it may be necessary to more critically assess the range of metrics being made available to researchers through automated tools in order to better understand how distinct they truly are and what benefit is gained through each new offering.

A critical evaluation of these metrics may also lead to the development of new measures that help highlight alternative aspects of lexical sophistication that are more distinct from current offerings, thereby increasing explanatory power. In close relation to this, although we deliberately focused on one specific construct of writing quality in this study, lexical sophistication, additional constructs should continue to be targeted, with each metric evaluated under the same critical gaze. As 36% of total variation in essay scores was predicted by the metrics used here, further variance remains to be accounted for and additional variables, both established metrics related to other writing constructs (e.g., syntax, cohesion) and new lexical sophistication measures, would prove beneficial in the ongoing goal of accurately modeling human assessments of L2 writing quality on integrated writing tasks.

In the same vein, a vast amount of currently used statistical models (e.g., multiple regression) in this line of research, assume a certain level of independence among variables and breaching these assumptions can lead to the spurious interpretation of the results. Thus, careful attention must be paid to the manipulation of variables, data screening, relevance of the model, and the assumptions underlying statistical models in order to develop and apply metrics that will be of actual value in research of this kind. In relation to this point, it is not only the volume of metrics that must be increased, but also the quality and uniqueness.

With the limited number of multi-word measures currently available (at least in comparison to single-word measures), this category appears to offer a fruitful area for expansion. In fact, with regard to multi-word measures, Paquot (2019) has recently demonstrated that new phraseological metrics can indeed be used to distinguish between levels of the Common European Framework of References for Languages (CEFR). Thus, although multi-word metrics did not factor into the final regression in the present study, this category of variables still holds potential and should not be abandoned.

Pedagogical Implications

The most important finding from a pedagogical perspective is the high amount of variance in writing proficiency that could be explained by total word count. This finding, though not entirely surprising given results from previous studies, is important in that it suggests it is not only text quality that contributes to assessments of L2 writing ability, but also quantity. In light of the importance of total word count in assessed proficiency on this task, and other similarly timed tasks from previous research, it may be beneficial for teachers to incorporate more fluency-based activities to boost speed of written production. For instance, timed activities in which students are encouraged to write as much as possible on a given topic within a short time frame (e.g., 10-15 minutes) could be used as a way of promoting greater writing fluency and prepare students for similar testing conditions.

Although grammatical accuracy and appropriate academic vocabulary should not be ignored, it may be helpful to encourage students to focus more on quantity, as opposed to a strict adherence to accuracy, at least in the early stage of the writing process, in order to increase comfort producing longer texts that offer more linguistic data. This may also help students make use of a wider range of sophisticated lexis, as a fear of making mistakes may be reduced due to the shift in focus to text quantity. As students should already be encouraged to take a process view of writing that involves revision of each draft, an early focus on 'getting their ideas out' in the initial stages of the writing process could benefit later drafts as students will have a greater volume of text to revise (Elbow, 1973). Thus, initial drafts should be viewed as a first step that allows students to express their ideas fully without an overly oppressive fear of making mistakes – a factor that may limit production in favor of an emphasis on correctness.

While a focus on writing fluency by way of timed activities is likely to offer some benefit to learners, this recommendation is hedged by the fact that the results from this study, as well as those of many others in this area, are based on exam style writing produced under testing conditions. In this type of task, time constraints are likely a major factor that feeds into the subsequent assessment. In other words, the existence of time constraints may encourage raters to place greater emphasis on text length in the evaluation of each sample since lower-level writers may struggle to spontaneously produce large amounts of text without sufficient planning/drafting time.

However, these time pressures may be less important in assignments included as part of normal course activities (e.g., take-home assignments). This is particularly true in those assignments that make use of minimum/maximum word counts, since all students are expected to produce a similar volume of text. Therefore, while we believe a focus on improving students' writing fluency is a worthwhile goal that could lead to proficiency gains, it is important to keep in mind that improvements in writing fluency may not equally benefit all types of writing tasks, and fluency is only one aspect of a multi-faceted issue. Thus, time restricted activities may lead to proficiency gains on similar writing tasks; however, these activities may be less effective in achieving comparable gains on compositions that make use of more liberal time constraints (e.g., take-home assignments).

Despite the importance of writing fluency in assessments of this kind, the combination of word naming response time, orthographic neighborhood frequency, and lexical decision time (combined 10% of variance accounted for) reemphasize the importance of not simply producing a large volume of writing but incorporating genre and register appropriate lexis that demonstrates mastery of advanced vocabulary. Thus, although a major pedagogical implication of this study is that it is necessary to incorporate speed of production activities into the language classroom in order to prepare students to be able to write effectively under the time constraints common in testing conditions, appropriate register and genre specific language cannot be ignored. As a result, it would seem important for language instructors to gear their teaching approach to the goals of each specific group of students. For example, for test preparation, fluency activities may prove exceptionally beneficial. However, for students aiming to prepare for other settings (e.g., post-secondary studies), it may be more helpful to focus on activities that could lead to an increase in the level of lexical sophistication present in each student's text.

CONCLUSION

Two main conclusions can be drawn from the current study. First, although substantial advances have been made in the automated analysis of written discourse, further developments are needed to more deeply explore the existing relationship between lexical measures and human scores of L2 proficiency. As previous research also found a relatively low amount of variance accounted for by lexical measures in a similarly designed source-based writing task, there is a need for the development of further metrics to explore this issue. The methodological caveats we encountered can be a limiting factor to the development of automated analyses of written discourse in corpus research and this calls for the development of new measures or techniques that account for such limitations. Second, in light of the identified importance of text length in holistic assessments in the type of integrated writing task targeted here, it may be necessary for writing teachers to offer further opportunities for their L2 English learners to focus on writing fluency if they are to better prepare their students for the type of writing they will be asked to perform in assessments of this kind.

Automated text analysis in applied linguistics is blooming, but more work is needed to further understand the intricacies of linguistic measures in writing ability. New measures need to be developed to overcome the current issues associated with similar measures in different transformations (e.g., logs, standardized, raw), which is by and large the cause of multicollinearity in linear models. Thus, more advanced statistical models need to be explored to overcome this issue. Perhaps machine learning, and multilevel models are a good starting point.

FUNDING

This research was funded by a Waseda University Grant for Special Research Projects (Project # 2017S-180).

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Randy Appel: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing-review and editing.

Angel Arias: Statistical analyses.

REFERENCES

Adelman, J. S., & Brown, G. D. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14*, 455–459. https://doi.org/10.3758/BF03194088

Appel, R., Trofimovich, P., Saito, K., Isaacs, T., & Webb, S. (2019). Lexical aspects of comprehensibility and nativeness from the perspective of native-speaking English raters. *ITL-International Journal of Applied Linguistics*, *170*, 24-52. https://doi.org/10.1075/itl.17026.app

Appel, R., & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: Differences between high-and low-proficiency levels. *Language Assessment Quarterly*, *13*, 55-71. https://doi.org/10.1080/15434303.2015.1126718

Balota, D. A, Yap, M. J., Cortese, M. J., Hutchison, K. A, Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459. https://doi.org/10.3758/BF03193014

Berger, C. M., Crossley, S., & Kyle, K. (2017). Using native-speaker psycholinguistic norms to predict lexical proficiency and development in second language production. *Applied Linguistics, 40*(1), 22-42. https://doi.org/10.1093/applin/amx005

Bi, P., & Jiang, J. (2020). Syntactic complexity in assessing young adolescent EFL learners' writings: Syntactic elaboration and diversity. *System*, *91*(4), 1-10. https://doi.org/10.1016/j.system.2020.102248

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904–911. https://doi.org/10.3758/s13428-013-0403-5

Casal, J.E., & Lee, J.J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing*, *44*, 51-62. https://doi.org/10.1016/j.jslw.2019.03.005

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology Section A, 33*, 497–505. https://doi.org/10.1080/14640748108400805

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-238. https://doi.org/10.2307/3587951

Crossley, S. (2020). Lingusitic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415-443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S., Cai, Z., & McNamara, D. (2012). Syntagmatic, paradigmatic, andautomatic n-gram approaches to assessing essay quality. In P. M. McCarthy & G.M. Youngblood (Eds.), *Proceedings of the 25th International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 214–219). AAAI Press.

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26*, 66–79. https://doi.org/10.1016/j.jslw.2014.09.006

Crossley, S., Kyle, K., McNamara, D. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods, 48*, 1227-1237. https://doi.org/10.3758/s13428-015-0651-7

Crossley, S. A., Roscoe, R. D., McNamara, D. S., & Graesser, A. (2011) Predicting human scores of essay quality using computational indices of linguistic and textual features. In G. Biswas, S. Bull, J. Kay, and A. Mitrovic (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (pp. 438-440). Springer. https://doi.org/10.1007/978-3-642-21869-9_62

Dascalu, M., McNamara, D. S., Crossley, S. A., & Trausan–Matu, S. (2016). Age of exposure: A model of word learning. In *The 30th Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence* (pp. 2928–2934). AAAI Press.

Elbow, P. (1973). *Writing without teachers*. Oxford University Press.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. The MIT Press.

Fitzgerald, J., & Spiegel, D. L. (1986). Textual cohesion and coherence in children's writing.*Research in the Teaching of English*, *20*, 263-80.

Fleckenstein, J., Meyer, J., Jansen, T., Keller, S., & Köller, O. (2020). Is a long essay always a good essay? The effect of text length on writing assessment. *Frontiers in Psychology*, *11*, 562462. https://doi.org/10.3389/fpsyg.2020.562462

Guo, L., Crossley, S., & McNamara, D. (2013). Predicting human judgements of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing, 18*, 213-238. https://doi.org/10.1016/j.asw.2013.05.002

Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Houghton Mifflin.

Howell, D. C. (2013). *Statistical methods for psychology* (8th ed.). Wadsworth.

Johns, A. M. (1993). Reading and writing tasks in English for academic purposes classes: Products, processes and resources. In J. G. Carson & I. Leki (Eds.), *Reading in the composition classroom: Second language perspectives* (pp. 274–289). Heinle & Heinle.

Kamimura, T., & Oi, K. (2001). The effects of differences in point of view on the story production of Japanese EFL students. *Foreign Language Annals, 34(2)*, 118 - 128. https://doi.org/10.1111/j.1944-9720.2001.tb02817.x

Keith, T. Z. (2019). *Multiple regression and beyond*: *An introduction to multiple regression and structural equation modeling* (3rd ed.). Routledge.

Kim, M., Crossley, S. (2018). Modelling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing, 37*, 39-56. https://doi.org/10.1016/j.asw.2018.03.002

Kim, M., Crossley, S., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal, 102*, 120-141. https://doi.org/10.1111/modl.12447

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitkin, R. W. Bailey, & N. Hamilton–Smith (Eds.), *The computer and literary studies* (pp. 153–165). Edinburgh University Press.

Kyle, K., & Crossley, S. (2015) – Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*, 757-786. https://doi.org/10.1002/tesq.194

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and sources-based writing. *Journal of Second Language Writing, 34*, 12-24. https://doi.org/10.1016/j.jslw.2016.10.003

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: a usage-based approach. *Language Testing, 34(4)*, 513-535. https://doi.org/10.1177/0265532217712554

Kyle, K., Crossley, S., & Berger, C. (2017). The tool for the automatic analysis of lexical sophistication. *Behavior Research Methods, 50*, 1040-1046. https://doi.org/10.3758/s13428-017-0924-4

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English. English*. Brown University Press.

Kuperman, V., Stadthagen–Gonzales, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods, 44*, 978–990. https://doi.org/10.3758/s13428-012-0210-4

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Lawrence Erlbaum.

Lei, L., Wen, J., & Yang, X. (2023). A large-scale longitudinal study of syntactic complexity development in EFL writing: A mixed effects model approach. *Journal of Second Language Writing*, *59*, 1-14. https://doi.org/10.1016/j.jslw.2022.100962

Li, Y., Lin, S., Liu, Y., & Lu, Z. (2023). The predictive powers of fine-grained syntactic complexity indices for letter writing proficiency and their relationship to pragmatic appropriateness. *Assessing Writing*, *56*, 1-15. https://doi.org/10.1016/j.asw.2023.100707

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly 45(1)*, 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Journal, 96*, 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44*, 295–322. https://doi.org/10.1177/00238309010440030101

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, *23*, 35–59. https://doi.org/10.1016/j.asw.2014.09.002

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research, 35*, 121-145. https://doi.org/10.1177/0267658317694221

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *The Modern Language Journal, 102*, 713-731. https://doi.org/10.1111/modl.12509

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912. https://doi.org/10.1111/lang.12079

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: an exploratory study. *Journal of Second Language Writing, 9(3)*, 259-291. https://doi.org/10.1016/S1060-3743(00)00028-X

Simpson–Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*, 487–512. https://doi.org/10.1093/applin/amp058

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. Bureau of Publications, Teachers College.

Wen, Q., Wang, L., & Liang, M. (2005). *Spoken and written English corpus of Chinese learners*. Foreign Language Teaching and Research Press.

Wolfe-Quintero, K., Inagaki, S., & Kim, H-Y (1998). *Second language development in writing: measures of fluency, accuracy, & complexity*. University of Hawaii Press.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

APPENDIX A

Writing Performance Band Score Criteria

Score	Criteria
10-20	***Very Limited:***
	Generally unable to express ideas effectively
	Very restricted and/or ungrammatical language
	Uses words randomly and without overall coherence
30	***Limited:***
	Attempts to write something which is related to the topic but the writing is not predictable
	Restricted and/or ungrammatical language
	Seems to understand the topic, but is unable to develop ideas because language constrains or distorts expression
40	***Marginally Competent:***
	Makes links among ideas and addresses the topic but the writing lacks clarity and cohesiveness
	Displays elements of control in the writing (e.g. a thesis statement, an introduction and conclusion) but internal coherence is lacking
	Uses little or no support (i.e., quotations, examples, etc.) to develop the thesis
50	***Competent but Limited:***
	Addresses the topic to a degree but with somewhat limited clarity and cohesiveness
	Uses some support to develop the thesis
	Control of the argument is limited by poor comprehension of the readings and lecture, and/or poor understanding of the requirements of academic writing
60	***Competent:***
	Develops a thesis using a range of support
	Uses language that is generally accurate but is constrained by a somewhat limited vocabulary
	Demonstrates general understand of the requirements of academic writing
70	***Adept:***
	Responds readily to the demands of the topic and presents information clearly and logically
	Uses the readings and lecture effectively to support the thesis
	Demonstrated understand of the requirements of academic writing
80-90	***Expert:***
	Demonstrates mastery of appropriate, concise, and persuasive academic writing
	Writes with authority and style

Note. The scale presented above was used to evaluate all essays included in this study (i.e., a retired version of the CAEL Assessment). However, this rating scale is no longer in use and has been replaced with an alternative version.

https://doi.org/10.17323/jle.2023.12663

The Interplay between Corrective Feedback, Motivation and EFL Achievement in Middle and High School Education

Emnijeta Ahmetovic, Senad Becirovic, Vildana Dubravac, Amna Brdarevic-Celjo

International Burch University, Sarajevo, Bosnia and Herzegovina

ABSTRACT

Introduction. Despite the fact that error correction has significant and long-term effects on facilitating language learning and development, there has not been any research that investigates its influence on learners' motivation within the classroom context of Bosnia and Herzegovina.

Purpose. This research aims to examine the impact of written and oral corrective feedback on students' motivation and achievement within this EFL context.

Method. For this quantitative study, the questionnaire has been used to collect the data from 160 middle and high school students in central Bosnia and Herzegovina.

Results. The findings indicated that the respondents generally like to be corrected and they are moderately to highly motivated to speak and write in English as a foreign language. Furthermore, learners with positive attitudes towards the received feedback feel significantly more motivated to keep learning than those with negative attitudes.

Conclusion. The study is expected to provide teachers with suggestions on how to transform their classrooms into an environment conducive to the development of higher levels of writing and speaking motivation and how to provide corrective feedback that will positively influence students' EFL achievement.

KEYWORDS

corrective feedback, motivation, students' achievement, attitudes

INTRODUCTION

Learning motivation is learners' desire or willingness to put time and effort into initiating or sustaining the process of language learning (Ortega, 2009) and can eventually determine their success or failure in learning a foreign language. Thus, motivated individuals will learn a foreign language at a faster pace and to a larger degree than individuals lacking motivation (Gass & Selinker, 2001, p. 349). As such, motivation has attracted a keen interest of researchers (e.g. Deci & Ryan, 1991; Dörnyei, 2000; Gardner & MacIntyre, 1993), who have tried to develop a suitable model explaining the role of motivation in the process of second language learning. One of these models is the Self-Determination Theory which takes human interaction with the social environment as a way of satisfying the needs of autonomy, competence and relatedness as its starting point (Lagaut, 2017). Thus, if learners perform an activity with an internal drive for knowledge, pleasure and feeling of satisfaction, they are considered to be intrinsically motivated. Conversely, if the reason for doing the task is controlled from the outside world by the means of rewards and punishments, that is referred to as extrinsic motivation.

However, as indicated by previous researchers, the construct of motivation is a dynamic variable (Dervić & Bećirović, 2019; Dörnyei, 2005; Mašić et al., 2020), and is influenced by other individual and environmental factors, one of which is corrective feedback (CF) (Vásquez & Harvey, 2010) or a teacher's response

Citation: Ahmetovic E., Bećirovic S., Dubravac V., & Brdarevic-Celjo A. (2023). The Interplay between Corrective Feedback, Motivation and EFL Achievement in Middle and High School Education. *Journal of Language and Education, 9*(1), 26-40. https://doi.org/10.17323/jle.2023.12663

Correspondence:
Emnijeta Ahmetovic,
emnijeta.ahmetovic@stu.ibu.edu.ba

Received: June 29, 2021
Accepted: Mach 15, 2023
Published: March 31, 2023

to learners' incorrect utterances of some sort. Since CF occurs in a response to learners' errors, it is believed to be beneficial to students because it gives them negative evidence and helps them to see the difference between the target language they have acquired and the target-like L2 (Second Language) form (Long, 1996). Russell and Spada (2006) maintained that CF in L2 classroom instruction is an apt tool for improving language production despite the existing differences in teachers' instruction modes. Still, in spite of such positive perceptions of CF as a practice prompting successful L2 acquisition (e.g., Doughty & Long, 2003; Doughty & Williams, 1998; Gass, 2003; Long, 1996; Schachter, 1991), it is very frequently observed as a needless practice with little impact on L2 learning (e.g., Flynn, 1996). Nevertheless, in teaching schemes, feedback plays an essential role in maintaining the teacher-student or peer-student interaction (Polz & Bećirović, 2022; Hyland & Hyland, 2006) and also in enhancing students' autonomous learning (Hyland, 2003) and its efficacy relies on the nature of linguistic features, teaching practices, and learners' individual characteristics, motivation in particular (Schachter, 1991).

CF has also been frequently pointed out as a factor which can negatively influence motivation (DeKeyser, 1993; Jean & Simard, 2011), but also as a process appreciated by language learners and observed as a factor strengthening motivation (Jean & Simard, 2011; Lee, 2013; Yoshida, 2008). In the majority of recent studies feedback is seen as beneficial (Evans et al., 2010; Ferris, 2006; Kang & Han, 2015; Nassaji, 2016). However, there is still some controversy about what form of correction is more influential (Al-Rubai'ey & Nassaji, 2013; Bitchener, 2008; Chen et al., 2016; Nassaji, 2015), which has proved to be at least partially dependent on learners' individual differences (Zamel, 1985), and their interest in the learning process (Papi et al., 2019).

Hence, teachers should have a deep insight into theoretical as well as practical guidelines related to this issue (Schulz, 2001), and they are expected to take learners' beliefs into account as well (Jean & Simard, 2011; Yaman & Bećirović, 2016), since the difference between their and their students' beliefs could have detrimental effects on motivation (Kalaja & Barcelos, 2003). Therefore, to foster learners' motivation and ensure linguistic accuracy, teachers need to be familiar with their students' perceptions of the best way of correcting mistakes. In order to be extrinsically motivated, students need encouragement and support which can be provided by teachers through feedback as a reward. However, if learners are aware of their own learning progress, it can lead to greater intrinsic motivation and students' engagement in learning. Thus, as found by previous researchers (e.g., Choi, 2013; Gardner, 2011), the use of CF might decrease the risk of lowering students' motivation.

The studies determining how and to what extent CF and motivation interact with each other to affect language learning outcomes in a foreign language learning context seem to be relatively scant. Moreover, it has also been pointed out that such studies have mainly focused on adults and that there is a need for research revealing the impact of CF on younger learners (Jean & Simard, 2011). Therefore, the current study aims to investigate the relationship between CF and motivation taking into account their interactive effect on middle and high school students' English language performance. Moreover, it aims to help teachers, students, and future researchers understand whether the way of providing feedback in middle and high school motivates students and to what degree. It also reveals students' attitudes towards the received feedback as well as to its frequency, and the impact of error correction on students' achievement. Such a comprehensive approach can further our understanding of why CF is so important and can open new ways of research on how to make feedback more effective in order to motivate students.

LITERATURE REVIEW

Corrective Feedback

Although research into the role of CF and its impact on L2 acquisition is steadily progressing, there is no single study that covers all aspects of understanding its controversial function. In general, CF has been found beneficial to SLA (Second Language Acquisition) but in order to gain a better understanding of its controversial role as well as to identify research orientations for L2 scholars, it is important to present a comprehensive overview of its effectiveness.

The usefulness of both written and oral correction for learners' L2 improvement has been frequently confirmed (Ellis, 2009; Li & Vuono, 2019) and negative evidence has emerged as a necessary factor for language learning advancement (Gooch et al., 2016; Li, 2010; Lyster & Saito, 2010; Lyster et al., 2013; Nassaji, 2016, 2017; Yang & Lyster, 2010). Several recent meta-analyses (e.g., Kang & Han, 2015; Karim & Nassaji, 2019; Lyster & Saito, 2010; Plonsky & Brown, 2015; Russell & Spada, 2006) have also shown that CF significantly impacts L2 learners' performance, with long-lasting effects (Russel & Spada, 2006). Correction frequently appears in numerous classroom events (Brown, 2016; Wang & Li, 2020) and it has been particularly effective in constituting students' achievement (Bitchener et al., 2005; Choi, 2013; Li & Vuono, 2019; Sheen, 2007). Patra et al. (2022) showed that immediate feedback positively affects learners' academic performance, since it triggers positive attitudes towards continuous learning and learners get more involved in learning.

Some researchers find oral correction feedback unneeded and useless. According to Krashen (1982), error treatment is maleficent to the whole learning process, as it makes learners avoid complex tasks, and "even under the best

of conditions, with the most learning-oriented students, teacher corrections will not produce results that will live up to the expectations of many instructors" (Krashen, 1982, p. 119). Truscott (1999) believes that correction provided during spoken activities should be avoided as it does not enhance learners' willingness to speak accurately and mentions some of its disadvantages, such as the difficulty of giving clear explanation for an error and the uncertainty as to whether feedback will help learners in acquiring the target language. Lyster et al. (1999) disagree with Truscott's findings, claiming that effectiveness of oral CF is certain and correction is necessary in some cases. Through correction, students engage in their L2 "communicatively-authentic discourse" (Lyster et al., 2013), which in return improves their accuracy, fluency as well as automaticity (Saito, 2021, p. 408). Both students and teachers in Ha et al. (2021) study conducted in the Vietnamese EFL context felt very positive about explicit feedback types, namely explicit correction and metalinguistic feedback, but they differed in their beliefs on feedback timing. While the students favor immediate feedback, the teachers worry about their emotional state and are concerned that feedback may disrupt the flow of their speech (Ha et al., 2021).

In addition to studies investigating spoken error correction, there have been a few studies investigating whether CF of written errors can help students to improve their writing accuracy. In some earlier studies, corrective feedback of written errors did not prove particularly useful as the students who received correction and those who did not differed insignificantly in writing accuracy (e.g. Robb et al., 1986; Sheppard, 1992). In some more recent studies, feedback has proved to be beneficial to students' writing accuracy (Bitchener, 2017; Bitchener & Storch, 2016; Bruton, 2010; Chandler, 2009; Chen & Nassaji, 2018; Ferris 2006; Kang & Han, 2015; Karim & Nassaji, 2019), though the degree to which it can improve L2 writing is still debatable (Karim & Nassaji, 2019). Karim and Nassaji (2019) also emphasized that students need to provide output with the correct form on their own as such practices could be more useful for their interlanguage development. Several studies have investigated the implementation of written corrective feedback (WCF) over a longer time period (i.e., weeks or months) and have come to different conclusions. Some studies provide support for WCF (see. Ashwell, 2000; Ferris & Roberts, 2001), while some others fail to see any advantages on the side of the group that received WCF (Polio et al., 1998). This might be assigned to the non-existence of a control group in some previous studies (Bitchener, 2008; Storch, 2010; as cited in Karim & Nassaji, 2019) or to the measurement of the usefulness of correction only in terms of the production of revised utterances (e.g., Ferris, 2010). More recent studies have investigated the effect of correction not only on revised but also on new pieces of writing (Hartshorn et al., 2010; Karim & Nassaji, 2018; Lopez et al., 2018) and their findings varied depending upon the ways of providing feedback. Xu's (2021) findings showed that Chinese learners were generally well-disposed towards WCF in an online environment in this case, but what they particularly liked is the clarity of teachers' feedback and its potential for being analyzed further or as marked in their responses 'indefinitely' (Xu, 2021, p. 568).

When grading students' written compositions, teachers are obliged to provide correction but are also advised to provide feedback "an objective that may operate at cross-purposes with the evaluative goal" (Hedgcock & Lefkowitz, 1996, p. 288). Teachers sometimes tend to be impacted by the belief that the more errors they correct the more responsible teachers they are (Lee, 2013, p. 113), which is potentially dangerous and can impact the learning process (Hyland & Hyland, 2006). It is generally suggested that not all errors should be corrected but correction should be directed towards a few errors so that learners pay more attention to the feedback thus provided (Ellis *et al*., 2008).

Corrective Feedback and Motivation

Although it has been confirmed that both feedback and motivation are important for language learning, empirically very little is known about their interplay in SLA (Ellis, 2010). More specifically, some recent studies conducted by the leading CF investigators (Lyster et al., 2013; Nassaji, 2016) do not pay so much attention to the impact of motivation on the usefulness of correction (Bitchener, 2017) and neither do L2 motivation investigators focus so much on this relation (see Dörnyei & Ushioda, 2011).

However, numerous experimental and non-experimental research studies have found CF and motivation to be positively correlated (e.g., DeKeyser 1993; Guilloteaux & Dörnyei, 2008). Dörnyei and Csizer (1998) interviewed 200 Hungarian language teachers asking them to rank a selection of 51 teaching strategies considered as motivational factors and the results showed that one of the strategies ranked high was teacher feedback which is seen as an influential factor in students' motivation and as a stimulator of learners' motivation and self-confidence. Hence, consistent feedback is considered to be essential in facilitating L2 motivation since it carries a clear message about the teacher's priorities and is reflected in the student motivation (Dörnyei, 1994, p. 278). A similar study conducted by Guilloteaux & Dörnyei (2008) among EFL students in South Korea also indicated that CF in the form of prompts and students' motivation for language learning are strongly and positively correlated.

The correlation between CF and motivation was also found in a quasi-experimental study carried out by DeKeyser (1993), which examined the influence of feedback on L2 grammar and oral performance. In this study, the participants were learners of French who went through a ten-month treatment in Belgium, with one group obtaining constant correction and the other group not obtaining any. DeKeyser (1993) found that students with lower motivation achieved better results

on oral fluency tests as well as accuracy tests after constantly receiving correction, which suggests that correction might directly affect motivation and eventually have positive outcomes for the learning process. Thus, DeKeyser's original hypothesis stating that students with low motivation would take CF as criticism and that it would not be beneficial for them as much as for students with high motivation was refuted since the results confirmed that correction had an impact on students' L2 motivation regardless of its initial level. Likewise, Hamidun et al. (2012) examined if teacher's feedback can foster students' motivation. As it was initially found through classroom observation that students were not motivated or had little motivation to engage in language activities, explicit correction was used, and the results indicated that direct feedback increased the level of motivation, with participants responding well to this type of error correction. Al-Darei & Ahmed (2022) also researched the effect of feedback but also of feedback type on students' motivation and achievement in the e-learning environment. Their findings also indicated an increase in motivation and achievement after the implementation of feedback, particularly interpretative feedback.

The Present Study

The present study aims to explore whether written and oral corrective feedback lead to an increase in students' motivation for learning English as a foreign language and better EFL achievement in the middle- and high-school milieu of Bosnia and Herzegovina. Based on the theoretical and empirical findings presented above, the research will respond to the following research questions:

(1) Does oral CF significantly correlate with motivation for EFL speaking in the classroom?

(2) Does written CF significantly correlate with motivation for EFL writing in the classroom?

(3) Does gender significantly influence motivation for EFL speaking and writing when the English study duration is controlled?

(4) Does oral and written CF in English as a foreign language significantly predict the students' EFL achievement?

METHOD

Participants

The research sample consisted of 160 middle and high school students. Different grade levels were included with randomly selected participants, and a stratified sampling method was employed. Two classes (the first and the third grade) from high school and five classes (from the fifth to the ninth grade) from middle school, both located in central Bosnia, participated in this research. The sample consisted of 90 female participants (56.3%) and 70 male participants (43.8%) aged between 10 to 17 (M = 13.83, SD = 1.998). Table 1 gives an accurate description of the representative selection.

Table 1

Descriptive Analysis of the Research Sample

		N	%
School	Middle school	118	73.8
	High school	42	26.3
Gender	Female participants	90	56.3
	Male participants	70	43.8
Grade levels	Fifth	16	10.0
	Sixth	16	10.0
	Seventh	29	18.1
	Eighth	27	16.9
	Ninth	30	18.8
	First	21	13.1
	Third	21	13.1
Total		***160***	***100.0***

Instruments and Procedures

After the schools' administration as well as the participants themselves granted the approval, the data collection tools were prepared, administrated and adjusted accordingly to the middle and high school by the researchers themselves. The participants were informed that the data obtained from these surveys will be anonymous, confidential and voluntary and were provided with an appropriate clarification on how to fill in a Likert-type scale.

The questionnaire comprised three core parts. The first part collected statistical data about the participants' individual characteristics, such as gender, age, GPA in English, grade level and the duration of English language learning process. The second part consisted of the motivation scale to speak and write in English as a foreign language designed and authorized by Ryan & Connell (1989). The aim of this survey was to gain more in-depth information about how participants are motivated to speak and write in English as well as its correlation with other variables such as gender, achievement and the duration of learning English. The instrument comprised 33 items for speaking and 33 items for writing motivation, separated into three

crucial categories, namely intrinsic, extrinsic motivation and amotivation, which are further divided into seven subscales (amotivation e.g. Doing these activities is not interesting for me; external regulation e.g. Because I know I will get in trouble if I do not; introjected regulation e.g. Because it is absolutely necessary to do these activities if one wants to be successful in language learning; identified regulation e.g. Because it is a good way to gain lots of skills that could be useful to me in other areas of language learning and my life; intrinsic motivation for knowledge e.g. Because I experience a great pleasure while discovering new techniques of expression of ideas and feelings through these activities; intrinsic motivation for accomplishment e.g. Because I think carrying out hard tasks in these activities will improve my performance; and intrinsic motivation for stimulation e.g. Because it makes me happy). The Relative Autonomy Index (RAI) was used to measure both self-determined motivation for speaking and writing by using the original RAI formula: *RAI = (2 * intrinsic + identified) - (2 * external + introjected*) (Ünlü, 2016, 2019). The instrument showed the overall consistency reliability, motivation to write, in total (introjected regulation, identified regulation, intrinsic motivation for knowledge, intrinsic motivation for accomplishment and intrinsic motivation for stimulation dimension) α = .81 as well as motivation to speak, in total α = .79. Also, the compounded writing intrinsic motivation α = .75 and speaking intrinsic motivation α = .72 were consistent.

The third and the last part of the survey was the Corrective Feedback Scale, adopted from Calderón (2013). It employed a five-point Likert scale, ranging from strongly disagree to strongly agree and it contained 15 statements related to written feedback and 15 statements related to spoken feedback, divided into the following four categories: a type of feedback (e.g. I like it when my teacher explicitly tells me I made a mistake and gives me the right version of it), the frequency of given feedback (e.g. Teachers should correct students every time they make a mistake when speaking/ writing English), positive attitude towards CF (e.g. Error correction is good for language learning) and negative attitude towards CF (e.g. I usually feel embarrassed when my teacher corrects me in front of the whole class). Each of the mentioned constructs of the survey has three to four questions addressing distinct focus (Calderón, 2013). This part of the survey was used to investigate the students' perception of CF and its association with learners' motivation to speak and write in English as a FL, as well as their EFL achievement.

To fill in the items provided in the questionnaire, the participants were kindly asked to read all the statements carefully and without exception and to choose the preferred number (1-5) which demonstrates their beliefs about the given items. To complete the survey, the participants needed approximately 25 minutes.

Data Analysis

The Statistical Package for Social Science (SPSS) version 23.0 was computed for examining the data, and four different statistical procedures were applied. Before testing the hypothesis, the normality test was performed (Byrne, 2010; Chua, 2013; Hair et al., 2010). To determine the extent of the students' motivation to speak and write in English as well as their attitude towards CF in the English language classroom, means and standard deviation were calculated. Then, the Person correlation coefficients were calculated to explore the relationship between the constructs of CF and motivation. One-way MANCOVA was performed in order to determine the impact of gender on writing and speaking motivation with the impact of English study duration being controlled. Finally, standard multiple regression was employed to examine students' motivation and the impact of corrective feedback on students' English language achievement.

RESULTS

Initial Analyses

The initial analysis measured the mean values and standard deviation on all subscales of motivation, i.e. amotivation, external regulation, introjected regulation, identified regulation, intrinsic motivation for knowledge, intrinsic motivation for accomplishment, and intrinsic motivation for stimulation. It also measured the mean values and standard deviation of all variables of corrective feedback, i.e. frequency of given feedback, positive attitude towards corrective feedback, and negative attitudes towards corrective feedback (Table 2).

In terms of speaking motivation subscales, the analysis revealed that students experienced identified regulation as the highest level of motivation with a rather high mean value, followed by intrinsic motivation for accomplishment, introjected regulation, intrinsic motivation for knowledge, external regulation and intrinsic motivation for experiencing stimulation. Rather expectedly, the lowest mean was measured for amotivation. With regards to corrective feedback, the findings related to three categories of the CF scales revealed that participants generally like to be corrected frequently and they were quite positive about it, with positive attitudes being rather high and negative attitudes moderate. Skewness and kurtosis values showed that all the variables are normally distributed since the range of normality of skewness and kurtosis is from -2 to +2 (Garson, 2012; Hair et al., 2010).

Moreover, the mean values and standard deviation of all of subscales of motivation for writing and written corrective feedback were also measured and the data are presented in Table 3.

Table 2

Descriptive Analysis of Motivation Subscales and Categories of CF Scale for Speaking

Measure	M	SD	Skewness	Kurtosis
Amotivation	2.58	.79	-.052	-.458
External regulation	3.34	.63	-.150	.343
Introjected regulation	3.67	.66	-1.001	2.045
Identified regulation	3.76	.64	-.545	1.011
Intrinsic motivation for knowledge	3.65	.86	-.400	-.492
Intrinsic motivation for accomplishment	3.72	.77	-.926	1.604
Intrinsic motivation for stimulation	3.30	.89	-.156	-.514
Frequency of given CF	3.15	.78	-.145	.167
Positive attitude towards CF	3.78	.64	-.333	.499
Negative attitude towards CF	2.93	.88	-.228	-.219

Table 3

Descriptive Analysis of Motivation Subscales and Categories of CF Scale for Writing

Measure	M	SD	Skewness	Kurtosis
Amotivation	2.60	.81	-.159	-.603
External regulation	3.19	.63	-.122	.185
Introjected regulation	3.55	.68	-.653	1.103
Identified regulation	3.74	.67	-.679	1.233
Intrinsic motivation for knowledge	3.65	.82	-.400	-.395
Intrinsic motivation for accomplishment	3.70	.81	-.924	.911
Intrinsic motivation for stimulation	3.40	.85	-.036	-.964
Frequency of given CF	3.27	.75	-.377	.829
Positive attitude towards CF	3.70	.67	-.310	-.429
Negative attitude towards CF	2.83	.92	.173	-.289

The findings related to motivation for writing indicated that the participants scored the highest mean on identified regulation, followed by intrinsic motivation for accomplishment, intrinsic motivation for knowledge, introjected regulation, intrinsic motivation for stimulation, and external regulation. The participants again obtained the lowest mean on amotivation. Considering the frequency of receiving feedback, the current study students generally like to be corrected when they make an error in their writing. Overall, they appear to have positive attitude towards error correction and when they are corrected in writing, they feel that they learn more. As for their negative attitude towards correction, the mean value was not rather high. The skewness and kurtosis values again show that all the variables are normally distributed (Hair et al., 2010).

The Relationship between Oral CF and EFL Motivation to Speak

The analysis further measured the relationship between the subscales of oral corrective feedback and motivation for speaking English as a foreign language (Table 4). The interrelationship analyses revealed that the frequency of CF (r = -.16 p = .048) as well as negative attitude (r = -.33 p = .000) towards correction are negatively and significantly correlated with EFL motivation for speaking, while positive attitudes towards CF and motivation for speaking are positively and significantly correlated (r = .19, p =.017). Therefore, those students who have rather negative attitudes towards correction and who are corrected more frequently seem to be less motivated for speaking English, while those students who have positive attitudes towards

CF are more motivated to speak in a foreign language. The correlation between negative attitudes towards CF and positive attitudes towards CF is negative and statistically significant (r = -.20, p = .01). In addition, it is found that positive attitudes towards CF (r = .00, p > .05) and negative attitudes towards CF (r = .15, p > .05) are positively but insignificantly related to the frequency of CF.

The Relationship between Written CF and EFL Motivation to Write

The relationship between the subscales of oral corrective feedback and motivation for writing was also measured (Table 4). The correlation analyses show that writing motivation is positively and significantly correlated only with positive attitudes towards error correction (r = .21 p = .008). It is negatively but statistically insignificantly correlated with the frequency of given feedback (r = -.12, p > .05) and negatively and statistically significantly correlated with negative attitudes towards feedback (r = -.22, p = .004). This indicates that the students who have higher negative attitudes towards CF are less motivated to write. The results also revealed a negative and insignificant correlation between positive attitudes towards CF and its frequency (r = -.11, p > .05), as well as negative and statistically significant correlation between positive and negative attitudes towards corrective feedback (r = -.23, p = .004). The correlation between the frequency of given feedback and negative attitudes towards error correction was positive and statistically significant (r = .17, p = .030).

The Relationship between Gender and EFL Motivation to Speak and Write

A Multivariate analysis of covariance (MANCOVA) was performed to examine the effect of gender on EFL motivation to speak and write (RAI variables), while the impact of the duration of studying English was controlled. The MANCOVA results revealed insignificant differences between female and male participants on the combined depended variable [Wilks' Λ = .996, F (2, 156) = .324, p = .724, multivariate η *2* = .004]. The results were also insignificant on the covariate duration of studying English [Wilks' Λ = .948, F (2, 156) = 2.435, p = .091 >.001, multivariate η *2* = .030]. Likewise, differences based on gender were insignificant on speaking motivation F (1, 909) = .286, p = .593, η *2* = .002, as well as on writing motivation F (1, 005) = .002, p = .966, η *2* = .000. The differences were also insignificant on the controlling variable for speaking motivation F (7, 619) = 2.400, p = .123, η *2* = .015 as well as for writing motivation F (4, 895) = .000, p = .997, η *2* = .000.

Corrective Feedback as a Language Achievement Predictor

With the intention of determining the accuracy of the frequency of CF, positive attitudes towards CF, and negative attitudes towards CF in speaking and writing in forecasting the learners' accomplishment visible in terms of the English average grade, the standard multiple regression was conducted. The regression analysis showed that error correction scale insignificantly predicted attainment in mastering English as a foreign language R^2 = .056, R^2 *adj.* = .019 F (6,153) = 1.768, p = .173. The regression coefficient displayed in Table 5 also demonstrates that higher levels of positive attitudes towards speaking and writing are associated with higher levels of students' EFL achievement despite its insignificance. Likewise, the frequency of CF and negative attitudes towards CF in terms of both speaking and writing do not significantly predict the students' EFL performance but are negatively related to the student's achievement, as seen in Table 5.

DISCUSSION

Descriptive Data

The current paper aimed to explore Bosnian EFL learners' speaking and writing motivation and their stance towards oral and written CF received during English classes. The results showed that these learners in general like to be corrected since the majority of them have positive attitudes towards given feedback, which implies that they are convinced that CF accelerates their learning. Such findings are closely aligned with those reported by Schulz (1996), who found that 90% of the participants hold rather positive attitudes towards correction and grammar teaching, as well as those presented by Hyland (1998) and Ferris and Roberts (2001), who claimed that students expect to be corrected

Table 4

Correlations for Speaking/Writing

	N	1	2	3	4
1. Speaking motivation (RAI)	160	1/1			
2. Frequency of given CF	160	-.16*/-.12	1/1		
3. Positive attitudes toward CF	160	.19*/.21**	.00/-.11	1/1	
4. Negative attitudes toward CF	160	-.33**/-.22**	.15/.17*	-.20*/-.23**	1/1

Table 5

Multiple Regression CF Speaking and Writing

	B	B	T	P	Bivariate r	Partial r
Frequency CF speaking	-.095	-.068	-.428	.669	-.116	-.035
Frequency CF writing	-.023	-.015	-.098	.922	-.109	-.008
Positive attitudes speaking	.082	.048	.350	.727	.092	.028
Positive attitudes writing	.012	.007	.052	.958	.101	.004
Negative attitudes speaking	-.159	-.128	-.647	.518	-.214	-.052
Negative attitudes writing	-.080	-.067	-.339	.735	-.211	-.027

and they perceive CF as beneficial while learning the target language.

The current study participants' answers linked to the frequency of given feedback were quite diverse. Overall, it was shown that the participants like to be frequently corrected because error correction helps them to learn more by transforming a wrong utterance they have produced. These results are in accordance with those reported in the study conducted by Kim and Mathes (2001), who argued that continued feedback is needed when compared to limited feedback, especially if an error occurs more frequently. What is rather interesting is that the participants' negative attitudes towards both oral and written error correction were measured as moderately low, which also corroborates some previous findings suggesting that error correction has a remarkable role in the process of teaching and learning a second or foreign language (see Ellis, 2009; Hattie & Timperley, 2007; Li & Vuono, 2019; Ramsden, 1992).

In terms of motivation to speak and write in a foreign language, the results revealed that it is somehow high though the participants of the present study seem to be more motivated to speak in English, which might be attributed to the fact that writing is frequently cited as the most challenging proficiency target (Laličić & Dubravac, 2021) as well as to the fact that the majority of language learners wish to develop their speaking skills at a rather fast rate, since "someone's fluency in speaking measures his/her proficiency in that language" (Martin, 2011, p. 237). These findings are consistent with previous research on motivation in the same context. For instance, Bećirović and Brdarević-Čeljo (2018) and Ahmetović et al. (2020) also indicated that Bosnian students are highly motivated to learn the English language. This can be ascribed to the fact that Bosnia and Herzegovina is a country where the English language is widespread (Brdarević-Čeljo & Dubravac, 2022; Brdarević-Čeljo et al., 2018; Brdarević-Čeljo et al., 2021) and presents an effective means of establishing good communication as well as close cooperation (Brdarević-Čeljo & Asotić, 2017; Bećirović, 2023; Dubravac & Latić, 2019; Dubravac & Skopljak, 2020, Ribo & Dubravac, 2021). Being influenced by the media as well as by economic and technological advancements, these Bosnian students feel more motivated to put more effort into learning English as a foreign language (Brdarević et al., 2021; Delić et al., 2018; Dubravac, Brdarević-Čeljo & Bećirović, 2018; Dubravac, Brdarević-Čeljo & Begagić, 2018; Kovačević et al., 2018).

The Relationship between Oral CF and EFL Motivation to Speak

The first research question sought to identify whether oral CF significantly correlates with motivation for speaking in English. The current study correlation results indicated that positive attitudes towards error correction are significantly correlated with motivation, which means that more positive attitudes towards receiving feedback help develop the participants' proficiency and those who complete the task successfully are expected to continue working hard to improve in the areas where they are less skillful and will be motivated to increase their effort. These results are in line with those in some other studies which also showed a positive and significant correlation between CF and motivation (e.g., DeKeyser, 1993; Guilloteaux & Dörnyei, 2008). On the other hand, negative attitudes towards correction correlated negatively and significantly with motivation, which shows that the current study participants who are not well motivated will adopt negative stances towards their tutors' correction and the other way round. The frequency of given feedback also significantly correlated with motivation, but the correlation coefficient was negative. Even though it is generally believed that correction enhances language proficiency, still, teachers should be aware of the fact that frequent error correction in spoken activities disturbs the flow of speech and is the cause of learners' demoralization while performing classroom activities (Lasagabaster & Sierra, 2005).

The Relationship between Written CF and EFL Motivation to Write

To respond to the second research question, the correlation between motivation for writing in English and

three subcategories of written corrective feedback, i.e. frequency of CF, negative and positive attitude towards error corrections, was investigated. In particular, positive attitude towards written CF significantly correlates with the participants' motivation to write, indicating that the current study participants who are well disposed towards written CF have stronger motivation to write. On the contrary, two other subcategories of CF correlated negatively with motivation to write, namely frequency towards given CF non-significantly, and negative attitudes towards CF significantly, which shows that the current study participants who are not well disposed towards CF do not feel motivated particularly if they are corrected frequently. These results are aligned with some other studies (e.g., Bitchener, 2017; Bitchener & Storch, 2016; Ferris, 2003; Leki, 1991), which showed that learners consider feedback valuable and if they do not receive feedback they want, they may lose motivation (Ferris, 2003). Additionally, learners who enjoy learning English, despite having negative attitudes towards correction and constantly being corrected, may find learning challenging and thus become more motivated. Bearing in mind that writing is a difficult (Graham & Harris, 2013; Ogrić-Kevrić & Dubravac, 2017) and an unattractive task (Boscolo & Gelati, 2013; Hamzić & Bećirović, 2021) involving multiple steps, teachers should nurture and support students by giving meaningful feedback on the areas that need improvement. Hence, the feedback provided by teachers may raise learners' motivation. Though the frequency of given feedback is positively correlated with learners' motivation, constant and extensive correction and taking notice of every mistake the learner makes is needless (Graham et al., 2011). Overall, as one of the major factors influencing students' motivation, feedback needs to be an essential part in acquiring writing skills (Schunk et al., 2014), but teachers should be cautious while employing it as in some research studies no advantages of WFC were observed (e.g. Polio et al., 1998).

The Relationship between Gender and EFL Motivation to Speak and Write

The third question researched the impact of gender on the participants' motivation for speaking and writing in English when the duration of studying English is controlled. Overall, the current findings did not find that gender had a significant impact on the participants' writing and speaking motivation as suggested by previous research (Bećirović et al., 2022; Dörnyei & Csizer, 2005; Harthy, 2017; Xiong, 2010). Still, the current findings are in accordance with the findings of some other studies (Akram & Ghani, 2013), in which no gender-based differences were measured in terms of motivation. Despite the non-existent gender differences in the current study, the male participants had an insignificantly higher level of speaking motivation, while the female participants were more motivated in terms of writing.

Corrective Feedback as Language Achievement Predictor

The fourth research question asked whether oral and written CF in English significantly predict the participants' language achievement. In the current study, neither oral nor written CF significantly predicted the participants' EFL achievement. Taking into consideration different components of oral as well as written corrective feedback (frequency of given feedback, as well as negative and positive attitude towards error correction), the current study showed that positive attitude is insignificantly and positively associated with the learners' attainment, which indicates that the current study participants who are well disposed towards error correction might be more successful in learning the target language since they believe that CF facilitates their progress in developing greater EFL proficiency. This finding is in accordance with the past research discoveries which demonstrated correction effectiveness in improving students' achievement (Bitchener et al., 2005; Choi, 2013; Hyland, 1998; Lee, 2004; Sheen, 2007). The frequency of given feedback and negative attitude towards error correction are insignificantly but negatively related to the participants' attainment, which suggests that these participants find continuous correction very annoying, distracting and ineffective and some of them would even stop participating in the classroom activities to avoid being frequently corrected. Thus, too much feedback can hinder students' achievement as they might feel controlled, which results in putting less effort into learning activities. Therefore, as it was confirmed previously by the great majority of studies, correction should take place in EFL classrooms, since it improves students' accuracy (see Kepner, 1991; Sheppard, 1992), but teachers should be cautious as to how frequently they employ it especially during speaking activities. To sum up, it seems that learners with positive attitude towards CF will benefit more from correction and will, thus, succeed in increasing their language proficiency than those with negative attitudes as correction can impact the scope of their engagement in learning activities (Sheen, 2007).

CONCLUSION

Error making is an unavoidable aspect of language learning, which should not be hindered by error correction. In general, students tend to believe that they should receive correction when they make a mistake, and that teachers are tasked with the responsibility of providing learners with correct information. Even though implementing error correction properly can be challenging, it is crucial that teachers are knowledgeable of learners' attitudes towards given feedback so that they can use correction effectively while addressing students' errors and not decrease their motivation.

Therefore, the aim of the current study was to investigate students' motivation to speak and write in English and its correlation with their attitude towards error correction and frequency of given feedback based on their gender, achievement, and duration of EFL learning. On the one hand, learners in Bosnia have different attitudes towards provided feedback and its frequency, so the students with positive attitudes towards CF and frequency of given feedback feel more motivated to continue improving in the target language in comparison with the students who are not so well disposed towards the received feedback. Generally, the participants of the current study react positively to error feedback, which means that they take it seriously, pay attention to it and appreciate it highly. Thus, they appear to believe in the benefits of CF, and they agree that error correction is necessary to help them see the difference between a wrong utterance and the right version of it. Furthermore, the results suggest that those students who are more negatively oriented towards CF might feel more embarrassed, which could eventually diminish their self-confidence and hinder their language learning achievement.

Led by these research results, teachers can apply different strategies of error correction with the aim of motivating learners in an attempt to obtain significant results in acquiring English as a foreign language. For instance, while doing oral activities, teachers can provide feedback in various ways: teachers may record learners while speaking and at the end of the activity, they might firstly ask students whether they are aware of their own errors or write both the correct and incorrect versions (the one that students have used) on the board and ask students to choose the right one. When doing this, teachers should not point to the person who made the mistake in order to avoid increasing students' sense of insecurity. To achieve this, teachers should concentrate on those mistakes made by more than one person. If teachers want to focus on individual students' mistakes, they can write individual notes to students about the mistakes they have made, with explanations or suggestions as to where to find more information about those mistakes. On the other hand, giving feedback on the written language production will depend on the characteristics of a writing task. For example, in workbook exercises, teachers can mark the responses as either correct or incorrect, while in more creative or communicative writing, such as reports, essays, stories, and so on, teachers need to approach the task with caution demonstrating their curiosity in the content of the learners' assignment rather than focusing on grammar. This could be done by responding to students' work in a form of written feedback emphasizing strong and weak sides of the composition rather than just assessing the work by a numerical grade. Although all of these methods are time-consuming, they are very useful and more beneficial than the other methods used in error correction. Additionally, in writing activities teachers can use codes or symbols and put them in the body of writing. This makes correction more helpful and less threatening than responding comments, and they are very useful while referring to issues such as spelling, verb tenses, word order, etc. However, students' interest in improving their language skills plays an important part in the process of learning as it can lead to a positive attitude towards teachers' correction and towards learning English in general. Taking this into account, teachers should find appropriate measures to arouse students' curiosity and develop their interest in EFL learning.

Limitations

The present study has some limitations. Firstly, only students and not teachers were included in the research, although teachers' attitudes towards error correction play a significant role in error correction and in L2 motivation. Consequently, as the present study did not include teachers as participants, the obtained results might not be fully representative of the reality in this foreign language context especially in terms of teachers' beliefs. Additionally, as the study did not include any classroom observation, learners' views about error correction as well as the frequency of given feedback might have been purely subjective and the results could not have been confirmed through a more objective measurement method. Thus, longitudinal observation as well as a mixed method design (both quantitative and qualitative) might lead to more concrete results about the relationship between EFL motivation and CF.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Emnijeta Ahmetovic: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing-review and editing.

Senad Bećirović: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Supervision; Validation.

Vildana Dubravac: Conceptualization; Data curation; Investigation; Methodology; Project administration; Supervision; Validation; Visualization; Writing-original draft.

Amna Brdarević Čeljo: Conceptualization; Data curation; Investigation; Methodology; Project administration; Supervision; Validation; Writing-original draft; Writing-review and editing.

REFERENCES

Ahmetović, E., Bećirović, S., & Dubravac, V. (2020). Motivation, anxiety and students' performance. *European Journal of Contemporary Education, 9*(2)*,* 271-289. https://doi.org/10.13187/ejced.2020.2.271

Al-Darei, I. S. & Ahmed, A. M. (2022). The effect of feedback type in the e-learning environment on students' achievement and motivation. *Journal of Educational Technology & Online Learning*, *5*(3), 694-705. http://doi.org/10.31681/jetol.1111527

Al-Rubai'ey, F., & Nassaji, H. (2013). Direct and indirect metalinguistic feedback: A matter of suitability rather than superiority. In M. Mahmoud & R. Al Mahrooqi (Eds.), *TEFL in the Arab World* (pp. 28–43). Sultan Qaboos University Press. https://doi.org/10.1558/isla.37949

Akram, M., & Ghani, M. (2013). Gender and language learning motivation. *Academic Research International*, *4*(2), 536-540.

Ashwell, T. (2000). Patterns of teacher response to student writing in a multi-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*(3), 227–257. https://doi.org/10.1016/j.sbspro.2014.03.484.

Bećirović, S. (2023). *Digital pedagogy: The use of digital technologies in contemporary education*. Springer.

Bećirović, S., & Brdarević-Čeljo, A., (2018). Exploring and assessing cross-cultural sensitivity in Bosnian tertiary education: Is there a real promise of harmonious coexistence? *European Journal of Contemporary Education,* 7(2), 244-256. https://doi.org/10.13187/ejced.2018.2.244

Bećirović, S., Dubravac, V., & Brdarević-Čeljo, A. (2022). Cooperative learning as a pathway to strengthening motivation and improving achievement in an EFL classroom. *SAGE Open*, *12*(1). https://doi.org/10.1177/21582440221078016

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*(3), 191–205. https://doi.org/10.1016/j.jslw.2005.08.001

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102–118. https://doi.org/10.1016/j.jslw.2007.11.004

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development.* Multilingual Matters. https://doi.org/10.21832/9781783095056

Bitchener, J. (2017). Why some L2 learners fail to benefit from written corrective feedback. In H. Nassaji & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning* (pp. 129–40). Routledge. https://doi. org/10.4324/9781315621432-10

Boscolo, P., & Gelati, C. (2013). Best practices in promoting motivation for writing. In S. Graham, C.A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (2nd ed., pp. 202- 222). The Guilford Press.

Brdarević-Čeljo, A., & Asotić, M. (2017). The influence of social context, grade level and gender on the use of language learning strategies in primary schools*. Imperial Journal of Interdisciplinary Research*, *3*(12), 7-14.

Brdarević-Čeljo, A., Bećirović, S., & Bureković, M. (2018). The use of imaginative conditional clauses by Bosnian university-level English-majoring students. *The Journal of Linguistic and Intercultural Education, 11*(2), 25-40. https://doi.org/10.29302/jolie.2018.11.2.2

Brdarević-Čeljo, A., Ahmetović, E., & Bajić, E. (2021). Variations in attitudes towards codeswitching and codeswitching frequency among multilingual speakers. *Journal of Multilingual and Multicultural Development.* Advance online publication. https://doi.org/10.1080/01434632.2021.1983580

Brdarević-Čeljo, A., Bećirović, S., & Dubravac, V. (2021). An examination of perceived reading strategy use among university level students. *European Journal of Contemporary Education*, *10*(3), 595–608. https://doi.org/10.13187/ejced.2021.3.595

Brdarević-Čeljo, A., & Dubravac, V. (2022). *Engleski u BiH: Između uglađenog i ležernog* [English in B&H: Posh or casual?]. Multidisciplinary Academic Publishing.

Brown, D. (2016). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research, 20*(4), 436-458. https://doi.org/10.1177/1362168814563200

Bruton, A. (2010). Another reply to Truscott on error correction: Improved situated designs over statistics. *System, 38*(3), 491–498. https://doi.org/10.1016/j. system.2010.07.001.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming* (2nd ed.). Routledge.

Chandler, J. (2009). Response to Truscott. *Journal of Second Language Writing*, *18*(1)*,* 57–58. https://doi.org/10.1016/j.jslw.2008.09.002

Calderón, N. (2013). *EFL learner and teacher perspectives on corrective feedback and their effect on second language learning motivation* [Unpublished Master's thesis]. McGill University.

Chen, S., Nassaji, H., & Liu, Q. (2016). EFL learners' perceptions and preferences of written corrective feedback: A case study of university students from Mainland China. *Asian-Pacific Journal of Second and Foreign Language Education, 1*(5), 1–17. https://doi.org/10.1186/s40862-016-0010-y

Chen, S., & Nassaji, H. (2018). Focus on form and corrective feedback at the University of Victoria. *Language Teaching, 51*(2), 278–83. https://doi.org/10.1017/ S026144481800006X

Choi, S. (2013). *The effects of written corrective feedback on second language writing focused on the English writing system* [Unpublished doctoral dissertation]. ProQuest Digital Dissertations.

Chua, Y. P. (2013). *Mastering research statistics*. McGraw-Hill Education.

DeKeyser, R.M. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *The Modern Language Journal, 77*(4), 501-514. https://doi.org/10.1111/J.1540-4781.1993.TB01999.X

Deci, E.L., & Ryan, R.M. (1991). A motivational approach to self: Integration in personality. In R. Dienstbier (Ed.), *Nebraska symposium on motivation*. *Perspectives on motivation* (pp. 237-288). University of Nebraska Press.

Delić, H., Bećirović, S., & Brdarević Čeljo, A. (2018). Effects of grade level and gender on foreign language learning process in Bosnian high schools. *International Journal of Educational Policy Research and Review*, *5*(6), 83-89. https://doi.org/10.15739/IJEPRR.18.010

Dervić, M., & Bećirović, S. (2019). Native and non-native EFL teachers dichotomy: Terminological, competitiveness and employment discrimination. *Journal of Language and Education*, *5*(3), 114-127. https://doi.org/10.17323/jle.2019.9746

Doughty C., & Long M. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language Learning & Technology, 7(3),* 50-80. http://dx.doi.org/10125/25214

Doughty, C. J., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. Cambridge University Press.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal, 78*(3)*,* 273-284. https://doi.org/10.1111/j.1540-4781.1994.tb02042.x.

Dörnyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualization of student motivation. *British Journal of Educational Psychology*, *70*(4), 519-538. https://doi.org/10.1348/000709900158281

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Routledge.

Dörnyei, Z., & Csizer, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research, 2*(3)*,* 203-229. https://doi.org/10.1191/136216898668159830

Dörnyei, Z., & Ushioda, E. (2011). *Teaching and Researching Motivation* (2nd ed.). Pearson.

Dubravac, V., Brdarević-Čeljo, A., & Bećirović, S. (2018). The English of Bosnia and Herzegovina. *World Englishes, 37*(4), 635-652. https://doi.org/10.1111/weng.12347

Dubravac, V., Brdarević-Čeljo, A., & Begagić, M. (2018). The presence of the British and American English variety among the first-year students at the University of Zenica. In A. Arnaut (Ed.), *Saznanje, Zbornik radova 7. međunarodnog naučno - stručnog skupa: Obrazovanje, jezik, kultura: tendencije i izazovi* [Cognition, Proceedings of the 7th International Scientific and Expert Meeting: Eduation, Language, Culture: Tendencies and Challenges] (pp. 518-528). University of Zenica.

Dubravac, V., & Latić, E. (2019). The plasticity of students' language learning beliefs: Interplay of gender, grade and educational level. *Journal of Language and Education Education*, *5*(4), 36-53. https://doi.org/10.17323/jle.2019.9732

Dubravac, V., & Skopljak, N. (2020). Foreign and multilingual language play on social sites as an identity marker. *Journal of Multicultural Discourses, 15(*1), 61-79. https://doi.org/10.1080/17447143.2019.1701678.

Ellis R., (2009). Corrective feedback and teacher development. *L2 Journal, 1*(1), 3–18. https://doi.org/10.5070/l2.v1i1.9054

Ellis, R. (2010). A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition, 32*, 335–349. https://doi.org/10.1017/s0272263109990544

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, *36*(3), 353-371. https://doi.org/10.1016/j.system.2008.02.001

Evans, N. W., Hartshorn, K. J., & Tuioti, E. A. (2010). Written corrective feedback: Practitioners' perspectives*. International Journal of English Studies, 10*(2), 47-77. https://doi.org/10.6018/ijes.10.2.119191

Ferris, D. R. (2003). *Response to student writing: Implications for second language students.* Lawrence Erlbaum Associates.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge University Press. https://doi.org/10.1017/CBO9781139524742.007

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA. Intersections and practical applications. *Studies in Second Language Acquisition, 32*(2), 181-201. https://doi.org/10.1017/S0272263109990490

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes. How explicit does it need to be? *Journal of Second Language Writing, 10*(3)*,* 161-184. https://doi.org/10.1016/S1060-3743(01)00039-X

Flynn, S. (1996). Nature of development in L2 acquisition and implications for theories of language acquisition in general. In S. Flynn & W. O'Neill (Eds), *Linguistic theory in second language acquisition* (pp. 277-294). Kluwer. https://doi.org/10.1017/S0272263100009050

Gardner, R. (2011). Language learning motivation: The student, the teacher and the researcher. *Foreign Language Education, 6(1),* 1-18.

Gardner, R. C., & MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, *43*(2), 157-194. https://doi.org/10.1111/j.1467-1770.1992.tb00714.x

Garson, G. D. (2012). *Testing statistical assumptions.* Statistical Associates Publishing.

Gass, S. M., & Selinker, L. (2001). *Second Language Acquisition, an Introductory Course*, (2nd ed.). Lawrence Erlbaum Associates.

Gass, S. (2003). Input and interaction. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 104-129). Blackwell Publishing Ltd. https://doi.org/10.1111/b.9781405132817.2005.00001.x

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciationdevelopment: Teaching English /ɹ/ to Korean adult EFL learners. *System*, *60*, 117-127. http:// dx.doi.org/10.1016/j.system.2016.06.007

Graham, S., Harris, K., & Hebert, M.A. (2011). *Informing writing: The benefits of formative assessment. A Carnegie Corporation Time to Act report.* Alliance for Excellent Education.

Graham, S., & Harris, K.R. (2013). Designing an effective writing program. In S. Graham, C.A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 3-25). The Guilford Press.

Guilloteaux, M.J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly, 42,* 55-77. https://doi.org/10.1002/j.1545-7249.2008.tb00207.x

Ha, X. V., Nguyen, L. T., & Hung, B. P. (2021). Oral corrective feedback in English as a foreign language classrooms: A teaching and learning perspective. *Heliyon*, *7*(7), e07550. https://doi.org/10.1016/j.heliyon.2021.e07550

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective*. Pearson Education International.

Hamidun, H., Hizvari, Sh., & Othman, N. F. (2012). Enhancing student's motivation by providing feedback on writing: the case of international students from Thailand. *International Journal of Social Science and Humanity, 2*(6)*,* 591-594. https://doi.org/10.7763/IJSSH.2012.V2.179

Hamzić, U., & Bećirović, S. (2021). Twice-exceptional, half-noticed: The recognition issues of gifted students with learning disabilities. *MAP Social Sciences*, *1*(1), 13–22. https://doi.org/10.53880/2744-2454.2021.1.1.13

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1)*,* 81-112. https://doi.org/10.3102/003465430298487

Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOLQuarterly, 44*(1), 84–109. https://doi.org/10.5054/tq.2010.213781

Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *The Modern Language Journal, 80*(3), 287-308. https://doi.org/10.1111/j.1540-4781.1996.tb01612.x

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge*. TEXT, 18*(3), 349-382. https://doi.org/10.1515/text.1.1998.18.3.349

Hyland, K. (2003). *Second language writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667251

Hyland, K., & Hyland, F. (2006). *Feedback on second language students' writing* (vol. 39). Cambridge University Press. https://doi.org/10.1017/S0261444806003399

Jean, G., & Simard, D. (2011). Grammar learning in English and French L2: Students' and teachers' beliefs and perceptions. *Foreign Language Annals, 44*(4), 465–492. https://doi.org/10.1111/j.1944-9720.2011.01143.x

Kalaja, P., & Barcelos, A. M. F. (Eds.). (2003). *Beliefs about SLA: New research approaches.* Kluwer. http://dx.doi.org/10.1007/978-1-4020-4751-0

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *Modern Language Journal*, *9*(1), 1-18. https://doi.org/10.1111/modl.12189

Karim, K., & Nassaji, H. (2018). The revision and transfer effects of direct and indirect comprehensive corrective feedback on English-as-a-second-language (ESL) students' writing. *Language Teaching Research*, *3*(1), 29-69. https://doi.org/10.1177/1362168818802469

Karim, K., & Nassaji, H. (2019). The effects of written corrective feedback: a critical synthesis of past and present research. *Instructed Second Language Acquisition, 3*(1), 28-52. https://doi.org/10.1558/isla.37949

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. The Modern Language Journal, 75(3), 305–313. https://doi.org/10.2307/328724

Kim, H., & Mathes, G. (2001). Explicit vs. implicit corrective feedback. *The Korea TESOL Journal, 4*(1), 1-15.

Kovačević, F., Brdarević-Čeljo, A., & Bećirović, S. (2018). Opportunities and challenges facing Bosnian high-school EFL learners. *European Researcher. Series A, 9*(4), 298-306. https://doi.org/10.13187/er.2018.4.298

Krashen, S. (1982). *Newmark's «Ignorance Hypothesis» and current second language acquisition theory* [Unpublished manuscript].

Laličić, A., & Dubravac, V. (2021). The role of reading in English language classrooms. *MAP Social Sciences*, *1*(1), 23-36. https://doi.org/10.53880/2744-2454.2021.1.1.23

Lasagabaster, D., & Sierra, J. M. (2005). What do students think about the pros and cons of having a native speaker teacher? In E. Llurda (Ed.), *Non-Native language teachers* (pp. 217-241. https://doi.org/10.1007/0-387-24565-0_12

Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing 13*(4), 285–312. https://doi.org/10.1016/j.jslw.2004.08.001

Lee, E. J. (2013). Corrective feedback preferences and learner repair among advanced ESL students. *System*, *41*(2), 217-230. https://doi.org/10.1016/j.system.2013.01.022

Leki, I. (1991). The preferences of ESL students for error correction in college level writing classes. *Foreign Language Annals, 24,* 203-218. https://doi.org/10.1111/j.1944-9720.1991.tb00464.x

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309–365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback. *System, 84*, 93–109. https://doi.org/10.1016/j.system.2019.05.006

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press.

Lopez, M. B., Steendam, E. V., Speelman, D., & Buyse, K. (2018) The differential effects of comprehensive feedback forms in the second language writing class. *Language Learning*, *68*(3), 813–850. https://doi.org/10.1111/lang.12295

Lyster, R., Lightbown, P. M., & Spada, N. (1999). A response to Truscott's `What's wrong with oral grammar correction'. *The Canadian Modern Language Review, 55*(4), 457-467. https://doi.org/10.3138/cmlr.55.4.457

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA. *Studies in Second Language Acquisition, 32*(2)*,* 265–302. https://doi.org/10.1017/S0272263109990520

Lyster R., Saito K., & Sato M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*(1), 1–40. https://doi.org/10.1017/S0261444812000365

Mašić, A., Edda P. E., & Bećirović, S. (2020). The Relationship between Learning Styles, GPA, School Level and Gender. *European Researcher. Series A, 11*(1), 51-60. https://doi.org/10.13187/er.2020.1.51

Nassaji, H. (2015). *The interactional feedback dimension in instructed second language learning: Linking theory, research, and practice.* Bloomsbury.

Nassaji, H. (2016). Interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research, 20*(4), 535–562. https://doi.org/10.1177/1362168816644940

Nassaji, H. (2017). The effectiveness of extensive versus intensive recasts for learning L2 grammar. *The Modern Language Journal*, *101*(2), 353–368. https://doi.org/10.1111/ modl.12387

Nassaji, H., & Kartchava, E. (2020). Corrective feedback and good language teachers. In C. Griffiths & Z. Tajeddin (Eds.), *Lessons from good language teachers* (pp. 151-163). Cambridge University Press. https://doi.org/10.1017/9781108774390.015

Ogrić-Kevrić S., & Dubravac V. (2017). Students' and teachers' perceptions on skills acquisition and error correction treatment in Bosnian EFL context. *Imperial Journal of Interdisciplinary Research, 3*(9)*,* 187-194.

Ortega, L. (2009). *Understanding second language acquisition*. Hodder Education.

Papi M., Rios A., Pelt H., & Ozdemir E. (2019). Feedback-seeking behavior in language learning: Basic components and motivational antecedents. *The Modern Language Journal, 103*(1), 205–226. https://doi.org/10.1111/modl.12538

Patra, I., Alazemi, A. F., Al-Jamal, D., & Gheisari, A. (2022). The effectiveness of teachers' written and verbal corrective feedback (CF) during formative assessment (FA) on male language learners' academic anxiety (AA), academic performance (AP), and attitude toward learning (ATL). *Language Testing in Asia*. https://doi.org/10.1186/s40468-022-00169-2

Polio, C., Fleck, C., & Leder, N. (1998). 'If I only had more time': ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, *7,* 43–68. https://doi.org/10.1016/S1060-3743(98)90005-4

Polz, E., & Bećirović, S. (2022). *Competency based teaching and learning*. Multidisciplinary Academic Publishing.

Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 metanalyses of feedback have different results). *Second Language Research*, *31*(2), 267–278. https://doi.org/10.1177/0267658314536436

Ramsden, P. (1992). *Learning to teach in higher education. learning to teach in higher education.* Routledge. https://doi.org/10.1080/03075079312331382498

Ribo, R., & Vildana D. (2021). The influence of the English language on BCS with the focus on the business register. *MAP Education and Humanities*, *1*(1), 20-27. https://doi.org/10.53880/2744-2373.2021.1.1.20

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, *20*(1)*,* 83–93. https://doi.org/10.2307/3586390

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for second language acquisition: A meta-analysis of the research. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 131–164). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.13.09val

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*(5), 749-761. https://doi.org/10.1037/0022-3514.57.5.749

Saito, K. (2021). Effects of corrective feedback on second language pronunciation development. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge handbook of corrective feedback in language learning and teaching* (pp. 407-428). Cambridge University Press. https://doi.org/10.1017/9781108589789.020

Schachter, J. (1991). Corrective feedback in historical perspective. *Second Language Research*, *7*(2), 89-102. https://doi.org/10.1177/026765839100202

Schulz, R. A. (1996). Focus on form in the foreign language classroom: Students' and teachers' views on error correction and the role of grammar. *Foreign Language Annals*, *29*(3), 343- 364. https://doi.org/10.1111/j.1944-9720.1996.tb01247.x

Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar teaching and corrective feedback: USA-Colombia. *The Modern Language Journal*, *85*(2), 244-258. https://doi.org/10.1111/0026-7902.00107

Schunk, D. H., Meece, J. R., & Pintrich, P. R. (2014). *Motivation in education: Theory, research and applications* (4th ed.). Pearson.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2)*,* 255 –283. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*(1), 103–110. https://doi.org/10.1177/003368829202300107

Truscott, J. (1999). The case for "the case for grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing*, *8*(2)*,* 111- 122. https://doi.org/10.1016/S1060-3743(99)80124-6

Ünlü, A (2016). Adjusting potentially confounded scoring protocols for motivation aggregation in organismic integration theory: An exemplification with the relative autonomy or self-determination index. *Frontiers in Psychology*, *7*, 1-4. https://doi.org/10.3389/fpsyg.2016.00272

Ünlü, A. (2019). Self-determination motivation theory in R: The software package SDT. *Archives of Psychology*, *3*(9). https://archivesofpsychology.org/index.php/aop/article/view/131.

Wang, W. & Li, S. (2020). Corrective feedback and learner up take in American ESL and Chinese EFL classrooms: A comparative study. *Language, Culture, and Curriculum*, *34*(1), 1-16. https://doi.org/10.1080/07908318.2020.1767124.

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past-tense forms. *Studies in Second Language Acquisition*, *32*(2), 235–263. https://doi.org/10.1017/S0272263109990519

Yaman, A., & Bećirović, S. (2016). Learning English and media literacy. *Imperial Journal of Interdisciplinary Research*, *2*(6), 660-663.

Yoshida, R. (2008). Teachers' choice and learners' preference of corrective feedback types. *Language Awareness, 17*(1), 78-93. https://doi.org/10.2167/la429.0

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, *19*(1), 79-101. https://doi.org/10.2307/3586773

https://doi.org/10.17323/jle.2023.13736

Is It True They Negatively Engage? Mixed Method Research of Student Engagement in EFL Online Classrooms

Fahad Alzahrani

University of Tabuk, Tabuk, Saudi Arabia

ABSTRACT

Background. A leading concern in teaching and learning is how to increase the degree of student engagement in learning. Within the virtual educational environment, student engagement is a real issue facing instructors and teachers. Students in online classrooms are not able to engage in the same manner as in face-to-face settings.

Purpose. This study aims to explore the impact and reception of online education on student engagement in English as a foreign language (EFL) classroom.

Method. This study adopts a mixed-method approach, in order to understand student engagement online. Longitudinal self-report surveys (SRS) filled out by 127 undergraduate students after each class session throughout a four-week period were used to assess their engagement in online language classrooms. Focus-group interview transcriptions were used to triangulate the data and provide further information about student engagement in terms of gender difference, engagement growth over time, and engagement fostering or hindrance factors in virtual learning classrooms.

Results. Analysis showed that students were generally engaged during the weeks with some variances. Cognitive-social learning engagement showed dynamics among students in virtual language classrooms. Factors such as place of engagement and students' choice of device used to access the virtual session were found to influence student engagement in online classroom learning. Male and female students generally showed similar learning engagement in the virtual classes with disparities occurring over the study period.

Conclusion. The study results will be beneficial for researchers, instructors, and policymakers who are interested in understanding student engagement and who seek to improve the teaching experience.

KEYWORDS

student engagement, negative engagement, EFL, remote learning, online education

Citation: Alzahrani F. (2023). Is It true they negatively engage? Mixed method research of student engagement in EFL online classrooms. *Journal of Language and Education, 9*(1), 41-58. https://doi.org/10.17323/jle.2023.13736

Correspondence:
Fahad Alzahrani,
fahad_alzahrani7@hotmail.com

Received: January 15, 2022
Accepted: Mach 22, 2023
Published: March 31, 2023

INTRODUCTION

There is a strong consensus that engagement is essential for individual learning. Many online language teaching instructors, however, complain that student enthusiasm and engagement in virtual classrooms are not as lively as in traditional classrooms. Students who typically engage in the traditional mode face distinctive challenges specific to e-learning. The flexibility and convenience offered in online learning are acknowledged, but students may hold negative attitudes and resistance to successful learning in virtual environments. As Oraif and Elyas (2021) claim, the language learning experience could be challenging for students and ineffectual in meeting their needs. Consequently, they might not engage with the course. This has raised a leading concern related to keeping students as motivated and engaged in online classrooms as they are in the traditional setting (Bolliger & Martin, 2018).

Online student engagement differs from engagement in face-to-face learning in numerous ways. O'Shea *et al*. (2015) asserted, "When shifting to online contexts, engagement takes on different manifestations, due to the lack of face-to-face contact and the ways in which teaching and learning are mediated through tech-

nology" (p. 43). As a result, without this critical component in place, students in the online environments report a lack of interest, and thus, a lower quality of work and less overall satisfaction (Martin, 2019).

Defining Engagement

Engagement is a psychological construct that is considered "a proverbial new kid on the block" (Reschly & Christenson, 2012, p.14), especially when compared with motivation, which has received a high level of interest (Boo et al., 2015). The vast body of research in motivation is claimed to have "stolen the show and diverted attention away from engagement" (Mercer & Dörnyei, 2020, p. 5). Mercer and Dörnyei contended that "motivation is undoubtedly necessary for 'preparing the deal', but engagement is indispensable for sealing the deal" (p.6).

The term "learner engagement" has been construed differently. One key feature of engagement that has consistently reiterated across its conceptualizations is 'action'. The action concept of engagement resonates with definitions such as that of Skinner et al. (2009) who viewed engagement as "energized, directed, and sustained actions" (p. 225) and that of Reeve who described, "the extent of a students' active involvement in a learning activity" (2012, p.150). In the arena of second language (L2) learning, engagement was construed by Ellis (2010) to describe learner responses to corrective feedback.

The multifaceted nature of engagement has been agreed upon. As Fredricks et al. (2004) proposed, the multidimensional construct of engagement comprises three main components: behavioral, cognitive, and emotional. The behavioral dimension involves the active participation of students. The cognitive dimension comprises a psychological investment by trying to understand complex ideas, self-regulation, exerting effort for solving challenging tasks and using deep learning. In classroom settings, the emotional dimension is "often manifested in learner's personal affective reactions in the target language-related activities or tasks" (Zhou et al., 2021, p. 88). An additional aspect of engagement was suggested as one main aspect of engagement. This aspect underlines the social interaction with peers and teachers, collaborative learning, sharing ideas and maintaining the relationships in the class. Importantly, these four dimensions are interconnected, despite the possibility of becoming positively engaged in one dimension, while disengaged or negatively engaged along another dimension.

In the realm of online education, and since many students tend not to use their web cameras (Castelli & Sarvary, 2020), many professors and instructors seem to believe that students are actually faking their engagement. Fake engagement in this manner is "reflected in behaviors that are made consciously or unconsciously, by learners to achieve an outside appearance of being attentive and on-task; however, in reality, their internal states, are not congruent and, for diverse motives, they may be complying or just merely pretending compliance" (Mercer et al., 2021, p. 145). To put it differently, fake engagement occurs when there is behavioral engagement without being cognitively and emotionally engaged.

Related Works

Increased attention to the study of learner/student engagement is reflected in a recent large and growing body of research. A handbook-length work, such as Christenson et al. (2012) or richly edited books, such as Mercer & Dörnyei (2020), Hiver et al. (2020), and Quaye et al. (2019) have been published on this area. This underlines the issue that understanding how to engage learners remains a pressing concern.

In the language classroom, several studies have made notable contributions related to the domain of learner engagement (see Oga-Baldwin, 2019; Philip & Duchesna, 2016, for review). In L2 classrooms, researchers such as Svalberg (2009, 2018) was one of the pioneers who attempted to understand the role of engagement in language acquisition and learning. He examined engagement from a very narrow-angle, namely, engagement with language (EWL), a model through which learners can develop language awareness.

Similar to the situation in traditional classrooms, one of the most prominent questions involves keeping students motivated and engaged in the online setting (Bolliger & Martin, 2018). Thus, numerous studies have targeted student engagement in the EFL virtual classroom (e.g., Bolliger & Martin, 2018; Almossa, 2021; Kim, 2021; Martin, 2019; Oraif & Elyas, 2021; O'Shea et al., 2015; Soffer & Cohen, 2019; Yundayani et al., 2021). The results of these studies suggest that engagement in a non-face-to-face learning environment is impacted. Therefore, several strategies can be utilized in the online learning environment to enhance student engagement (Bolliger & Martin, 2018).

Studies such as Almossa (2021), Oraif and Elyas (2021), Yundayani et al. (2021), and Khlaif et al. (2021) investigated student engagement during the shift to emergency remote teaching. Numerous contexts have been targeted in exploring student engagement. In the Saudi context, a few contributions, such as Al-Bogami & Elyas (2020), Almossa (2021), and Oraif & Elyas (2021) were made. Almossa (2021) investigated the experiences and opinions of college students shared on their Twitter accounts about their engagement with online learning and assessment. By analyzing tweet posts, the findings suggested that cognitive, behavioral, and affective engagement with learning have been considerably impacted by online assessment. According to Almossa, this was attributed to several challenges that arose due to the

sudden shift in the learning mode, such as communication, fairness, technical and assessment issues.

Among the few studies investigating online student engagement in the Saudi context, there is the study by Oraif & Elyas (2021) which explores learners' level of engagement in 'My School' (Madrasati). This is a school platform in Saudi Arabia, where the results uncover a high level of engagement among EFL learners. Applied exclusively to female learners, Al-Bogami & Elyas (2020) also investigated the effect of handheld devices in EFL classrooms on student engagement. Through a mixed-method approach, they studied the use of the iPad and its apps for EFL engagement and learning among 20 middle school students. The findings suggest that the apps bolster engagement level and learning compared to traditional teaching paradigms. The focus in both studies - Bogami & Elyas (2020) and Oraif & Elyas (2021) - was on young female learners (i.e., middle school and high school age students). Some studies focused on gender differences and online EFL classroom engagement. These studies have revealed that there are no significant impacts on gender variation and engagement in EFL virtual learning environments (Almusharraf & Almusharraf, 2021; Devrim & Irem, 2020; Benhadj, 2021). There is a persisting stereotype that girls are better at language learning than boys. Is this stereotypic belief still in place is thus one of the aims of the present study: to further investigate gender differences and engagement for college level students learning English in online classrooms.

In terms of the assessment of student engagement, Zhou et al. (2021, p. 80) confirmed that "[t]o date there is no single instrument that is accepted for use across contexts – just as there is none that is accepted as a field-specific measure of engagement." Yet, different approaches were followed in the sciences of education and learning for measuring student engagement. Examples of these measurement methods include self-reported surveys and questionnaires (e.g., Oraif & Elyas, 2021), direct observations, and interviews (e.g., Yundayani et al., 2021). The most frequent methods used for measuring student engagement are self-reported surveys (SRS) and questionnaires, as used in the current study. In SRS and questionnaires, students are "presented with items describing different facets of engagement and are directed to choose the response from a range of possibilities that best describes them" (Zhou et al., 2021, p. 80).

Other approaches employed in previous literature for the purpose of assessing student engagement include traditional methods, such as self-reports, teacher ratings of students, observations of students' performances, and measurement of bio-signals. Alternatively, students' reflections on their experiences and opinions on social media platforms such as Twitter, which are likely to mirror their engagement explicitly or implicitly with online learning, have given rise to some attention (for example, Almossa, 2021). A new strategy which uses learners' facial expressions to assess the level of engagement in non-face-to-face learning situations was also proposed by Kim (2021). For further measures, Zhou et al. (2021) explored the past, present, and future of how learner engagement can be assessed in the L2 classroom.

That being said, this study casts light on the engagement of college level students in the online EFL classroom in the Saudi context. Specifically, this study aims to explore the effect of online teaching and virtual reality on language classroom engagement of college students during the pandemic era. This study was motivated by one of the key findings of Alzahrani et al. (2022) study, which recommended that the reasons behind students not engaging positively in EFL online classrooms be explored.

Research Questions

The study aims to gain comprehensive answers and insights to the following research questions (RQ):

(1) How do Saudi students evaluate their engagement in an online classroom?

(2) Which aspect of engagement is reported by Saudi students to be highest in the online classroom, and which is the lowest?

(3) Is there a gender difference in student engagement within these online classrooms?

(4) Does student engagement grow over time in online classrooms?

(5) What factors do Saudi students report to be enhancing or hindering their engagement in online classrooms?

METHOD

Background

This study was conducted in a Saudi public university's compulsory junior-year writing course. The data was drawn from an eight-week summer term, from May to July 2021. The course was offered in two sections: one section was for female students and taught by a female instructor, whereas the other section was for male students and taught by the researcher, as a male instructor. The female instructor's role was only to collect the data and provide it to the researcher; she was not involved as a researcher in this study. The time of each of the sessions was 3 hours, and twice a week. This course was offered fully online by the Language and Translation Department, for both computer science major students and information technology major students.

In order to carry out this research, a completely online course was designed in which course content as well as

all online materials and sources were developed by the researcher who taught the male section. Both sections shared the same course content, assessment structure, and tools. After setting up the course on Blackboard, the university's official learning management system (LMS), a course copy was exported and shared with the female instructor, in order to maintain the same quality of instruction. All course sessions were synchronously taught. Typical pedagogical activities in almost all sessions included instructor presentations, group discussions, question sessions with instructor feedback, quizzes, and exams, YouTube video clips, assignment tutorials, Collaborate Ultra Breakout group, and reference to the course textbook.

Study Sample

Convenience sampling was used in the study. The students were all of the Saudi nationality whose second language is English. They had different levels of English proficiency, but were mostly intermediate learners. This was determined by the successful completion of the English language program, offered by the university as a requirement, qualifying students for B 1 (intermediate English) in the Common European Framework of Reference (CEFR). The students' native language was Arabic. Total enrollment at the time of the research (i.e., summer 2021) was 127 students (both sections).

Data Collection and Procedures

This was a mixed-method study which utilized both quantitative and qualitative approaches. The mixed methods used in this research are primarily for a triangulation strategy, which seeks corroboration, convergence, and correspondence of results from different methods (Greene *et al*., 1989). The following three instruments were used:

Self-Report Surveys (SRS)

For the quantitative analysis, a longitudinal survey, aka self-report survey (SRS), was used. A sample of items is in Appendix (A). A longitudinal survey is defined as "one that collects data from the same sample elements on multiple occasions over time" (Lynn, 2009, p. 1). In the literature, this self-report instrument has been used commonly and shown good reliability (e.g., Fuller et al. (2018); Hiver et al. (2020) and Oga-Bakdwin & Fryer (2021).

The SRS was distributed over the first four weeks of the term amongst students who were enrolled in the course. The SRS consisted of a total of 34 five-point Likert scale statements divided into sections according to the four sub-scales (i.e., behavioral, cognitive, emotional, and social), as well as four questions at the very beginning for eliciting demographic information. Responses ranged from 1 (referring to "Very untrue of me") to 5 ("Very true of me"). Items of the SRS were adapted from validated instruments used in two recent research studies: Hiver et al. (2020) and Oga-Bakdwin & Fryer (2021).

The SRS was created online using Google forms in which all students could complete the questionnaire online. Students were asked to take the survey after each class session throughout the first 4 weeks (i.e., eight class sessions in total). This period was selected in the aims of examining this phenomenon in no less than half of the semester. In addition, as suggested by Lynn (2009), "the longer the period of time over which a longitudinal survey collects data, the richer and more valuable the data are likely to be." (p. 14). The SRS was developed bilingually (in English and Arabic). Finally, this instrument was valid and reliable as shown by Cronbach's alpha coefficient testing for reliability as internal consistency (Cronbach, 1951) (score= .837 which is high reliability (>.7)).

Focus Group Interviews

Students from both sections were invited immediately after the last session of the term to participate in a focus group. The purpose of these exploratory focused, in-depth group interviews was to further understand and triangulate the quality and character of learner engagement, in particular, their perspectives on why and when they were engaged or not.

Two interviews were conducted in week 8. One interview was dedicated to male students, and another was exclusive to female students. This was to encourage the participants to speak freely and comfortably. The interviews were semi-structured in which open-ended questions were formulated and used. Each interview lasted about 20 minutes in length. Interview questions were in the students' native language of Arabic, and the students were encouraged to answer in Arabic or English, whichever was convenient for them, in order to ensure full understanding and expression. Generally, students preferred to use English to answer interview questions.

Interview questions were piloted on two students with similar characteristics, in order to check for meaning, obtain feedback on how interview questions come across, help revise the question structures, decide whether more need to be included, or some deleted, and "to learn about [my] effectiveness as a moderator—do [I] need to modify the amount of involvement [I am] having in the interview?" (Breen, 2006, p. 471).

Both interviews were audio-recorded. In order to minimize interviewer effects, students were told that their responses will not have any effect on their course grades. As an instructor and a researcher, my own point of view on the topic or my relation to the students did not influence my way of moderating the interviews. In fact, the female students were interviewed by me and the male students were interviewed

by the other instructor. As recommended by Breen (2006), "it is always important to demonstrate reflexivity as an interviewer" (p. 473). Interview transcriptions were later looked at for theme extraction in light of the research questions.

Data Analysis

For the quantitative data, descriptive statistics were employed using SPSS® (v. 24), in order to analyze students' responses to the SRS items in terms of frequency, mean, percentage, and standard deviation to assess student engagement level during online learning. T-test, one-way ANOVA, repeated measures ANOVA, and correlation analyses were used in the analysis of the data as well. Further analysis was carried out, in order to determine different sources and post-hoc multiple comparisons were utilized using the Least Significant Difference (LSD) test.

In order to identify and organize shared experiential themes in qualitative data, thematic analysis, as proposed by Braun & Clarke (2006), was used. This methodology was used to interpret data collected from the semi-structured interviews. Interview data was analyzed and reported using six stages for developing thematic analysis: (1) establishing familiarity with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report.

Ethical Considerations

A high standard of ethical considerations was followed throughout the process of conducting this research. All data was generated with students' explicit permission. No participant was coerced into participating in the study. Every student could withdraw from the study at any point in time without repercussions.

According to Cohen et al. (2007), since interviews might be considered an intrusion into the personal lives of participants, especially with regard to the sensitivity of questions, ethical considerations were maintained at all stages of the interview process. Further oral consent was obtained at the beginning of the interview session, and each participant was informed that the session was being audio/video recorded. The follow-up interviews were conducted with the participants one week after the course finished, in order to reduce the possibility that students would be unduly influenced. In addition, for purposes of anonymity, pseudonyms were used to identify participants in all research processes, and any identifying markers were removed to protect students' identities.

RESULTS

This section commences with the quantitative statistical results and presents a detailed exposition of the observed data.

Results of SRS

Participants in the SRS

A total of 127 students (72 females and 55 males) took part in the surveys of the current study. No less than 50% of students attending each class session participated in the SRS. In fact, in the end-of-class surveys their participation reached 95% in many of the class sessions. Table 1 presents student frequency of participation in filling out the survey over the course of four weeks. No student's responses were excluded from the analysis.

"In-home" engagement versus "out-of-home" engagement were possible factors that might have affected student online engagement. The majority of the students reported they attended the online sessions from home (92.5%). Only 4.67% accessed sessions in their cars during class time or

Table 1

Student Record of Participation in In-Class Surveys Based on Class Sessions

Session#	Male students (/55)	Female students (/72)	Total
1	40 (91%)	62 (95%)	102
2	50 (91%)	62 (95%)	112
3	38 (73%)	43 (66%)	81
4	42 (93%)	58 (90%)	100
5	41 (85%)	46 (70%)	87
6	43 (86%)	42 (60%)	85
7	41 (91%)	38 (61%)	79
8	41(82%)	34 (50%)	75
			721

from public places such as coffee shops, hospitals, or restaurants. On the other hand, in all four weeks, a high percentage of students (49.1%) used personal computers (i.e., desktop or laptop) to access the virtual sessions. 24.7% of them used cell phones, while 7.9% used tablets to access the virtual sessions. The following findings are organized according to the guiding research questions.

Analysis of SRS

RQ1 which asked: *How do students evaluate their engagement in online classrooms?* was answered through the production of means, and standard deviations for the survey statements. As displayed in Table 2, students showed a generally high classroom engagement. That is, engagement in writing classes among Saudi college learners, when online learning was adopted, was generally at a high level of 'True of me' (*mean*=3.88, SD=0.44). Here, students had a similar engagement in general in all four weeks, indicating "True of me" with a mean score of 3.84 (SD=0.37) in week one, 3.89 (SD=0.45) in week two, 3.89 (SD=0.46) in week 3, and 3.91 (SD=0.51) in week four.

Table 2

Means and Standard Deviations for SRS Data

Weeks	Mean	SD
One	3.84	0.37
Two	*3.89*	*0.45*
Three	*3.89*	*0.46*
Four	*3.91*	*0.51*
All 4 weeks	*3.88*	*0.44*

To answer RQ2, it was found that, generally speaking, student behavioral engagement was rated the highest in all four weeks indicating "True of me" with a mean score of 4.16 (*SD*=0.63). This was followed by emotional engagement (*mean*=3.98; *SD*=0.48) and then social engagement (*mean*=3.74; *SD*=0.88). SRS data also showed that students were cognitively the least engaged during the eight virtua- classes (*mean*=3.43; *SD*=0.42). Nevertheless, this indicates a "True of me" (see Table 3). A closer look into the data showed that although student cognitive engagement was ranked the highest in week 1 (*Mean*=4.33, *SD*=0.43), it declined to the lowest rate in weeks two, three, and four (*Mean*=3.45, 3.48 and 3.49, respectively). However, the behavioral type of engagement was the highest across weeks 2, 3, and 4 (*Mean*=4.14, 4.17, and 4.18, respectively). In fact, all four student engagement aspects ranked similarly across weeks 2, 3, and 4.

To answer RQ3: *Is there a gender difference in students' engagement*?, analysis of an independent sample t-test (see Tables 4, 5, 6, and 7) showed that, in week 1, there was no significant difference between male and female students regarding engagement in online classroom as a whole (P>.05). However, significant differences were found between male and female students in the behavioral type of engagement (*P*<.05). The difference was in favor of females (*mean*=4.26) as opposed to (*mean*=4.02) for males (Table 4). Put simply, female participants showed higher engagement in behavioral activities such as effort, participation, initiative-taking, and persistence. Moreover, in week 1, significant differences were found between male and female students in other types of engagement such as cognitive and social engagements (*P*<.05). This time, male students were cognitively and socially more engaged than their female counterparts (*mean*=3.43, 3.87, respectively for male students) (*mean*=3.27, 3.39, respectively for female students).

In week 2, as shown in Tables 6, no significant difference was found between males and females in regard to their engagement in the online classroom as a whole (P>.05). However, significant differences were found between male and female students in emotional engagement (*P*<.05). The difference was in favor of females (*mean*=4.08) in contrast to (*mean*=3.90) for males. In week 3, significant differences were found between male and female students in cognitive engagement (*P*<.05). The difference was in favor of females (*Mean*=3.55) in contrast to (3.41) for males (see Table 7). However, in week 4, no significant differences were found

Table 3

Student Engagement in Online Classroom Based on Engagement Domains

Weeks	Week 1			Week 2			Week 3			Week 4			Total		
Subscales	**Mean**	**SD**	**Rank**	**Mean**	**SD**	**Rank**	**Mean**	**SD**	**Rank**	**Mean**	**SD**	**Rank**	**Mean**	**SD**	**Rank**
Behavioral	4.16	0.56	2	4.14	0.63	1	4.17	0.64	1	4.18	0.70	1	4.16	0.63	1
Cognitive	4.33	0.43	1	3.45	0.40	4	3.48	0.41	4	3.49	0.45	4	3.43	0.42	4
Emotional	3.97	0.39	3	4.00	0.49	2	3.97	0.51	2	3.98	0.56	2	3.98	0.48	2
Social	3.59	0.92	4	3.82	0.85	3	3.73	0.86	3	3.84	0.86	3	3.74	0.88	3

Table 4

Independent Sample T-Test Results (Week 1)

	Students	N	Mean	SD	t	P value
General	Male	90	3.84	0.38	0.093	.926 ns
	Female	124	3.83	0.36		
Behavioral	Male	90	4.02	0.63	3.054	.003**
	Female	124	4.26	0.48		
Cognitive	Male	90	3.43	0.38	2.272	.006**
	Female	124	3.27	0.44		
Emotional	Male	90	3.95	0.38	0.647	.518 ns
	Female	124	3.98	0.40		
Social	Male	90	3.87	0.87	3.918	.000**
	Female	124	3.39	0.91		

*Note: ** significant at .01 level , * significant at .05 level , ns= not significant*

Table 5

Independent Sample T-Test Results (Week 2)

	Students	N	Mean	SD	t	P value
General	Male	80	3.82	0.46	1.804	.073 ns
	Female	101	3.95	0.45		
Behavioral	Male	80	4.10	0.62	0.766	.445 ns
	Female	101	4.17	0.63		
Cognitive	Male	80	3.41	0.41	1.248	.214 ns
	Female	101	3.48	0.39		
Emotional	Male	80	3.90	0.50	2.459	.015*
	Female	101	4.08	0.48		
Social	Male	80	3.75	0.84	1.016	.311 ns
	Female	101	3.87	0.85		

*Note: * significant at .05 level , ns= not significant*

between male and female students regarding engagement in the online classroom as a whole (*P*>.05). This indicates that both male and female participants showed similar engagement in the online classroom.

ANOVA was employed to examine any significant difference between students regarding engagement in online classrooms, i.e. in answers to RQ4: Does student engagement grow over time? As shown in Table 8, analysis did not report a significant F statistic effect regarding engagement in online classrooms as a whole and regarding behavioral and emotional engagement. Nevertheless, there was a significant F statistic effect reported in regard to the cognitive engagement type and the social engagement (*P*<.05) indicating significant differences in engagement across all weeks.

As evident in Table 9, the post-hoc test revealed that significant differences in cognitive engagement occurred between week 1, on the one hand, versus weeks 2, 3, and 4, on the other hand. The difference trend was in favor of cognitive engagement in week 2, week 3 and week 4 due to the highest mean (3.45, 3.48, 3.49) respectively. There is a steady increase in cognitive engagement from week to week, reaching the highest average in week 4. The conclusion can be drawn that cognitive engagement grows over time. Furthermore, results analysis (Table 9) reported significant differences in social engagement occurring between week 1, versus weeks 2 and 4. The difference trend was in favor of social engagement in week 2 and week 4 due to the highest mean (3.82, 3.84), respectively. This concludes that there is a steady increase in cognitive engagement from week to week, reaching the highest average in week 4. Students' social engagement does grow over time as well.

Another layer of inferential analysis was performed, in order to identify any significant differences that may exist be-

Table 6

Independent Sample T-Test Results (Week 3)

	Students	N	Mean	SD	t	P value
General	Male	84	3.82	0.45	1.786	.076 ns
	Female	88	3.95	0.45		
Behavioral	Male	84	4.11	0.64	1.379	.170 ns
	Female	88	4.24	0.64		
Cognitive	Male	84	3.41	0.39	2.236	.027*
	Female	88	3.55	0.42		
Emotional	Male	84	3.91	0.51	1.435	.153 ns
	Female	88	4.02	0.49		
Social	Male	84	3.67	0.88	0.860	.391 ns
	Female	88	3.78	0.85		

*N.B. *significant at .05 level, ns= not significant*

Table 7

Independent Sample T-Test Results (Week 4)

	Group	N	Mean	SD	t	P value
General	Male	82	3.90	0.49	0.252	.801 ns
	Female	72	3.92	0.52		
Behavioral	Male	82	4.22	0.69	0.891	.374 ns
	Female	72	4.12	0.71		
Cognitive	Male	82	3.46	0.44	0.879	.381 ns
	Female	72	3.52	0.45		
Emotional	Male	82	3.96	0.54	0.600	.549 ns
	Female	72	4.01	0.59		
Social	Male	82	3.79	0.91	0.793	.429 ns
	Female	*72*	*3.90*	*0.79*		

Note: ns= not significant

tween the participants in terms of the specific locations of attending the online session (e.g., from home or outside the house), as well as choice of device (e.g., laptop, tablet, cellphone) used to access the virtual session. The results are presented below.

On conducting a one-way ANOVA test, as seen in Table 10, a relationship was established between student engagement and the location where they attended the virtual sessions. There is a significant F statistic effect regarding engagement in online classrooms, depending on the location of attending the online sessions ($P<.05$). In this regard, the ANOVA test reported a significant F statistic effect regarding the two types of engagement: behavioral and emotional ($P<.05$). No significant F statistic effect was reported in cognitive engagement or social engagement (P>.05).

Results of post hoc comparisons revealed that significant differences in engagement in online classrooms occurred between students attending the online sessions from home, and those from other locations such as a car or public places. Students who reported they were at home when attending the online session had higher average points (*mean*=3.90), when compared to students attending from other locations. Significant differences were also found in both the behavioral and the emotional engagement types. The differences in both were in favor of students who reported attending the virtual classes from their homes, with mean scores of 4.20 and 4.00, respectively. It can thus be concluded that students attending the online sessions from home had a higher engagement level in online classrooms, and that their behavioral and emotional engagements are significantly associated with the location where they attend the virtual classroom.

Table 8

ANOVA Results

		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	.554	3	.185		
General	Within Groups	140.754	717	.196	.940	.421
	Total	141.307	720			
	Between Groups	.164	3	.055		
Behavioral	Within Groups	281.835	717	.393	.139	.936
	Total	282.000	720			
	Between Groups	2.982	3	.994		
Cognitive	Within Groups	126.652	717	.177	5.628	.001**
	Total	129.634	720			
	Between Groups	.149	3	.050		
Emotional	Within Groups	168.786	717	.235	.211	.888
	Total	168.936	720			
	Between Groups	7.366	3	2.455		
Social	Within Groups	550.657	717	.768	3.197	.023*
	Total	558.023	720			

*Note: ** significant at .01 level , * significant at .05 level , ns= not significant*

Table 9

Post-HOC Multiple Comparisons

		Mean	Week 1	Week 2	Week 3	Week 4
	Week 1	3.33	-	-	-	-
	Week 2	3.45	0.12*	-	-	-
Cognitive	Week 3	3.48	0.15**	-	-	-
	Week 4	3.49	0.15**	-	-	-
	Total	3.43		-	-	-
	Week 1	3.59		-	-	-
	Week 2	3.82	0.22*	-	-	-
Social	Week 3	3.73		-	-	-
	Week 4	3.84	0.25**	-	-	-
	Total	3.74		-	-	-

*Note: ** significant at .01 level , * significant at .05 level*

The subsequent analysis examined student engagement in relation to their choice of device (e.g., laptop, tablet, cellphone). Analysis revealed an association between student engagement and the devices used by the student to access virtual classes. ANOVA results, in Table 11, reported a significant F statistic effect regarding engagement in online classroom according to devices used by students to access virtual sessions, F(10.243), (*P*<.05) and significant F statistic effect in three types of engagement, namely, behavioral, emotional, and social engagement (*P*<.05). There was no significant main effect on student cognitive engagement.

An LSD post-hoc analysis (Table 12) showed that students who attended online sessions using a personal computer, whether a laptop or a desktop, were significantly more engaged (*m*=3.93) than those who joined the online classroom from cellphones (m=3.78). In fact, data showed that those who reported they used tablets to log in to sessions had the highest levels of engagement (*mean*=4.00), when compared

Table 10

Differences in Engagement according to the Location of Attending the Online Sessions

		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	4.111	7	.587		
General	Within Groups	137.143	711	.193	3.045	.004**
	Total	141.254	718			
	Between Groups	14.204	7	2.029		
Behavioral	Within Groups	267.127	711	.376	5.401	.000**
	Total	281.331	718			
	Between Groups	1.112	7	.159		
Cognitive	Within Groups	128.486	711	.181	.879	.522
	Total	129.598	718			
	Between Groups	4.805	7	.686		
Emotional	Within Groups	163.454	711	.230	2.986	.004**
	Total	168.259	718			
	Between Groups	4.395	7	.628		
Social	Within Groups	553.266	711	.778	.807	.582
	Total	557.661	718			

*Note: ** significant at .01 level , ns= not significant*

Table 11

ANOVA Results

		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	3.912	2	1.956		
General engagement	Within Groups	129.095	676	.191	10.243	.000**
	Total	133.008	678			
	Between Groups	8.430	2	4.215		
Behavioral	Within Groups	250.979	676	.371	11.354	.000**
	Total	259.410	678			
	Between Groups	.655	2	.328		
Cognitive	Within Groups	122.234	676	.181	1.812	.164
	Total	122.890	678			
	Between Groups	4.235	2	2.118		
Emotional	Within Groups	155.905	676	.231	9.182	.000**
	Total	160.140	678			
	Between Groups	5.308	2	2.654		
Social	Within Groups	525.482	676	.777	3.414	.033**
	Total	530.790	678			

*Note: ** significant at .01 level , ns= not significant*

to users of mobile phones or even personal computers. Participants, thus, differed in their learning engagement according to their choices of device to attend the online classroom.

Significant differences were reported in behavioral, emotional, and social engagement aspects in connection to students' choice of devices. As displayed in Table 14, students who used tablets to join the sessions showed better engagement in terms of behavioral (*mean*=4.31), emotional (*mean*=4.09), and social engagement (*mean*=3.90). Students who used cellphones were the least engaged behaviorally, emotionally, and socially compared to those using other device types (*mean*= 4.02, 3.88, 3.61, respectively).

An LSD post-hoc analysis (Table 12) showed that students who attended online sessions using a personal computer, whether a laptop or a desktop, were significantly more engaged (m=3.93) than those who joined the online classroom from cellphones (m=3.78). In fact, data showed that those who reported they used tablets to log in to sessions had the highest levels of engagement (*mean*=4.00), when compared to users of mobile phones or even personal computers. Participants, thus, differed in their learning engagement according to their choices of device to attend the online classroom.

Significant differences were reported in behavioral, emotional, and social engagement aspects in connection to students' choice of devices. As displayed in Table 14, students who used tablets to join the sessions showed better engagement in terms of behavioral (*mean*=4.31), emotional (*mean*=4.09), and social engagement (*mean*=3.90). Students who used cellphones were the least engaged behaviorally, emotionally, and socially compared to those using other device types (*mean*= 4.02, 3.88, 3.61, respectively).

Since there are multiple measurement opportunities for the dependent variables (i.e., engagement in online classroom and its subscales), GLM repeated measures was employed to examine student engagement growth over time. The results of GLM repeated measures are presented in Table 13.

Gender was selected as a between-subjects factor. Table 13 contains data about the within-subject factor, time, and its interactions with the independent variable (gender). The main effect for engagement in online classroom is a test of the null hypothesis that all levels of within-subjects factor are equal. It is a test of the hypothesis that engagement in online classroom levels are equal across the four weeks. There appear to be statistically significant differences between engagement in online classroom levels across the four weeks. The F statistic (350.320) and its associated significance level (P<.001) indicate that can reject this hypothesis as false.

Results of Focus Group Interviews

Description of Participants

In total, 18 students participated in two focus group online interviews. Eight participants were female, and ten were male, all of them aged between 18 and 24 years. The findings presented below encapsulate the experiences of online students in engaging with the undergraduate writing course.

Table 12

Results of Post-HOC Multiple Comparisons Using (LSD) Test

		Mean	Differences
General	Cell phone	3.78	1 vs. 2 (.15**) 1 vs. 3 (.22**)
	Computer	3.93	
	Tablet	4.00	
Behavioral	Cell phone	4.02	1 vs. 2 (.23**) 1 vs. 3 (.28**)
	Computer	4.25	
	Tablet	4.31	
Emotional	Cell phone	3.88	1 vs. 2 (.16**) 1 vs. 3 (.22**)
	Computer	4.04	
	Tablet	4.09	
Social	Cell phone	3.61	1 vs. 2 (.16*) 1 vs. 3 (.29*)
	Computer	3.77	
	Tablet	3.90	

*Note: **significant at .01 level , *significant at .05 level*

Table 13

Statistics for the Effects in the Model

Tests of Within-Subjects Effects						
	Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Time	Sphericity Assumed	98.389	3	32.796	350.320	.000**
	Greenhouse-Geisser	98.389	2.013	48.884	350.320	.000**
	Huynh-Feldt	98.389	2.021	48.679	350.320	.000**
	Lower-bound	98.389	1.000	98.389	350.320	.000**
Time *Gender	Sphericity Assumed	1.198	3	.399	4.266	.005**
	Greenhouse-Geisser	1.198	2.013	.595	4.266	.014*
	Huynh-Feldt	1.198	2.021	.593	4.266	.014*
	Lower-bound	1.198	1.000	1.198	4.266	.039*
Error (Time)	Sphericity Assumed	201.935	2157	.094		
	Greenhouse-Geisser	201.935	1447.128	.140		
	Huynh-Feldt	201.935	1453.222	.139		
	Lower-bound	201.935	719.000	.281		

*Note: ** significant at .01 level, * significant at .05 level*

Table 14

Themes of interview transcripts

#	Theme	*Frequency*	%
1	Students' self-evaluation of engagement in the online course	18	100
2	Factors attributed to students' high engagement	6	33
3	Students' perspectives on engagement in online classrooms versus in a non-face-to-face classrooms	13	72
4	Students' engagement throughout the semester	11	61
5	Reflective Moments when students felt fully engaged	8	44
6	Influence of location/place of class attendance on students' engagement	15	83

Analysis of Semi-Structured Interviews

The primary purpose of conducting student interviews was to answer RQ5 and to triangulate the data of other research questions through different sources. From the focus group interviews, six overarching themes relating to online student engagement were identified and constructed during the analytical process. As shown in Table 13, the themes were: 1) student self-evaluation of engagement in the online course; 2) Factors attributed to a high level of student engagement; 3) Student perspectives on engagement in online classrooms versus in a non-face-to-face classrooms; 4) Student engagement throughout the semester; 5) Reflective moments when students felt fully engaged; and 6) Influence of presence/place of class attendance on student engagement. These themes are presented below using paraphrases and individual direct quotes. Some quotes were altered to remove filler language (e.g., um, like) and provide clarity without changing the meaning of quotes.

Theme 1: student self-evaluation of engagement in online learning. Across the focus groups, students described how well they were engaged in the course. The students showed a high level of satisfaction of their engagement with the online class. In fact, they all reported that they were engaged well in the course. One student answered, "I personally find my engagement in this course to be excellent." Another student rated, "I would give my engagement 85%".

Theme 2: Factors attributed to a high level of student engagement in the online course. This theme accounts for students' perspectives on the factors that attributed to their high level of engagement in the online classroom. Based on interview data, pleasant experiences in the online course which led to a high level of engagement were attributed to four factors. These factors are discussed below.

1. The first factor is in regards to the engaging nature of teaching, which Hipkins (2012) discussed saying, "Teach-

ers' curricular intentions and the manner they construct learning opportunities in the classroom have an impact on engagement" (p. 441). In this respect, one student said, *"Class sessions tended to be enthusiastic in nature. And, each class is even more engaging than the one before".*

2. The interactive nature of the course and lectures was one point which facilitated their engagement. This course was demanding, interactive, and hands-on when compared to other courses. Students were expected, for example, to discuss the topic in hand, respond to discussion board questions and post their answers to Blackboard forums. This point was described by one of the interviewees *"Classes contained a lot of interactive components, such as homework you assign, screen sharing and its accompanying activities, and student discussion*".

3. One key factor underlined by the students in regard to their classroom engagement was the well-designed, pedagogical methods in the course. One student attributed his high level of engagement to course assignments and in-class activities in which he thought they *"motivate students to learn and participate, and ultimately, were reflected on students' outcomes".*

4. The instructor's role as a motivator was reported by students in both groups to be central in enhancing classroom engagement. One student believed "*The instructor plays a role in connecting students with the course and materials. Whenever the instructor acts harshly, students are more likely to feel alienated. However, in this course, I, frankly, can say I enjoyed the class. I made sure I do the in-class activities, because we were motivated to do so. So, yes, generally, it's all about the instructor's efforts in engaging the students*". Similarly, students valued the professor's role on this matter, and one student commented, "*Classroom engagement depends on the professor. If she is engaging, a student will feel motivated*." Another student further explained, "*You attract students' attention all the time. This way, students stay engaged all the time and never feel bored*". This seems to be in line with Tomlinson & Moon (2013) who contend that "Engagement in the classroom results when a student's attention is attracted to an idea or a task and is held there because the idea or task seems worthwhile (p.7).

Theme 3: Students' perspectives on engagement in online classrooms versus in face-to-face classrooms. This theme reflects students' views of whether an online classroom is as engaging as a physical classroom. Although all students reported they were highly engaged in the class, only three students believed it would have been more engaging, if the course was offered in the traditional classroom. To this extent, one student believed that being at home is not as conducive to study as the traditional classroom. He explained:

> *The whole atmosphere of the course was great, whether it was traditional or online. Yet, for me personally I sometimes felt distracted. However, the course was completely enjoyable. The online environment is not suitable for learning and studying. Sometimes, you get distracted by others who are surrounding you, yet, you try to stay focused in class. [student]*

According to students, examples of distractions included another person coming into the room or a phone ringing. Class interaction was raised as an issue by the female group. One said, *"In online classrooms, there is not much interaction. It's just the teacher speaking most of the time, and we just listen.".*

However, not all the participants held the same stance. The majority of students disagreed with the notion of how the traditional classroom would be more engaging for learning. For many, the convenience and flexibility offered online were the most virtues mentioned. One participant noted, *"I honestly was happy that this course was online. The instructor can stay connected with the students more. Students can easily interact and engage more in the class."*

The female student group held a similar opinion on the effectiveness of online learning in terms of engagement. One student said, "Due to being online, we concentrated more". Furthermore, the female focus group students argued that for the introverted type of students, online courses are more advantageous, since they would be more beneficial and more engaging. One explained, *"I favor online, because some students are shy to participate face-to-face, while online, everyone participates equally."* Another agreed and stated, *"Yes, true! A shy student would not feel nervous when speaking or asking.". Another supported, "I myself cannot participate similarly in the physical classroom."*

Theme 4: Engagement throughout the semester. This theme assesses how well students were engaged at the beginning of the term versus the middle or end. Almost all students in both focus groups believed their engagement was not as high when the term commenced, as towards the middle or the end of the term. Eleven students reported that they got connected and engaged more as the semester progressed. Examples included the following:

- *"At the beginning, a student may feel introverted, but, as time passes by, one will start to get engaged more as he becomes familiar with the rules, other students, and interaction. So, engagement grows this way.".*

- *"With time, there became more engagement in the course. We broke the ice!"*

- *"At the very beginning, there was not a full classroom engagement, but we engaged more as time went on".*

Theme 5: Reflective Moments when students felt fully engaged. This theme relates to students' perceptions of aspects which they found more engaging in the online classroom. As the semester progressed, students encountered variable levels of engagement. Students rated course requirements based on their usefulness in engaging students more. In other words, interviewees also reported on the times when they were highly engaged during the online class. Those times were as follows (listed from the most engaging to the least):

- *Discussion board (DB) questions during class time.*
- *Class participation with the instructor.*
- *Group work activity.*

Although all course requirements were perceived to be engaging from the point of view of students, student-controlled active learning activities, in particular, were highly appreciated. Participants believed that interaction and discussions undertaken as part of discussion board activities resulted in improved engagement. For example, one student asserted, *"We were engaged more during discussion board questions. Because we were allowed enough time to think and answer questions assigned.".* This highlights the connection between time allowed to do an activity and student cognitive engagement. Another student emphasized*:*

> *The fun part of discussion board activity is that when you allocate time for the activity, some students post their answers to the forum. This allowed us to brainstorm and check other students' answers. Then, you share and discuss answers along with hints. This motivated me to stay connected more."*

This attitude was also reinforced by the female focus group. One of them added, *"In other courses, DB activities are assigned like homework to be done later, not during class time. Your way was better because once finished, you go through students' answers and give feedback.".* Additionaly, information and activities that were connected to students' lives were rated as highly engaging. Female students, especially, were in favor of this. Giving an example, one student noted, *"When lessons contain information that can be applied to our personal lives, we stay focused and engaged more. For example, in one lesson, there was information on interviews; this was valuable and useful.".* Another one added, *"When lessons touch personal needs, we get more excited about the lesson and stay focused."*

In students' narratives, the role of their peers as part of their learning engagement in the course was not valued so much. Students considered working in peers or groups with other classmates to be the least contribution to engagement, making it less "effective" in terms of engaging them online. Insight was also gained into some of the challenges associated with group work. Some students offered explanations to account for their perception. One student from the male focus group said, *"I think discussion board was the most engaging, then class participation and the least was working in groups."* This was explained, *"When in groups, I noticed, students remain silent waiting for someone to start and take the lead. When there is no student to take the leader role, group discussion becomes less active."* Another student supported, *"One student should proactively initiate the discussion then other ones may start to engage."*.

Theme 6: Influence of location/place of class attendance on student engagement.

Similar to the results of the SRS data, in-home engagement versus out-home engagement appeared to affect engagement. There seemed to be a consensus among both focus groups on the influence of student presence/ location during the time of attending virtual classes on student engagement. The following two extracts described their opinions towards how their engagement was affected according to where they were during the class time.

> *Home is usually calmer, but if I attend the class from a public place, it would be hard to pay attention.*
>
> *In a coffee shop, I cannot stay focused 100%. So, yes, it differs based on my location.*

Even if the setting is home, students agreed that it should be a quiet place without distractions. One student stated, *"The more people exist around you, the less engaged you become. So, as possible, I should isolate myself"*. Therefore, it is suggested to *"Try to stay away from distractions."* This required some students to create a study environment. A male student described, *"I tried to create an environment similar to the university classroom."* Another from the female group supported, *"My room has become a study-specific environment. Full engagement requires certain characteristics of the environment."*

DISCUSSION

Keeping students involved, engaged, and actively learning in online learning environments has challenged educators all over the world. In this study, there are several key findings based on the data collected. First, in regard to RQs 1 and 2, SRS data reported that students showed a general high level of engagement. Student behavioral engagement was rated the highest across all four weeks, followed by emotional engagement, and then social engagement. SRS analysis also showed that students were cognitively the least engaged during the eight virtual classes. In particular, students showed more meaningful engagement in week 1. Their cognitive engagement was ranked the highest in week

1 and the lowest in the other three weeks of investigation, during which the behavioral engagement type was highest, followed by emotional and then social engagements.

Despite the variability in students' perceptions raised in focus group data, the majority appear to have had positive engagement in virtual classrooms with increased engagement linked towards the middle and the end of the term. Student academic performance data illustrates a high level of achievements among students in all course major requirements, with female students outperforming male students. This means that students were cognitively engaged, otherwise, they would have not been able to score high in course tests, exams and written assignments. These findings are in agreement with those of Oraif & Elyas (2021), which have shown a high level of engagement among students in the online classroom.

Furthermore, general engagement did not vary between male and female students during all four weeks of the online classroom which is in line with the aforementioned studies in the literature (Almusharraf & Almusharraf, 2021; Devrim & Irem, 2020; Benhadj, 2021). However, female students were behaviorally more engaged in week 1, whereas male students were cognitively and socially more engaged during the same week. In weeks 2 and 3, female students showed more emotional engagement and more cognitive engagement. In week 4, both groups of students showed similar learning engagement in the virtual classes. The dynamicity of engagement was shown in the dimensions of cognition and social interactions in which they were found to grow over the course period, but behaviors and emotions were less likely to grow.

The qualitative data-based findings of this study indicated that successful student engagement in the virtual environment was influenced by a number of factors. The first was related to the engaging nature of the course and its interactive nature, how well it is designed and prepared, and how well it is facilitated with a motivator instructor. A key influence in online student engagement was found to come from how well prepared the course needed to to be, in order to engage students, and that in-class group work does not guarantee enrolment and participation. Tomlinson & Moon (2013) asserted, "Engagement in the classroom results when a student's attention is attracted to an idea or a task and is held there because the idea or task seems worthwhile." (p.7). Engagement is also a result of environmental facilitators such as classroom interpersonal relations and instructional quality, as well as personal factors such as motivation and aptitude" (Oga-Baldwin, 2019, p. 5).

Likewise, the job of the instructor in encouraging student engagement is a key factor. Influenced by self-determination theory and instructor's relatedness – one of three basic psychological needs – instructors "need to put enthusiasm into lessons, show an open, honest, and caring attitude toward students, and encourage students to support each other" (Bao et al., 2021, p. 3). Jang et al. (2010) rightly pointed out, when students do engage in classroom learning, there is always some aspect of the instructor's behavior that plays a role in the initiation and regulation of student engagement. Consequently, the role of online instructors needs to change from being authoritative to being cooperative and engaging (Oraif & Elyas, 2021). Lack of teacher enthusiasm, as perceived by learners, was cited as the reason for social-behavioral learning disengagement in Chinese EFL classrooms (Dewaele & Li, 2021).

Learner–learner interaction is essential for the engagement of students in the online learning environment, as stressed by Bolliger and Martin (2018). However, interview and focus group transcripts have shown otherwise. Students did not seem to have engaged meaningfully online when they were assigned into groups. It could be that the strategies used in this course were not implemented effectively, in order to initiate and support learner-to-learner interaction. Although, in this course, several activities in establishing a supportive online environment for students were applied, as suggested by Ryle & Cumming (2007). Activities included doing icebreakers, posting welcome messages and regular informative announcements to establish expectations, posting discussions in advance to stimulate interest, and providing necessary resources. In fact, students were required to share and reflect on their own learning styles and those of other, as well as, preferences, and learning difficulties. Still, it could be that students did not feel they had an established sense of belonging to the class community.

The location when attending the online session and students' choice of device (e.g., laptop, tablet, cellphone) used to access the virtual session are two indicators which affect student engagement in virtual learning. With regard to the first, students who choose to log on to the online session while at home are more likely to engage well in the online course, especially in terms of behavioral and emotional engagements. The findings of Al Shammari's (2021) affirmed that although students may have used smartphones in their remote learning, they did not recommend them. They were rated the least preferred compared to laptops, tablets, and desktop computers. Furthermore, students who select laptops in accessing virtual sessions are more likely to show higher engagement than those with hand-held devices like tablets or smart phones, especially in terms of behavioral, emotional, and social engagement aspects.

The notion of location of engagement (i.e., home or out-of-home in places such as the car, coffee shops, malls... etc.) as another dimension in nurturing student engagement during virtual language classrooms was also emphasized by the thematic analysis of interviews. Public places were reported to be not ideal for a full and meaningful engagement in online classrooms. An environment suitable for meaningful

engagement is associated with calmness and free of distractions.

Limitations and Future Directions

Despite the strengths of this study, two limitations need to be acknowledged. First, the data in this study was based on a four-week period. In order to study the longer-term effects of the online mode on language learning, it may be necessary to consider a longer period. Second, data was obtained from one course, and generalizations could not be made. A variety of courses in the exploration of student engagement in the virtual space could help gain a deeper understanding of the impact and reception of online education and learning on student engagement. A factor to be explored in future research, could be the role of academic majors and their impact on language learning in student classroom engagement.

CONCLUSION

This research revisited the topical issue among language instructors that within the virtual educational environment student engagement is impacted. Both the technology-mediated learning/teaching and the lack of face-to-face- contact may have attributed to students exhibiting different manifestations of engagement in the online classroom. This study attempted to investigate the engagement level among students in an online intermediate-level writing class taking place in the Saudi EFL context. Its aim was to contribute to understanding the factors that could have an impact on engagement in online learning. It focuses specifically on the English language learning classroom.

The findings and analysis demonstrated several points. There are variances in the types of engagement (e.g., cognitive, behavioral, social, and emotional), and there are also aspects that may influence engagement in online classrooms as a whole, such as learner-to-learner interactions, the influence of the instructor, the location of attending online classes, and the teaching materials amongst others. All these factors affect student engagement, which is a vital part of the learning process, but do not only depend on student behaviors. Instructors also have an important role in encouraging students and implementing more engaging activities in virtual classrooms to make the language learning experience more interactive and lively. Students generally had favorable attitudes towards online learning, during the period of distance study. Lastly, student engagement during the virtual learning experience was not affected especially in terms of behavior and emotion engagement dimensions.

Notably, online learning was able to achieve similar student engagement and learning outcomes as face-to-face learning. In today's fast-changing world, online learning and teaching has become more prevalent. Essential components that guarantee successful learning experiences need to be fulfilled, in order to obviate any sacrifice in the quality of education. Students should not be skeptical about distance study and seek to produce equivalent or better performance.

DECLARATION

The author received no financial support for the research.

ACKNOWLEDGEMENT

The author would like to thank the participating students of this research, in particular for their honesty and willingness to participate and wholehearted engagement.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCE

Al Shammari, M. H. (2021). Devices and platforms used in emergency remote learning and teaching during Covid-19: A case of English major students in Saudi Arabia. *Arab World English Journal (AWEJ). Special Issue on Covid 19 Challenges,* (1), 80- 94. https://dx.doi.org/10.24093/awej/covid.6

Al-Bogami, B., & Elyas, T. (2020). Promoting middle school students' engagement through incorporating iPad apps in EFL/ESL classes. *SAGE Open*, *10*(2), 2158244020926570.

Almossa, S.Y. (2021). University students' perspectives toward learning and assessment during COVID-19. *Education and Information Technologies, 26*, 7163–7181. https://doi.org/10.1007/s10639-021-10554-8

Almusharraf, A., & Almusharraf, N. (2021). Socio-interactive practices and personality within an EFL online learning environments. *Education and Information Technologies, 26*, 3947–3966. https://doi.org/10.1007/s10639-021-10449-8

Alzahrani, F., Zamakshari, N., & Elshemy, R. (2022). Did they manage to meet students' needs? English language instructors' experiences of remote instruction differentiation. *World Journal of English Language*, *12*(7), 1- 28. 10.5430/wjel.v12n7p28

Bao, C., Zhang, L. J., & Dixon, H. R. (2021). Teacher engagement in language teaching: Investigating self-efficacy for teaching based on the project "Sino-Greece Online Chinese Language Classroom". *Frontiers in Psychology*, *12*, 3000. https://doi.org/10.3389/fpsyg.2021.710736

Benhadj, Y. (2021). A Quasi-experimental study on the impact of blended learning on EFL students' language proficiency. *International Journal of Language and Literary Studies*, *3*(3), 146–155. https://doi.org/10.36892/ijlls.v2i3.668

Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System*, *55*, 145-157. https://doi.org/10.1016/j.system.2015.10.006

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

Breen, R. L. (2006). A practical guide to focus-group research. *Journal of Geography in Higher Education, 30*(3), 463-475. https://doi.org/10.1080/03098260600927575

Bolliger, D. U., & Martin, F. (2018). Instructor and student perceptions of online student engagement strategies. *Distance Education*, *39*(4), 568-583. https://doi.org/10.1080/01587919.2018.1520041

Castelli, F. R., & Sarvary, M. A. (2021). Why students do not turn on their video cameras during online classes and an equitable and inclusive plan to encourage them to do so. *Ecology and Evolution*, *11*(8), 3565-3576. https://doi.org/10.1002/ece3.7123

Cohen, L., Manion, L., & Morison, K. (2007). *Research methods in education* (6th ed.). Routledge.

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of research on student engagement*. Springer Science & Business Media.

Devrim, H., & Irem, A. (2020). Is technology in our classrooms? EFL teachers' beliefs and engagement with technology in the classroom. *Journal of Educational Issues, 6*(2), 38-58. https://doi.org/10.5296/jei.v6i2.17326

Dewaele, J. M., & Li, C. (2021). Teacher enthusiasm and students' social-behavioral learning engagement: The mediating role of student enjoyment and boredom in Chinese EFL classes. *Language Teaching Research*, 13621688211014538. https://doi.org/10.1177/13621688211014538

Fuller, K. A., Karunaratne, N. S., Naidu, S., Exintaris, B., Short, J. L., Wolcott, M. D., & White, P. J. (2018). Development of a self-report instrument for measuring in-class student engagement reveals that pretending to engage is a significant unrecognized problem. *PloS One*, *13*(10), https://doi.org/10.1371/journal.pone.0205828

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109. https://doi.org/10.3102/00346543074001059

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Harper, S. R., & Quaye, S. J. (2009). Beyond sameness, with engagement and outcomes for all. In S. R. Harper & S. J. Quaye (Eds.), *Student engagement in higher education* (pp. 1–15). Routledge.

Hipkins, R. (2012). The engaging nature of teaching for competency development. In *Handbook of research on student engagement* (pp.441-456). Springer.

Hiver, P., Al-Hoorie, A. H., & Mercer, S. (Eds.). (2020). *Student engagement in the language classroom*. Multilingual Matters.

Khlaif, Z. N., Salha, S., & Kouraichi, B. (2021). Emergency remote learning during COVID-19 crisis: Students' engagement. *Education and Information Technologies*, *26*, 7033–7055. https://doi.org/10.1007/s10639-021-10566-4

Kim, P. W. (2021). Assessing engagement levels in a non-face-to-face learning environment through facial expression analysis. *Concurrency and Computation Practice and Experience*, *33*, e6182. https://doi-org.sdl.idm.oclc.org/10.1002/cpe.6182

Lynn, P. (2009). Methods for Longitudinal Surveys. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 1-19). John Wiley & Sons. DOI:10.1002/9780470743874

Martin, J. (2019). Building relationships and increasing engagement in the virtual classroom. Practical tools for the online instructor. *Journal of Educators Online, 16*(1). https://doi.org/10.9743/JEO.2019.16.1.9

Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms.* Cambridge University Press.

Mercer, S., Talbot, K. & Wang, I. (2021). 8 fake or real engagement – Looks can be deceiving. In P. Hiver, A. Al-Hoorie, & S. Mercer (Eds.), *student engagement in the language classroom* (pp. 143-162). Multilingual Matters. https://doi.org/10.21832/9781788923613-011

Oga-Baldwin, W. Q. (2019). Acting, thinking, feeling, making, collaborating: The engagement process in foreign language learning. *System*, *86*, 102128. https://doi.org/10.1016/j.system.2019.102128

Oga-Baldwin, W. & Fryer, L. (2021). Engagement Growth in Language Learning Classrooms: A Latent Growth Analysis of Engagement in Japanese Elementary Schools. In P. Hiver, A. Al-Hoorie, & S. Mercer (Ed.), *Student Engagement in the Language Classroom* (pp. 224-240). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781788923613-015

Oraif, I., & Elyas, T. (2021). The impact of COVID-19 on learning: Investigating EFL learners' engagement in online courses in Saudi Arabia. *Education Sciences*, *11*(3), 99. https://doi.org/10.3390/educsci11030099

Philip, J., & Duchesne, S., (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics, 36*, 50-72. https://doi.org/10.1017/S0267190515000094

Quaye, S. J., Harper, S. R., & Pendakur, S. L. (Eds.). (2019). *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations*. Routledge.

Reeve, J. (2012). A self-determination theory perspective on student engagement. In Christinson,S., Reschly, A., & Wylie, C. (2012). *Handbook of research on student engagement* (pp. 149-172). Springer.

Reschly, A. L., & Christenson. S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), Handbook of research on student engagement (pp. 3–19). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-2018-7_1

Ryle, A., & Cumming, K. (2007). Reflections on engagement in online learning communities. *International Journal of Pedagogies and Learning*, *3*(3), 35-46. https://doi.org/10.5172/ijpl.3.3.35

Skinner, E. A., Kindermann, T. A., Connell, J. P., & Wellborn, J. G. (2009). Engagement and disaffection as organizational constructs in the dynamics of motivational development. In K. R Wentzel & D. B Miele (Eds.), *Handbook of motivation at school* (pp. 223–245). Routledge.

Soffer, T., & Cohen, A. (2019). Students' engagement characteristics predict success and completion of online courses. *Journal of Computer Assisted Learning*, *35*(3), 378-389. https://doi.org/10.1111/jcal.12340

Svalberg, A. M. L. (2018). Researching language engagement; current trends and future directions. *Language Awareness, 27*(1-2), 21-39. https://doi.org/10.1080/09658416.2017.1406490

Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and student success in a differentiated classroom.* ASCD.

Yundayani, A., Abdullah, F., Tandiana, S. T., & Sutrisno, B. (2021). Students' cognitive engagement during emergency remote teaching: Evidence from the Indonesian EFL milieu. *Journal of Language and Linguistic Studies*, *17*(1), 17-33. https://doi.org/10.52462/jlls.2

Zhou, S., Hiver, P., & Al-Hoorie, A. H. (2021). Measuring L2 engagement: A review of issues and applications. In Ph. Hiver, A. H. Al-Hoorie, & S. Mercer (Ed.), *Student Engagement in the Language Classroom* (pp. 75-98). Multilingual Matters.

https://doi.org/10.17323/jle.2023.14069

Translanguaging Instruction and Reading Comprehension Skills of Japanese EFL Learners: A Quasi-Experimental Study

Alexis Goli

Sophia University Junior College Division, Kanagawa, Japan

ABSTRACT

Background. The adoption of the TOEIC Listening and Reading test as the main English competency measurement instrument for Japanese businesses has lead Japanese business people to invest in courses specialized in test-taking strategies which seem to improve test scores in the short term. Nevertheless, there is pressing need to adopt more reliable instructions for reading ability development. Translanguaging pedagogy, an instruction method that urges L2 learners to make use of all languages in their linguistic repertoire, has rapidly gained the interest of language researchers and educators worldwide. Various studies have been conducted at all level of formal education, from elementary to tertiary education, to evaluate how it could help learners develop their proficiency in the target L2, including reading comprehension ability. However, no study on translanguaging in continuing education in Japan could be found in the literature.

Purpose. This paper presents an investigation to assess the effectiveness of translanguaging pedagogy in nurturing the reading comprehension of a group of Japanese EFL learners in a continuing education context.

Method. The study adapted a quasi-experimental design with a control and an experimental group, as well as a reading comprehension improvement intervention course between pre-test and post-test. The experimental group received instruction based on translanguaging pedagogy, and the control group was restricted to using only English in their classes.

Results. It was found that both the control and experimental groups improved their reading comprehension, but improvement in the experimental group was moderately more substantial. This finding contributes to the literature on translanguaging pedagogy in Japan, especially in the context of continuing education.

Conclusion. Due to the small number of participants, the findings of this study cannot be generalized to EFL education in continuing education. Further research with a substantial number of participants and treatment over a longer period could help confirm that translanguaging pedagogy can effectively be implemented in this setting to assist learners become proficient in the target L2.

KEYWORDS

translanguaging pedagogy, reading comprehension instruction, English as a foreign language (EFL), bilingualism, multilingualism, continuing education

Citation: Goli A. (2023). Translanguaging instruction and reading comprehension skills of Japanese EFL learners: A quasi-experimental study. *Journal of Language and Education, 9*(1), 59-75. https://doi.org/10.17323/jle.2023.14069

Correspondence:
Alexis Goli,
golikid@gmail.com

Received: March 23, 2022
Accepted: March 15, 2023
Published: March 31, 2023

INTRODUCTION

Background

English Education and English Use in the Japanese Context

The end of the twentieth century witnessed technological advances and the globalization of the world economy, resulting in the emergence of a multicultural and multilingual global community where people from all corners of the earth are permanently in contact, whether in person or remotely. As a result, in order to communicate effectively in the global community, most people aspire to gain competency in two or more languages, including English, the de facto lingua franca of the new world commu-

nity. Japan, an insular country that prides itself as being a homogeneous, monocultural, and monolingual society (Carroll, 2013) puts pressure on its students and workforce to acquire reliable English abilities. In this way it can contribute more efficiently to the country's economy which has yet to recover from the economic burst of the early 1990s.

In response to the request of the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) to enhance their EFL programs and foster bilingual business professionals, Japanese education institutions have adopted the Test of English for International Communication (TOEIC) Listening and Reading as the countries' main English competency measurement instrument (Takahashi, 2012). Universities rely on TOEIC test performances for course placement, and run TOEIC courses to ensure students graduate university with business English proficiency (ibid). Moreover, businesses use the TOEIC score to make employee job placement and career advancement decisions (Tsedal, 2019). As a result, students in their last year of university and people in the workforce make substantial financial investment in courses specialized in test-taking strategies. Except for test-taking strategies, such courses tend to use the same vocabulary and reading instructions which most Japanese EFL learners have been exposed to throughout formal education. The courses use rote learning and memorization to get learners to internalize long lists of vocabulary, and have them practice reading comprehension through grammar-translation practices, as in Japanese junior and senior high schools (Egitim, 2020; Kumagai, 1994; Steele & Zhang, 2017). Consequently, these courses seem to help improve test scores in the short term, but the issue of English competency remains prevalent.

In their continued effort to adapt EFL education to its long-term educational objectives, in 2020 the MEXT introduced foreign language policy reforms which put the nurturing of oral communicative competence at the center of its EFL education. The MEXT has imposed English as the sole official language of instruction in EFL classes in formal education (Turnbull, 2018). However, in a survey which investigated Japanese EFL learners' self-perceived oral communication ability improvement following a one-on-one tutoring course, it was found that the need for Japanese to improve their reading comprehension skills seems comparatively more pressing than expanding their speaking skills. Currently, "most respondents read work-related English documents and correspondence more often than they communicate orally in English" (Goli, 2021). Additionally, as a native language (L1), the Japanese language is not only the preferred language of instruction in foreign language education (Turnbull, 2018), but also the literature abounds with research findings which sustain that the use of L1 in foreign language classrooms plays a key role in an effective acquisition (Baker, 2011; Cenoz & Gorter, 2011, 2022; Garcia & Li Wei, 2014; Lewis et al., 2012; Spinelli, 2017; Williams, 2012). In the light of the reasons aforementioned, it is clearly necessary to adopt reading ability development instructions in EFL which are more suitable to Japanese learners and their context. Consequently, this study was conducted to investigate translanguaging pedagogical approach as an effective reading instruction method to develop the reading skills of Japanese EFL learners. The following research question was formulated: To what extent can translanguaging pedagogical instruction improve the reading comprehension ability of Japanese EFL learners in a continuing educational context?

LITERATURE REVIEW

With globalization and advancements in communication technologies, most people around the world need to be able to read at high levels of proficiency in languages other than their L1, in order to achieve their personal and professional goals. However, developing strong reading abilities, especially in a L2, is both time-consuming and challenging (Grabe & Stoller, 2011). Pressley (2000) demonstrated that unless teachers overtly teach comprehension skills, students cannot easily acquire the ability to understand texts. King (2007) supports the notion that in order for readers to make meaning and comprehend texts, it is crucial for them to also acquire the skill to infer main ideas from a text. According to Grabe & Stoller (2011), effective reading instruction in L2 needs to prioritize guiding learners to grasp the main idea of texts through class discussions where participants learn to make connections between the text and prior knowledge. Hence, learners should work in groups and explain the main ideas to each other. Similarly, August & Shanahan (2006) support the notion that learners increase their literacy comprehension when they work collaboratively in groups. Grabe & Stoller (2011), however, insist that making learners use all the languages in their linguistic repertoire in group reading activities is extremely important. The instruction approach suggested by Grabe & Stoller (2011) and August & Shanahan (2006) have common characteristics with translanguaging pedagogy, a teaching and learning method in language and multilingual education. In translanguaging instruction, the teacher purposefully specifies the language(s) of input and output, and guides learners to use all linguistic and semiotic resources in their possession to optimize learning of the subject matter (Baker, 2011; Cenoz & Gorter, 2011, 2022; Garcia & Li Wei, 2014; Lewis et al., 2012; Williams, 2012). It combines several activities which stimulate learners to make use of all skills in the four language areas to communicate and make meaning. For example, in a language class, the teacher can have learners read a text in one language, investigate the topic and participate in group discussions, then make a summary and report their findings in the target language (Nagy, 2018; Cenoz & Gorter, 2011, 2022). Baker (2011) and Lewis et al. (2012) support the notion that learners can enhance literacy and oral competency in their weaker language when they simultaneously use those languages. Garcia & Li Wei (2014) argue that learners' engagement in translanguaging prompts interlanguage exchanges which

expand on their pre-existing linguistic and cognitive knowledge to instigate further learning and new knowledge. Focusing on the pedagogical aspect of translanguaging, Cenoz & Gorter (2017, 2020) put forward the concept of 'pedagogical translanguaging' which, they argue, goes beyond the original approach of translanguaging by including practices related to the development of metalinguistic awareness. They also explain that pedagogical theory and practice aims at developing multilingualism in two or more languages in content and language classes.

Translanguaging for Reading Skill Development in Tertiary Education in a Global Context

Various studies conducted in tertiary education suggest that translanguaging pedagogy helps learners more effectively develop proficiency in the target L2, in comparison to pedagogical instruction that restricts exchanges among class participants to the L2 (Bartlett, 2018; Hungwe, 2019; Makalela, 2015; Spinelli, 2017). Makalela (2015) researched the vocabulary gains and improvements in oral reading proficiency in Sepedi among 60 multilingual pre-service teachers who used Sepedi as an additional language. While the experimental group received instructions following translanguaging pedagogy, language use in the control group was restricted to the target language. Makalela (2015) concluded that the gain in vocabulary of the experimental group was significantly larger than the control group. However, with regard to oral reading competency, gains were similar for both groups. The author concluded that the vocabulary of adult bilingual language learners could more effectively be enhanced by means of translanguaging instructions, but improvement in oral reading by means of translanguaging instructions were inconclusive. Makalela (2015) established his study on the theory that word recognition skill is crucial for vocabulary and reading comprehension proficiency. In light of the findings of Spinelli (2017), Kor et al. (2014), and Grabe & Stoller (2011), Makalela (2015)'s conclusion on this point of investigation appears plausible. With regard to oral reading proficiency, Makalela (2015) does not provide or elaborate on any explicit theoretical or empirical evidence to support his assumption that word recognition and vocabulary skills affect oral reading competency. However, it is argued that oral reading proficiency is affected by automatic information processing or automaticity (LaBerge & Samuels, 1974; Logan, 1997). Non-fluent language learners tend to struggle with oral reading comprehension, since they have not yet acquired automaticity. Grabe & Stoller (2011) support the notion that non-fluent readers need thousands of hours of reading practice, in order to acquire word recognition automaticity. Consequently, the outcome of the study on oral reading proficiency appears reasonable.

Some researchers have explored the impact of translanguaging and paraphrasing on the development of reading comprehension. Hungwe (2019) investigated the impact of translanguaging and paraphrasing on the development of reading comprehension skill among 36 multilingual medical school students in a course on English for academic purposes. An analysis of the quality of summaries by the participants of a scientific article discussed during intervention led Hungwe (2019) to the conclusion that translanguaging practice used in combination with paraphrasing enhances the reading comprehension of English learners in comparison to conventional instruction. Hungwe (2019) based her argument on translanguaging theory literature and paraphrasing. Regarding translanguaging theory, she supports the idea that the use of L1 through translanguaging not only helps to develop the target language (Lewis et al., 2012), but also helps students work at a higher cognitive level than they would if restricted to the exclusive use of the target language (Storch & Wigglesworth, 2003). Furthermore, the learners are able to understand the meaning of new and difficult words, as well as explain complex syntactic rules (Hussein, 2013). According to the literature, paraphrasing is an excellent tool for reinforcing reading skills, such as identifying main ideas and finding supporting details (Fisk and Hurst, 2003). It has also been shown that by paraphrasing, students grasp and express the original ideas of the author with their own words, thus showing they understand a text (Hirvela & Du, 2013).

Another researcher who investigated translanguaging and other multilingual pedagogies, is Spinelli (2017). Spinelli (2017) researched the impact of translanguaging and other multilingual pedagogies, namely cross-linguistic comparison and inter-comprehension, and L3/Ln learners' reading and writing abilities in a multilingual setting at a university in the United States. Spinelli found that multilingual pedagogy facilitated positive interlingual transfers, leading to the development of both reading and writing skills, with relatively more gain in writing skills. Furthermore, in addition to the degree of proficiency of languages in the learners' linguistic repertoire, the improvement of writing skills seemed to have positively influenced their reading comprehension. Spinelli (2017) based her argument on theories and findings in reading comprehension research, according to which reading comprehension combines visual information processes (lower processes) with prior knowledge (higher lever process) applied by readers to their reading. She paid special attention to visual information processing, by focusing on word recognition and syntactic parsing processes, identified as two of the most important processes for reading comprehension (Adams, 1990; Perfetti, 1999). Spinelli supports the notion that while L2 reading is a dual-language process, L3/Ln reading is a multi-language involvement process (Cook, 1997; Koda, 2007). As a result, the richer the linguistic repertoire of language learners, the more their reading comprehension is expediated at the lower level of processing. Thus, parallel grammatical ordering in the languages in the multilingual speaker's repertoire, cognates across those languages, and similar syntactic information such as deter-

miners promote transfers, especially when reading instructions help learners gain awareness of aforementioned similarities and differences (Nagy et al., 1993; Cook & Bassetti, 2005; Koda, 2007; Grabe, 2009). In light of all the evidence provided by Spinelli (2017) in support of her argument, it appears reasonable that, with adequate reading comprehension training, multilingual learners can gain word recognition skills and more efficiently acquire high and reliable reading abilities for visual information processing.

Translanguaging for the Development of Reading Skills in Tertiary Education in the Japanese Context

In Japan translanguaging pedagogy is gaining attention in language education. Some researchers have started to experiment with regard to its suitability and effectivity in formal education. In his studies of reading comprehension of English, Bartlett (2018) investigated the effect of translanguaging instruction on reading comprehension in a Japanese tertiary educational setting with 107 second-year college students. Findings showed that the 56 participants in the experimental group demonstrated a higher level of retention and a more varied English language use in presentations, when compared to the 51 students in the control groups. Bartlett (2018) also found that the participants in the experimental group were more motivated to study English. Responding to recommendations from translanguaging literature encouraging educators to adopt alternative teaching approaches in the translanguaging classroom (Makalela, 2015), Bartlett (2018) placed multiple intelligences theories as the foundation of his study. He also investigated the effectiveness of applying these intelligences in translanguaging classrooms in promoting learners' learning and comprehension. It has been suggested that each learner possesses at least seven intelligences or independent ways of processing information, which would interact differently depending on the individual (Gardner & Hatch, 1989). Stimulation of the intelligences promotes a more personalized learning, and, in comparison with practices in conventional education, provides different learning opportunities to every learner (Csikszentmihalyi, 1990). Therefore, adopting activities designed to stimulate multiple intelligences along with the full linguistic repertoire of students in translanguaging classroom is consistent with the goals of integrating such activities in the teaching process, especially in the language classroom. This should provide each learner with a unique opportunity to reinforce their linguistic competence, including their reading comprehension, in accordance to the intelligences they possess.

All four studies introduced above concluded that translanguaging pedagogy contributes, to various extents, to EFL learners' reading comprehension improvement. However, although the studies targeted adult learners, they were all been performed in formal education settings, particularly in tertiary education. In light of the increasing need for Japanese business people and workers to improve their EFL reading comprehension business purposes, an evaluation of the effectiveness of translanguaging instruction in a continuing education setting is highly necessary.

METHOD

Design

This project adopted a quasi-experimental design, a pre-test—post-test control and experimental group design with a period of intervention between pre-test and post-test (Cohen et al., 2017, p.402)

Participants

21 participants were recruited by convenient sampling method. They all met the course participation criteria described as follows:

(1) Be 18 years old or older;

(2) Be a businessperson, a company employee, or self-employed person;

(3) Be aiming at improving one's reading comprehension skill;

(4) Have a TOEIC Listening and Reading score between 400 and 780.

All the participants m*et a*l the criteria. However, although all 21 participants sat for the pre-test and took part in the entire treatment course, 2 of them did not take the post-test and fill out the questionnaire about their personal details. As a result, 19 participants took part in the entirety of the study. Basic background information on participants is summarized in Table 1.

The Intervention

A reading comprehension instruction course was designed and administered by the researcher. The course was free of charge and was administered online via Zoom Meeting with a control group and an experiment group. The course consisted of 7 sessions of 120 minutes for each group held over the weekend. The reading tests were conducted in the first and 7th sessions, and the 5 sessions in-between were dedicated to reading comprehension skill building. Participants were assigned to the control and experimental groups randomly at registration online by means of the website registration system. In the reading comprehension instruction sessions with the control group, only the target language, English, was used for instruction and during discussion among the participants. The study materials made availa-

Table 1

Basic Background Information of Participants

Gender	Female			Male		
	15 (78.9 %)			4 (21.1%)		
Age	20 to 29	30 to 39	40 to 49	50 to 59	60 or more	
	0	2 (10.5 %)	13 (68.4%)	3 (15.8%)	1 (5.3 %)	
Education	Elementary school	Junior high school	Senior high school	Professional school	college	Graduate school
	0	0	0	4 (21.1%)	12 (63.2%)	3 (15.8%)
Years of school English education	1 to 3 years		4 to 6 years	7 to 9 years		10 years or more
	0		5 (26.3%)	11 (57.9%)		3(15.8%)
English Study in speaking countries	Within 1 month	Within 3 months	Within 6 months	Within 12 months	Within years	2 years or more
	9 (47.4%)	4 (21.1%)	0	4 (21.1%)	0	2 (10.5%)
Life in English speaking countries	Within 1 month	Within 3 months	Within 6 months	Within 12 months	Within years	2 years or more
	8 (42.1%)	1 (5.3%)	0	1 (5.3%)	2 (10.5%)	7 (36.8%)
Occupation	Company employee					8 (42.1%)
	an executive secretary, a web designer, a graphic designer, a nursery teacher, a university staff member, a Japanese language teacher, a painter, a nurse, a shop assistant, a housewife, and a self-employed person					11 (57.9) %

ble on the webpage were mainly in English, with a simple Japanese translation for relatively long and complex task instructions. In the reading comprehension classes with the experimental group, the researcher used both English and Japanese. The participants were encouraged to use both English and Japanese for in-class discussion, as well as during discussions among participants. The participants in the experimental group were also provided with a Japanese translation of all the English reading materials used in the class.

Data Collection

A pre-test and post-test approach was adapted to gather quantitative data. Thus, two sets of data were collected before and after the reading comprehension intervention. First, before the intervention, participants were asked to complete a 40-minute mockup TOEIC reading test. After attending a total of ten hours reading comprehension instruction classes, the participants took another 40-minute mockup reading test of the TOEIC. Both the pre-test and post-test were administered online. The TOEIC test was chosen for two reasons. First, it is a highly reliable test (ETS, 2007).[3] All questions in the TOEIC Reading test are in the form of multiple-choice. Secondly, the TOEIC reading test is highly reliable due to the nature of its test tasks which enable objective scoring, as opposed to task-based evaluations such as writing and speaking test.

The official TOEIC reading test has a total of 100 questions. The number of questions is different for each part. For the purpose of this study, short versions of the TOEIC reading test with 50 questions were used. Table 2 summarizes the abilities measured in the TOEIC Reading test (IIBC, n.d.) and the content of the official TOEIC reading test and the customized version for this study (pre-post tests).[4] The content of the pre-test and post-test were adapted from Educational Testing Service (2019) and Educational Testing Service (2020), respectively.[5] [6]

Data Analysis

Descriptive and inferential statistical (independent-samples and paired-sample t-tests) analyses were performed using IBM SPSS 27.0. Inferential statistical analyses were performed, in order to evaluate differences in performance

[3] ETS. (2007). *TOEIC Score User Guide-Listening & Reading*. https://www.ets.org/s/toeic/pdf/toeic-listening-reading-test-user-guide.pdf

[4] IIBC. (n.d.). Official score certificate format. https://www.iibc-global.org/english/toeic/test/lr/guide05/guide05_01.html

[5] Educational Testing Service. (2019). *Official TOEIC Listening & Reading preparation book 5*. Educational Testing Service.

[6] Educational Testing Service. (2020). *Official TOEIC Listening & Reading Preparation book 7*. Educational Testing Service.

Table 2

Content and Language Ability Measured in Each Part in the Reading Tests

Part (Test type)	Number of questions		Ability measured
	Official Test	Pre-post tests	
Part 5 (Incomplete passage)	30	15	• Ability to understand vocabulary in written texts. • Ability to understand grammar in written texts.
Part 6 (Incomplete text)	16	8	• Ability to understand vocabulary in written texts. • Ability to understand grammar in written texts. • Ability to make inferences based on information in written texts.
Part 7 (Single passage)	29	12	• Ability to understand vocabulary in written texts.
(Multiple passages)	25	15	• Ability to understand grammar in written texts. • Ability to make inferences based on information in written texts.

Table 3

Descriptive Analysis of Performance in Pre-test and Skewness and Kurtosis Statistics.

Control group									
		Scores			Skewness			Kurtosis	
Test rubrics	N	M	SD	Statistic	SE	Z	Statistic	SE	Z
Total score	10	25.40	9.755	0.817	0.867	0.942	-0.939	1.334	-0.704
Part 5 score	10	9.70	2.983	0.802	0.867	0.925	-0.620	1.334	-0.465
Part 6 score	10	5.40	1.430	-0.319	0.867	-0.368	-1.1663	1.334	-0.874
Part 7 score	10	10.30	7.134	1.367	0.867	1.577	0.483	1.334	0.362
Experimental group									
		Scores			Skewness			Kurtosis	
Test rubrics	N	M	SD	Statistic	SE	Z	Statistic	SE	Z
Total score	9	26.22	9.897	-0.396	0.717	-0.552	-0.687	1.400	-0.491
Part 5 score	9	9.44	2.698	0.373	0.717	0.520	0.433	1.400	0.309
Part 6 score	9	5.33	2.000	-0.469	0.717	-0.654	-0.844	1.400	-0.603
Part 7 score	9	11.44	5.855	-0.848	0.717	-1.183	0.464	1.400	0.331

Note. SE: Standard Error; Z: Z-score (Z=Statistic/SE)

between the two sets of tests and between the two groups (Dörnyei, 2007; Paltridge & Phakiti, 2015). The first, t-tests were performed on the participants' pre-test, in order to verify that there were no significant differences in English reading proficiency levels between the control and experimental groups before intervention. Next, t-tests of the participants' post-test were conducted to compare the performance of the control and experimental groups after intervention. Then, paired samples t-tests were conducted, in order to study the effect of the reading comprehension intervention on each group. All t-tests of the participants' performances included their performances in each part of the reading test, namely Parts 5, 6, and 7.

RESULTS

Descriptive Analysis with Skewness and Kurtosis Statistics

Descriptive statistics of participants' performance in the pre-test are shown in Table 3 with the results of skewness and kurtosis statistics. According to West *et al*. (1995) and Lowie & Seton (2012), if the absolute values of the skewness and kurtosis z-scores (Z) are both smaller than 1.96 for a sample size under 50, the sample is considered acceptably normally distributed. Values of z-sores for both the control

and experimental groups for the overall performance as well as performances in Part 5,6, and 7 are all less than 1.96. Consequently, it could be concluded that the samples used for this study was normally distributed.

Results of The Independent T-Test of The Pre-test

In order to compare the level of the two groups, independent-samples t-tests of their performances in the pre-test were conducted. Results of the total score and performance in each part of the reading test are shown in Table 4. They reveal that, out of a total of 50 possible score, both control (M = 25.40; SD = 9.755) and experimental (M = 26.22; SD = 9.897) groups achieved comparably similar mean scores values and standard deviations. The t-test pointed out that the difference between the two groups was not statistically significant (p > 0.05). Therefore, the control and experimental groups had similar language proficiency and were comparable.

For Part 5, out of a total 15 possible score, comparable means score and standard deviation values were obtained by the control (M = 9.70; SD = 2.983) and experimental (M = 9.44; SD = 2.698) groups. Similarly, a comparable mean score and standard deviation values were obtained by the control (M = 5.40; SD = 1.430) and experimental (M = 5.33; SD = 2.000) groups out of a total 8 possible score for Part 6. For Part 7, out a total 27 possible score, the control and experimental groups achieved similar means score values and standard deviations. The t-test for each part of the reading test indicates that the difference between the two groups were not statistically significant (p > 0.05). Consequently, the control and experimental groups had similar reading proficiency in each area of the test, and were therefore comparable.

Results of The Independent T-Test of The Post-Test

Independent-samples t-tests of participants' performances in the post-test were performed to evaluate the effect of the reading instruction on the two groups. Table 5 shows the t-test of the overall performance (total score) of the participants, as well as their performance in each part of the post-test. The mean difference of 4.389 points between the total score of the control (M = 26.50; SD = 10.277) and experimental (M = 30.89; SD = 9.532) groups was observed. Results of the t-test indicate that this difference in score between the two groups was not statistically significant (p > .05). However, the Eta squared value (the effect size describing the magnitude of the difference between two groups observed) suggests that the difference in score has a moderate effect, therefore, not negligible.

In Part 5, comparable means score and standard deviation values were obtained by the control (M = 10.30; SD = 2.584) and experimental (M = 9.22; SD = 2.819) groups, with a mean difference of 1.078 points. This difference was not statisti-

Table 4

Results of Independent-Samples T-Test of Participants' Pre-test Performance

	Pre-test total score					
	N	**M**	**SD**	*df*	*t*	**Eta squared**
Control group	10	25.40	9.755	9	.182	.002
Experimental group	9	26.22	9.897	8		
	Part 5					
	N	**M**	**SD**	*df*	*t*	**Eta squared**
Control group	10	9.70	2.983	9	-.195	.002
Experimental group	9	9.44	2.698	8		
	Part 6					
	N	**M**	**SD**	*df*	*t*	**Eta squared**
Control group	10	5.40	1.430	9	-.084	.000
Experimental group	9	5.33	2.000	8		
	Part 7					
	N	**M**	**SD**	*df*	*t*	**Eta squared**
Control group	10	10.30	7.134	9	.380	.008
Experimental group	9	10.44	5.855	8		

Note. p > .05

Table 5

Results of Independent-Samples T-Test of Participants' Post-test Performance

	Pre-test total score					
	N	**M**	**SD**	*df*	*t*	**Eta squared**
Control group	10	26.50	10.277	9	.962	.052
Experimental group	9	30.89	9.532	8		
	Part 5					
	N	**M**	**SD**	*df*	*t*	**Eta squared**
Control group	10	10.30	2.584	9	-.870	.043
Experimental group	9	9.22	2.819	8		
	Part 6					
	N	**M**	**SD**	*df*	*t*	**Eta squared**
Control group	10	5.30	2.214	9	-.578	.019
Experimental group	9	4.78	1.641	8		
	Part 7					
	N	**M**	**SD**	*df*	*t*	**Eta squared**
Control group	10	10.90	7.593	9	1.859	.169
Experimental group	9	16.89	6.294	8		

Note. $p > .05$

cally significant ($p > .05$), and the effect of the Eta squared is relatively small. As a result, there was no difference in the two groups' performance in Part 5 ($p > .05$). The control (M = 5.30; SD = 2.214) and experimental (M = 4.78; SD = 1.641) groups achieved comparable mean scores and standard deviation values in Part 6, with a mean difference of .522 points that was not statistically significant ($p > .05$), and has very small effect (Eta squared = 0.19). However, regarding the participants' performance in Part 7, there was a mean difference of 5.989 points between the control (M = 10.90; SD = 7.593) and experimental (M = 16.89; SD = 6.294) groups. This was evaluated as not statistically significant ($p > .05$), but with a large effect on the difference in performance between the two groups.

Results of The Paired T-Tests of the Pre-Test and Post-Test

Paired samples t-tests of the participants' pre-test and post-test performances were performed, in order to examine the effect of the reading comprehension intervention within the control and experimental groups. The test results of the control group are shown in Table 6, and they reveal that the control group achieved similar means score and standard deviation values in the pre-test (M = 25.40; SD = 9.755) and post-test (M = 26.22; SD = 10.277) with a mean paired difference of 1.100 points. However, the paired t-test shows statistically significant differences between the two tests with a substantial effect (p <.05; Eta squared = .417). This represents a 41.7 percent difference between the pre-test and post-test total scores in the control group.

In Part 5, comparable means score and standard deviation values were obtained in the pre-test (M = 9.70; SD = 2.983) and post-test (M = 10.30; SD = 2.584) in the control groups, with a mean paired difference of .60 point. This difference in score was not statistically significant ($p > .05$), and its effect is small (Eta squared = .045). Consequently, the score improvement of the control group in Part 5 alone was ignorable. Similar outcomes were obtained for Part 6 and 7. Both the mean paired differences of .10 point in Part 6 and .60 point in Part 7 were evaluated as not statistically significant ($p > .05$) with a considerably small effect (Eta squared = .005, Part 6; Eta squared = .049, Part 7). A grouped spaghetti plot of the performance of the control group is represented in Figure 1.

Table 7 displays the results of the paired samples t-test of the performance of the experimental group. It shows that there was an increase in the post-test (M = 30.89; SD = 9.532) in comparison to the pre-test (M = 26.22; SD = 9.897) with a mean paired difference of 4.667 points. The standard deviation values were comparable, and the paired t-test shows a statistically significant difference between the pre-test and post-test with a substantial effect (p <.05; Eta squared = .632). In other words, there was a 63.2 percent difference between the pre-test and post-test scores in the experimental group. A grouped spaghetti plot of the performance of the

Table 6

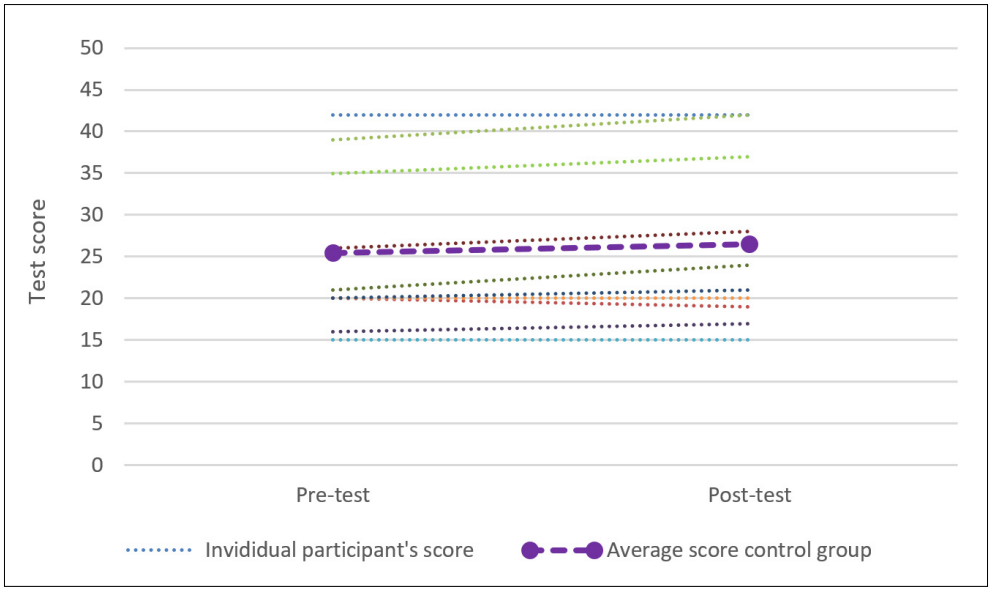
Results of Paired Samples T-Test of the Control Group on Pre-Test and Post-Test (N=10)

	Pre-test total score				
	M	**SD**	**df**	*t*	**Eta squared**
Pre-test	25.40	9.755	9	-2.538*	.417
Post-test	26.50	10.277			
	Part 5				
	M	**SD**	**df**	*t*	**Eta squared**
Pre-test	9.70	2.983	9	-.651	.045
Post-test	10.30	2.584			
	Part 6				
	M	**SD**	**df**	*t*	**Eta squared**
Pre-test	5.40	1.430	9	.218	.005
Post-test	5.30	2.214			
	Part 7				
	M	**SD**	**df**	*t*	**Eta squared**
Pre-test	10.30	7.134	9	-.678	.049
Post-test	10.90	7.593			

*Note. Total test score: *p < .05 (p = .032); p > .05*

Figure 1

Grouped Spaghetti Plot of the Performances of the Control Group



experimental group is represented in Figure 2. Figure 3 is a grouped spaghetti plot of the performance of participants of both the control and experimental groups.

In Part 5, similar means score and standard deviation values were obtained in the pre-test (*M* = 9.44; *SD* = 2.698) and post-test (*M* = 9.22; *SD* = 2.819) with a mean paired difference of .222 point. The paired samples t-tests results also indicate that the difference in score in Part 5 was not statistically significant ($p > .05$), and has small effect (Eta squared = .013). As a result, the decrease in score of the experimental group for Part 5 is considerably small.

Table 7

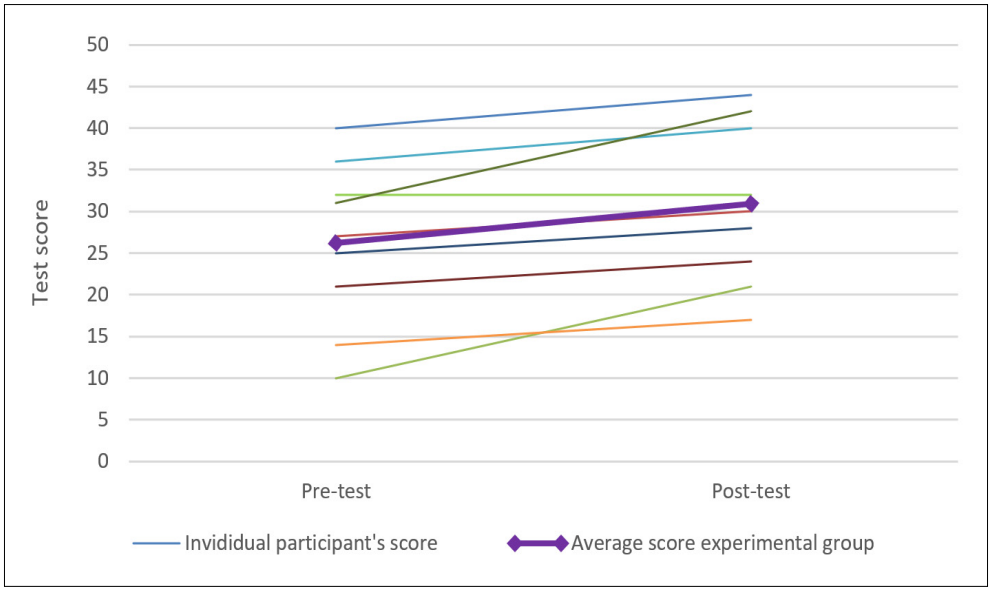
Results of Paired Samples T-Test of the Experimental Group on Pre-Test and Post-Test (N=9)

	Pre-test total score				
	M	**SD**	**df**	*t*	**Eta squared**
Pre-test	26.22	9.897	8	-3.709*	.632
Post-test	30.89	9.532			
	Part 5				
	M	**SD**	**df**	*t*	**Eta squared**
Pre-test	9.44	2.698	8	.326	.013
Post-test	9.22	2.819			
	Part 6				
	M	**SD**	**df**	*t*	**Eta squared**
Pre-test	5.33	2.000	8	1.104	.132
Post-test	4.78	1.641			
	Part 7				
	M	**SD**	**df**	*t*	**Eta squared**
Pre-test	11.44	5.855	8	-4.710**	.735
Post-test	16.89	6.294			

*Note. Total test score: ** p < .01 (p = .006); p > .05; ** p < .01 T*

Figure 2

Grouped Spaghetti Plot of the Performances of the Experimental Group



Regarding the performance of the experimental group in Part 6, comparable means score and standard deviation values were obtained in the pre-test (*M* = 5.33; *SD* = 2.0) and post-test (*M* = 4.78; *SD* = 1.641), with a mean paired difference of .503 point. The difference in the Part 6 score between the two tests was not statistically significant ($p > .05$), but it has a relatively large effect (Eta squared = .132). For Part 7, there was an increase in performance in the post-test (*M* = 16.89; *SD* = 6.294) in comparison to the pre-test (*M* = 11.44; *SD* = 5.855), with a mean paired difference of 5.444 points. The standard deviation values were comparable. The pairs t-test results show a statistically significant

difference between the pre-test and post-test with a large effect ($p < .05$; Eta squared = .735). In other words, there was a 73.5 percent difference between the pre-test and post-test in the experimental group in Part 7. Figures 4, 5, and 6 are grouped spaghetti plot of the performance of participants of both the control and experimental groups in Parts 5, 6, and 7, respectively.

Results of Multiple Linear Regression Test

A multiple regression test was performed, in order to examine the effect of the treatment groups and the English proficiency level of the participants on performance in the post-test. The multiple regression test was conducted with one dependent variable and three independent variables (predictors). The post-test score was set as dependent variable, and 'intervention group', 'proficiency level' of participants, and 'interaction' as independent variables. The 'intervention group' was a categorical variable consisting of the experimental and control groups. The experimental and control groups were divided into high and low proficiency levels which made up the "proficiency level" categorical variable (See Figure 7 for a grouped spaghetti plot of the performance of the low and high proficiency subgroups in the control and experimental groups). The third independent variable "interaction" was the product of 'intervention

Figure 3

Grouped Spaghetti Plot of the Performances of the Control and Experimental Groups

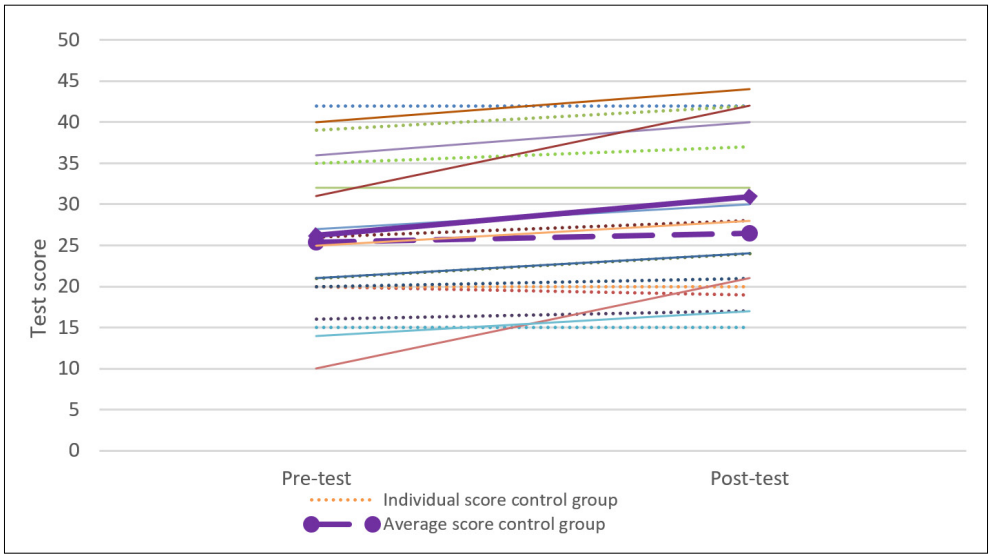


Figure 4

Grouped Spaghetti Plot of the Performances of the Control and Experimental Groups in Part 5

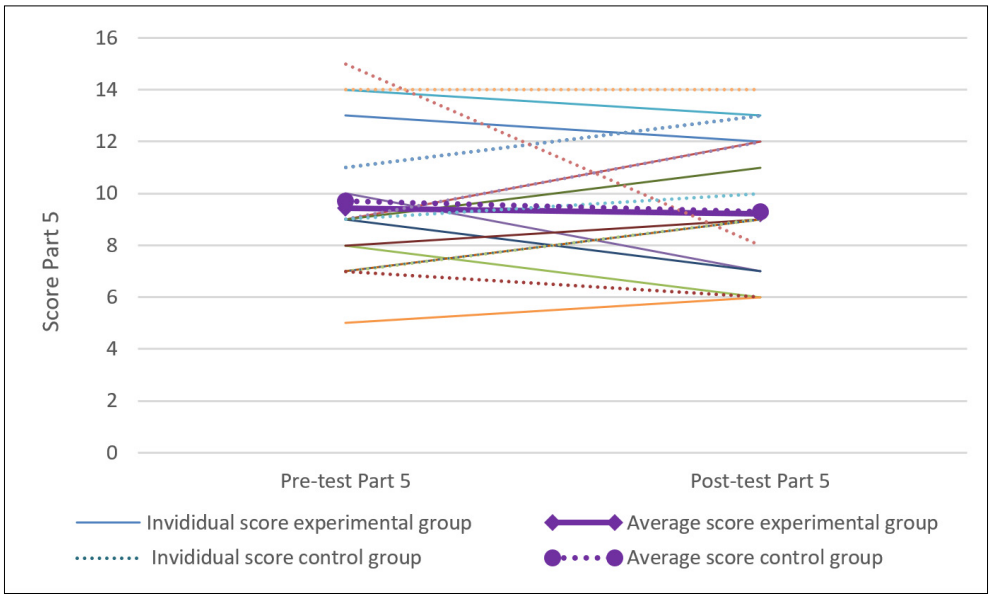


Figure 5

Grouped Spaghetti Plot of the Performances of the Control and Experimental Groups in Part 6

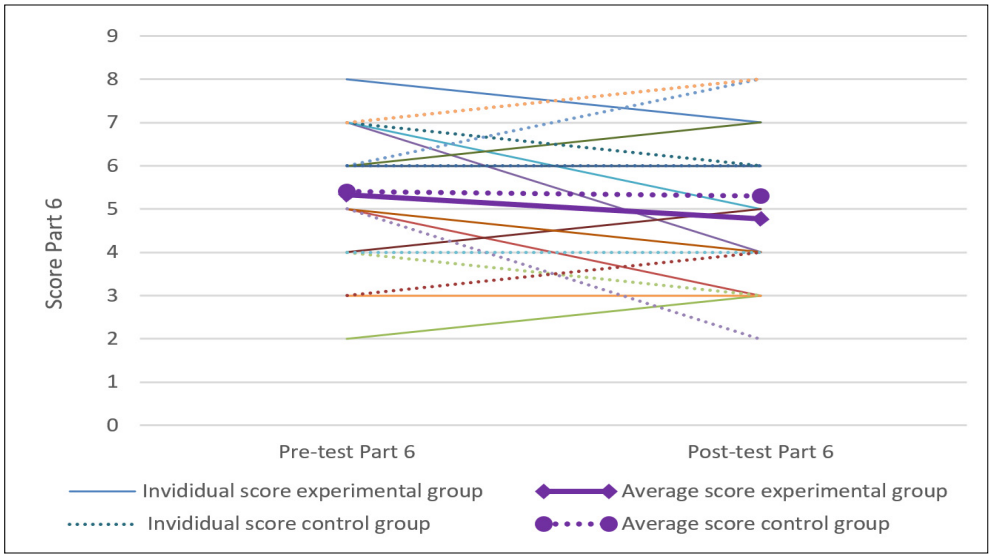
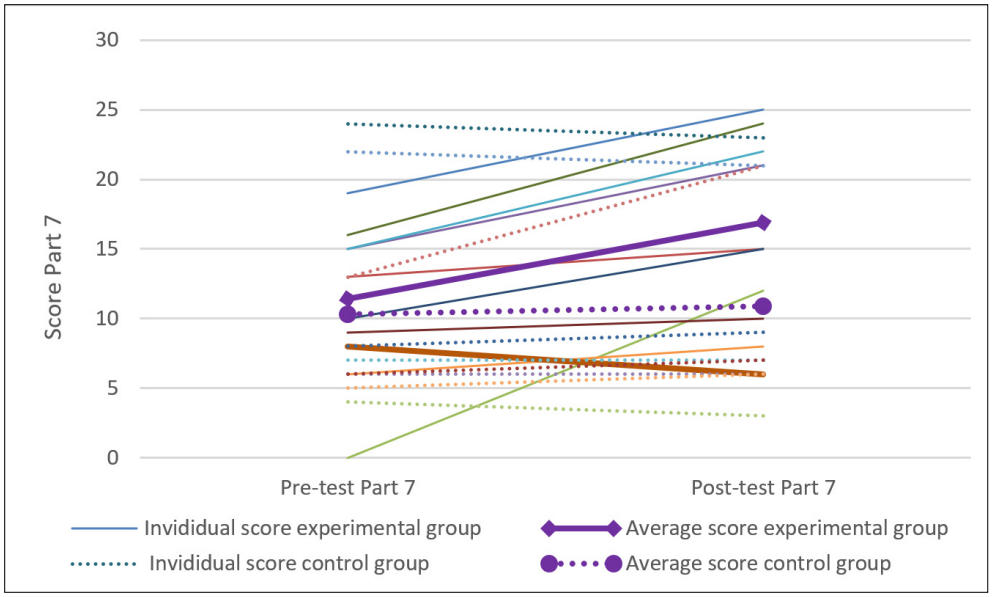


Figure 6

Grouped Spaghetti Plot of the Performances of the Control and Experimental Groups in Part 7



group' and 'proficiency level', in order to measure the combined effects of the two variables on performance after intervention.

Table 8 shows the results of the multiple regression test. With a $p < .001$, the F-test is statistically significant. This shows that the model itself is statistically significant, and the independent variables reliably predict the post-test score (the dependent variable). The R-squared is 0.817; meaning that approximately 82% of the variability of post-test score is accounted for by the variables. The adjusted R-squared shows that about 78% of the variability of post-test score is explained by the variables, even after taking into account the number of independent variables in the model. The Beta coefficients (Beta weighting) value for the 'intervention group', 'proficiency level', and 'interaction' are -.398, .562, and .417, respectively. Thus, for every unit of standard deviation increase in 'intervention group', a .398 standard deviation decrease in post-test score is predicted, if it is assumed that the other variables in the model are constant. Similarly, every unit standard deviation increases in 'proficiency level' leads to a .562 standard deviation increase in predicted post-test score. Every unit standard deviation increase in 'interaction' leads to a .417 standard deviation increase in predicted post-test score with the other variables remaining constant. Since the Beta coefficients of the independent variables are

computed relative to each other, the 'proficiency level' has the strongest positive effect on post-test score. The t-value for the 'intervention group', 'proficiency level', and 'interaction' are -1.185, 1.604, and 0.954, respectively. However, $p > .005$ for each of the predictors. Therefore, the coefficients for each of the predictors are not statistically significant. This outcome could be due to the relatively small size of the sample. Consequently, the effects of the predictors are true for this model, but cannot be generalized.

DISCUSSION

The analysis of the participants' total scores as well as their scores in the three sections (Part 5, 6 and 7) of the reading test showed that the two groups had comparable English reading comprehension skills before the experiment. Consequently, it is safe to conclude that the control and experimental groups involved in the experiment could be adequately compared, in order to achieve the purpose of this study (Cohen et al., 2017). After intervention, the experimental group achieved a higher score than the control group, and data analysis showed that the difference in performance was moderate and not ignorable. However, this outcome is true only for the sample of participants in this study and cannot be generalized.

The higher performance of the experimental group in the post-test was attained due to the high score of its performance in Part 7. The two groups achieved similar performances in Parts 5 and 6, with practically no difference in this

Table 8

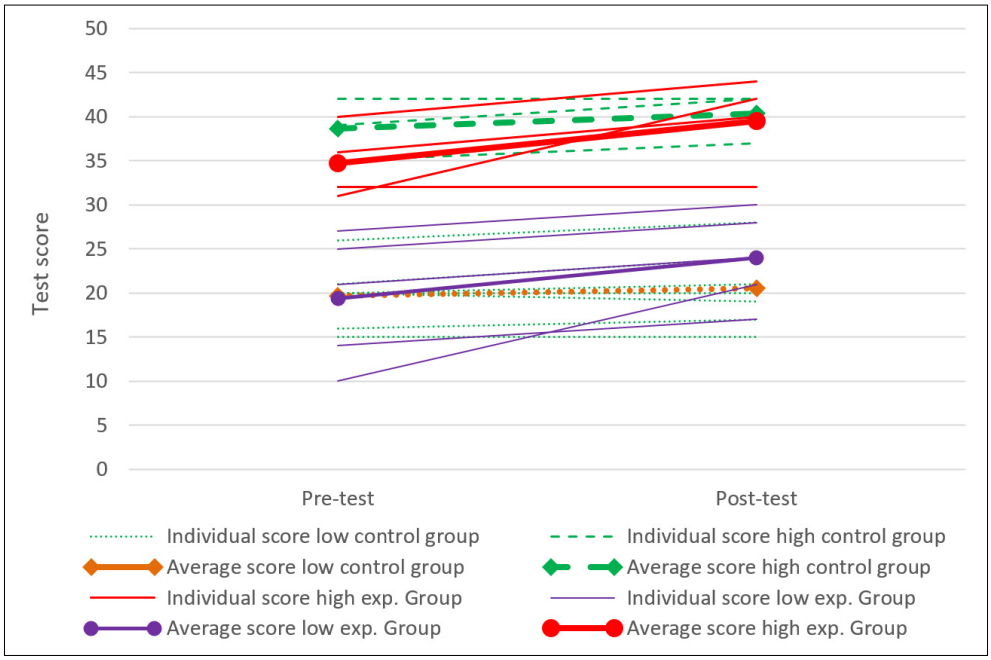
Multiple Regression Predicting the Effects of the Intervention and the Participants' Proficiency Level on Achievement in the Post-Test

Predictor	Zero-order	Unstandardized Coefficients		Standardized Coefficients		
	r	B	Std Error	β	t	Sig.
Intervention group	*-.227*	*-7.690*	*6.492*	*-.398*	*-1.185*	*.255*
Proficiency level	*.893*	*11.238*	*7.005*	*.562*	*1.604*	*.129*
Interaction	*.540*	*4.262*	*4.468*	*.417*	*.954*	*.355*
R	*= .904*	*Adjusted R = .781*				
R^2	*= .817*					
F-ratio	*= 22.346*	*p < .001*				
N	*= 19*					

Note: Intervention groups: experimental group =1; control group = 2; Proficiency level: love proficiency = 1; high proficiency = 2

Figure 7

Grouped Spaghetti Plot of the Performances of the Low and High Proficiency Subgroups in the Control and Experimental Groups



section in the pre-test and post-test. The lack of improvement in these sections of the test was unexpected considering that formal EFL education in Japan is highly dominated by grammar and grammar-translation classes (Egitim, 2020; Kumagai, 1994; Steele & Zhang, 2017). Another possible reasons could be that little time was dedicated to this section of the test during intervention (1/3 of intervention time), and the time was shared between discussing both vocabulary and grammar points that emerged in the questions. Since all participants are working professionals who have been out of formal education, they probably need more time to review and relearn detailed grammatical concepts.

Part 7 focuses on reading and answering literal and inferential questions on various types of texts. In order to help improve reading inferential skill, the reading instructions focused on identifying main ideas and generating paraphrased summaries. Performance improvement in this section of the test supports the importance of inferential skill in the reading process (Grabe & Stoller, 2011; King, 2007). The score improvement in Part 7 for both the control and experimental groups could be a reflection of the substantial time assigned to this section during treatment. The group sessions included individual work time in which each participant did a short test, followed by rereading comprehension passages, and completed a paraphrased summary. The individual work time was followed by a group activity session where the participants shared their summaries of reading texts and answers to reading comprehension questions, discussed main ideas and the meaning of complex words or phrases from the texts. The improvement of mean scores of the reading comprehension section for both groups is proof that collaborative work and group discussions during intervention have a positive effect on improving reading comprehension (August & Shanahan, 2006; Grabe & Stoller, 2011; King, 2007).

In comparison to the control group, the experimental group's improvement in the reading section of the test was relatively substantial, even though both groups were exposed to similar instructions, with the exception of rules related to language use. As a result, it could be said that the difference in performance between the two groups occurred because of the difference in language rules. The experimental group having been allowed to translanguage by using both their L1 and English, it could be deduced that translanguaging has contributed to the relatively high gain in mean score in the experimental group (Hungwe, 2019; Makalela, 2015).

The effect of language restriction was apparent throughout treatment. During group discussions in the control group, in most sessions, participants were generally hesitant to volunteer and express themselves. The more proficient participants were relatively proactive and volunteered to share their answers to questions from the tests. They read aloud the paraphrased English summary written on their notes, or attempted to explain the meaning of phrases from the reading passages. However, sometimes they would stop in the middle of their utterance when they were not sure how to complete their statement in English. Some participants would start a question in English, but give up halfway, or apologetically rephrase the whole questions in Japanese. Less proficient participants had a lot of difficulty in elaborating on the reasons why they selected a particular answer. They ended up either hesitantly reading an explanation from their notes or answering in just a few words in English. Furthermore, in the control group, participants often asked the teacher if they could express themselves just in Japanese.

The limitation of language to English apparently restricted the participants' freedom to express themselves. They were constrained to think and negotiate meaning in a language they were comparatively less proficient in. This probably worked against the development of reading comprehension skills since the proficiency level of language use during the reading skill development influences their improvement level (Spinelli, 2017).

In contrast to the control group, in the experimental group almost all participants talked for most of the intervention time, except in the first reading instruction session during which most participants were hesitant to volunteer and answer questions or express their opinions. The low frequency of interaction in the first session was probably because most participants were meeting for the first time. However, from the second session, every participant contributed to the class. The less proficient participants asked many questions about new vocabulary, phrases, or asked for clarification of the main ideas of passages, and highly proficient participants offered elaborated answers, mostly in Japanese, but sometimes in English, or both. In comparison to the control group, there was less hesitation from low proficiency participants to ask questions and share their opinions.

There was a significant increase in the total reading score of the experimental group. On the other hand, the standard deviation of the post-test (*SD* = 9.897) score decreased slightly in comparison to the pre-test (*SD* = 9.532). This outcome suggests that participants at all levels were able to increase their total score. It could be inferred that translanguaging instruction benefits also participants with lower linguistic skill by allowing them to improve linguistic skills in the weakest language in their linguistic repertoire (Baker, 2011; Garcia & Li Wei, 2014; Lewis *et al*., 2012). Translanguaging appears to generate a relatively calming and relaxing learning environment, which in turn, encourages learners to be confident and proactively use the whole of their linguistic repertoire (Chukly-Bonato, 2016).

Proficient participants used English and Japanese interchangeably, while the less proficient ones used their L1 at a much higher rate than English. At times, the instructor would encourage participants, especially less proficient

ones, to use English as much as they could. Such encouragement resulted in relatively more language mixing and alternation from less proficient learners. The use of the L1 and L2 interchangeably probably acted as scaffold for participants to express their opinion, ask questions for clarification and deepen their understanding of the passage, thus correcting misunderstanding (Motlhaka, 2021). In other words, through group discussion (August & Shanahan, 2006; Grabe & Stoller, 2011; King, 2007) and translanguaging (Makalela, 2015; Motlhaka, 2021) less proficient participants could learn from more proficient learners and make more accurate inferences from reading passages. Consequently, not only proficient participants, but also less proficient ones were able to achieve a relatively higher score in the post-test in comparison to participants in the control group, as shown in Figure 7.

Taken together, translanguaging instruction with group discussions enable participants to display a high level of motivation and confidence. They share their thoughts when they are allowed to express themselves freely, thus making use of all resources available in their linguistic repertoire (Motlhaka & Makalela, 2016). This probably applies also to group discussion where language learners try to make sense of written contents. When learners use their L1 and other languages they are proficient in, they can perform at a higher cognitive level and better understand text with difficult syntactic structures, vocabulary, and idioms. In translanguaging instruction, learners read a text in one or several languages and engage in a discussion where they use all languages in their repertoire (Garcia, 2009). Throughout this practice, the learners move freely between languages, making use of aspects of each language which can help them make sense of the text on focus (Garcia, 2009; Otheguy, Garcia & Reid, 2015; Tian et al., 2020). Thus, through group discussions and translanguaging processes, learners develop their overall linguistic skills, notably in the target language. Learners challenge themselves to understand paragraphs, sentences as well as idioms and vocabulary from the text. In order to better grasp the meaning of specific passage from the author, less proficient participants ask for help from their peers. From the attitude of the participants in the experiment, it can be inferred that group discussions through translanguaging reduces the anxiety of low proficiency learners and their reluctance to speak both the target language and their L1 and play an active role in their learning (Baker, 2011; Garcia, 2009; Motlhaka, 2021). In the meantime, it gives the opportunities to more proficient learners to support their peers by contributing with explanations of difficult sentences, idioms, and words. Thus, group discussions provide a collaborative learning opportunity which helps to scaffold reading comprehension, while translanguaging instruction enhances the scaffolding aspect of collaborative learning, thus facilitating learning for participants, especially less proficient ones (Motlhaka, 2011). The results of our experiments and analysis shows that participants in the experimental group in this study display all the characteristics and effects of translanguage aforementioned.

Despite the positive outcomes of translanguaging on reading comprehension obtained through this study, this project had some limitations. The most noticeable is the number of participants. A total of 19 people took part in the project, but this number remains considerably small in comparison to similar translanguaging studies in the literature which often enroll around 50 people or more. The low number of participants makes it difficult to generalize the finding of this study. Another limitation is, the reading comprehension intervention conducted during this study lasted a total of ten hours over a period of 18 days. However, in most of the research found in the literature, intervention time lasts at least a semester. Consequently, it is difficult to compare adequately the outcome of this research to the findings in the literature.

CONCLUSION

This study suggested that improvement in the experimental group was moderately more substantial. Consequently, it could be concluded that translanguaging instruction helped the Japanese EFL learners in the workforce in this study improve their reading comprehension to a moderate extent. However, since the score difference between the two groups was not statistically significant, it was difficult to generalize the result of the findings. Nevertheless, it was demonstrated that translanguaging instruction, which encourages participants to express themselves freely using all resources available in their linguistic repertoire, encourages participants to display a high level of motivation and confidence to share their thoughts in group discussions. The learners especially used their L1 to perform at a higher cognitive level and better understand difficult texts. They moved freely between languages and used aspects of each language in their linguistic repertoire which facilitated their making sense of the texts. Furthermore, translanguaging instruction reduced the low proficiency learners' anxiety and reluctance to participate in class and played an active role in their learning. Thus, less proficient learners could make the most of the enhanced scaffolding aspect of collaborative learning opportunities provided by translanguaging instructions.

Literature on translanguaging in Japan is scarce. Moreover, most of the literature available is focused on formal education, especially at college level, and literature on the practice of translanguaging practice in continuing educating is non-existent. Therefore, this paper not only contribute to the literature on translanguaging in Japan, but also shows the potential effectiveness of this emerging pedagogy in the continuing education context in Japan where learners unarguably need reliable ways to improve their English skills for career advancement. As it could be seen in the discus-

sion, this project displayed some limitations which has left much to be covered in research on translanguaging in the context investigated. A range of different projects could be designed based on these limitations, in order to study the effectiveness of translanguaging in EFL education in the continuing education context in Japan.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print*. MIT Press.

August D., & Shanahan T. (2006). Developing literacy in second-language learners: Report of the National literacy panel on language-minority children and youth. *Journal of Literacy Research, 42*(3), 341-348. https://doi.org/10.4324/9781315094922

Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Multilingual Matters.

Bartlett, K.A. (2018). Applying translanguaging techniques in Japanese EFL settings. In *The Asian Conference on Language Learning 2018. Official Conference Proceedings* (pp. 239–251). IAFOR.

Carroll, T. (2013). *Language planning and language change in Japan: East Asian perspectives*. Routledge.

Cenoz, J. & Gorter, D. (2011). Focus on multilingualism: A study of trilingual writing. *Modern Language Journal, 95*(3), 356–69. https://doi.org/10.1111/j.1540-4781.2011.01206.x

Cenoz, J., & Gorter, D. (2022). *Pedagogical translanguaging.* Cambridge University Press.

Chukly-Bonato, K. (2016). *Transferring knowledge through translanguaging: The art of multilingualizing the foreign language classroom*. McGill University Libraries.

Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education* (8th ed.). Routledge.

Cook, V. (1997). The consequences of bilingualism for cognitive processing. In A. M. B. De Groot, & J. F. Kroll (Eds), *Tutorials in Bilingualism* (279–299). Cambridge University Press.

Cook, V., & Bassetti, B. (2005). *Second language writing systems*. Multilingual Matters.

Csikszentmihalyi, M. (1990). The Domain of Creativity. In M. A. Runco & R. S. Albert (Eds.), *Theories of Creativity* (pp. 190–212). Sage.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Egitim, S. (2020). *Understanding Japanese university English teachers' experiences as collaborative leaders: engaging learners in teaching and classroom management* [Doctoral dissertation]. ProQuest Dissertations and Theses Global. https://doi.org/10.17760/D20394199

Fisk, C. & Hurst, B. (2003). Paraphrasing and comprehension. *The Reading Teacher, 57*(2), 182–185. https://www.jstor.org/stable/20205336

Garcia, O. (2009). *Bilingual education in the 21st century: A global perspective*. Wiley/Blackwell.

Garcia, O., & Li Wei (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan Pivot.

Gardner, H., & Hatch, T. (1989). Educational implications of the theory of multiple intelligences. *Educational Researcher, 18*(8), 4–10. https://doi.org/10.3102/0013189X018008004

Goli, A. (2021). One-on-one English conversation lessons: Evaluating learners' satisfaction and self-perceived oral communication improvement. *Shonan Journal, 12*, 51–61. https://doi.org/10.15034/00007768

Grabe, W. (2009). *Reading in a second language* (3rd ed.). Cambridge University Press.

Grabe, W. & Stoller, F.L. (2011). *Teaching and researching reading* (2nd ed.). Pearson Education.

Hirvela, A. & Du, Q. (2013). Why am I paraphrasing? Undergraduate ESL writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes, 12*(2), 87–98. https://doi.org/10.1016/j.jeap.2012.11.005

Hungwe, V. (2019). Using a translanguaging approach in teaching paraphrasing to enhance reading comprehension in first-year students. *Reading & Writing 10*(1), a216. https://doi.org/ 10.4102/rw.v10i1.216

Hussein, B. (2013). Teaching and learning English as a second/foreign language through mother tongue: A field study. *Asian Social Science, 9*(10), 175–180. https://doi.org/10.5539/ass.v9n10p175

King, A. (2007). Beyond literal comprehension: A strategy to promote deep understanding of text. In S. D. McNamara (ed.), *Reading comprehension strategies: Theories, interventions and technologies* (pp. 267–290). Lawrence Erlbaum.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning Supplement, 57*, 1–44. https://doi.org/10.1111/0023-8333.101997010-i1

Kor, C. P., Low, H. M., & Lee, L. W. (2014). Relationship between oral reading fluency and reading comprehension among ESL students. *GEMA Online Journal of Language Studies, 14*(03), 19–32. https://doi.org/10.17576/gema-2014-1403-02

Kumagai, Y. (1994). *The effects of culture on language learning and ways of communication: The Japanese case* [Master's Capstone Project, Centre for International Education]. University of Massachusetts Amherst.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323. http://dx.doi.org/10.1016/0010-0285(74)90015-2

Lewis, G., Jones, B., & Baker, C. (2012). Translanguaging: origins and development from school to street and beyond. *Educational Research and Evaluation, 18*(7), 641–654. https://doi.org/10.1080/13803611.2012.718488

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 13*(2), 123-146.

Lowie, W., & Seton, B. (2012). *Essential statistics for applied linguistics* (2012th ed.). Palgrave Macmillan.

Makalela, L. (2015). Moving out of linguistic boxes: the effects of translanguaging strategies for multilingual classrooms. *Language and Education, 29*(3), 200–217. https://doi.org/10.1080/09500782.2014.994524

Makalela, L. (2016). Translanguaging practices in a South African institution of higher learning: A case of ubuntu multilingual return. In C.M. Mazak, & K.S. Carroll (Eds), *Translanguaging in Higher Education* (65–78). Multilingual Matters. https://doi.org/10.21832/9781783096657- 004

Motlhaka, H. (2021). Translanguaging in collaborative reading activity: A multilingual perspective of meaning making. *Psychology and Education, 58*(5), 2683–2691.

Nagy, T. (2018). On translanguaging and its role in foreign language teaching. *Acta Universitatis Sapientiae, Philologica, 10*(2), 41–53. https://doi.org/10.2478/ausp-2018-0012

Nagy, W., Garcia, G.E., Hancing-Bhatt, B., & Durgunoglu, A. (1993). Spanish-English bilingual students' use of cognates in English reading. *Journal of Reading Behaviour, 25*, 241–259. https://doi.org/10.1080/10862969009547816

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

Paltridge, B., & Phakiti, A. (2015). *Research methods in Applied Linguistics: A practical resource (Research Methods in Linguistics)* (2nd ed.). Bloomsbury Academic.

Perfetti, C.A. (1999). Comprehending written language, a blueprint for readers. In C.M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). Oxford University Press.

Pressley, M. (2000). What should reading instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (vol. 3, pp. 545–563). Lawrence Erlbaum.

Spinelli, B. (2017). The multilingual turn in FL education. *Translation and Translanguaging in Multilingual Contexts, 3*(2), 184–209. https://doi.org/10.1075/ttmc.3.2.03spi

Steele, D. & Zhang, R. (2017). Enhancement of teacher training: Key to improvement of English education in Japan. *Procedia-Social and Behavioral Sciences, 217*(2016) 16–25. https://doi.org/10.1016/j.sbspro.2016.02.007

Storch, N. & Wigglesworth, G. (2003). Is there a role for the use of the L1 in an L2 setting? *TESOL Quarterly, 37*(4), 760–770. https://doi.org/10.2307/3588224

Takahashi, J. (2012). An overview of the issues on incorporating the TOEIC test into the university English curricula in Japan. *Departmental Bulletin Paper, Tama University Global Discussion Department, 4*, 127–138.

Tian, Z., Aghai, L., Sayer, P., & Schissel, J. L. (2020). Envisioning TESOL through a Translanguaging Lens in the Era of Post-multilingualism. In Z. Tian, L. Aghai, P. Sayer, & J. L. Schissel (Eds.), *Envisioning TESOL through a translanguaging lens: Global perspectives (Educational Linguistics, 45)* (1st ed., pp. 1–20). Springer. https://doi.org/10.1007/978-3-030-47031-9_1

Tsedal, N. (2017). *The language of global success*. Princeton University Press.

Turnbull, B. (2018). Is there a potential for a translanguaging approach to English education in Japan? Perspectives of tertiary learners and teachers. *JALT Journal, 40*(2), 101–134.

West, S.G., Finch, J.F., & Curran, P.J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (1st ed.) (56–75). Sage.

Williams, C. (2012). *A language gained: A study of language immersion at 11-16 years of age*. University of Wales.

https://doi.org/10.17323/jle.2023.13989

Enhancement of Academic Performance through Developing Cross-Cultural Communicative Competence: A Case Study of Students Majoring in Economics

Nataliia Guskova [1], **Elena Golubovskaya** [2]

[1] HSE University, Moscow, Russian Federation

[2] RUDN University, Moscow, Russian Federation

ABSTRACT

Background. The article questions the possibility to increase the level of foreign language command through developing cross-cultural communicative competence (CCC) in students of non-linguistic universities. Despite extensive literature on intercultural communication, there are obviously gaps in investigating the way it can and should be built and the potential impact it may have towards students' academic performance in general.

Purpose. The paper aims to find out the possible correlation between the level of CCC formation and a command of a foreign language in general. To pursue the goal, an in-depth research into the CCC structure was carried out and the idea to simultaneously develop all its components was proposed.

Method. The paper reports on the results of the mixed-method research aimed at gathering the data and evaluating them both qualitatively and quantitatively. Senior students of the Higher School of Economics, Moscow, Russia, were selected and divided into the control and experimental groups with the subsequent training on the pre-designed curricula, with a primary focus on developing all the components of the CCC in the latter. To assess their performance, the method of experimental verification, self- and peer evaluation, educational observation, questioning method were employed. In-depth data analysis and verification provided post-active phase of the experiment conducted.

Results. The results of the carried-out experiments, first, proved the hypothetical assumptions on the efficiency of developing all the four CCC simultaneously and, second, showed that the targeted CCC development contributes to improving foreign language acquisition in general, which is supported by the increase in 5 out of 6 didactic units of the final testing where the experimental group participants surpassed the students in the control group.

Conclusion. The study provides evidence for the impact CCC development has on the linguistic communicative competence. The devised methodology can be borrowed and customized for teaching foreign languages to university students and, in particular for developing intrinsic motivation through CCC. Further, future research should address particular components of the CCC.

KEYWORDS

cross-cultural communicative competence (CCC), cross-cultural communication, academic performance

Citation: Guskova N., & Golubovskaya E. (2023). Enhancement of Academic Performance through Developing Cross-Cultural Communicative Competence: A Case Study of Students Majoring in Economics. *Journal of Language and Education, 9*(1), 76-88. https://doi.org/10.17323/jle.2023.13989

Correspondence:
Nataliia Guskova,
hse_guskova@mail.ru

Received: Mach 03, 2022
Accepted: Mach 15, 2023
Published: March 31, 2023

INTRODUCTION

In the continuously developing educational settings with geopolitical relations and values recalibrated and new challenges arising, good awareness of cultural conventions and global tolerance is actual as never before (Tikhonova et al., 2021). That is why, developing cross-cultural communicative competence (CCC) is within the most necessary set of skills to be acquired for successful implemen-

tation of business and professional activities (Johnson, 2006; Martin, 2015; Liao, 2020; Bagiyan et al., 2021), starting with doing a degree (Gierke, 2018; Liao, 2020).

So, nowadays delivery of cross-cultural education in a modern University should become a focal point. It can prepare graduates for multicultural environment where they will be able to effectively interact with representatives of other cultures, develop the ability of cultural self-determination, and establish value orientations for themselves in relation to their own culture (Deardorff, 2006a) and the "other" communities (Arasaratnam & Doerfel, 2005; Arasaratnam & Banerjee, 2011). In addition, building CCC contributes to the development of a modern democratic state, which is characterized by tolerance of views and opinions, recognition and development of cultural pluralism in society, and equal opportunities for all citizens. Announced as a commitment back in 1994 by UNESCO in Geneva, multicultural education is likely to pave the way for acknowledging individuals of different societies and sexes, "encourage convergence of ideas" to strengthen brotherhood and solidarity, increase the ability to overcome conflicts and "promote peacefulness in the minds of students" (Zamroni et al., 2021, p. 598).

The issues related to CCC as an essential element of successful activity in the professional sphere are being widely discussed nowadays. This is due to a number of factors: firstly, the need for a more detailed study of approaches to teaching students, taking into account cross-cultural differences between communicants, the volume of background knowledge, communication topics, situational conditions in which the interaction is carried out (Andreyeva et al., 2015); secondly, the importance of studying the way semantic content of information is transmitted within cross-cultural communication between students belonging to different cultures and languages (Rasmussen et al., 2015); thirdly, the importance of cultural knowledge that specialists need for an adequate and maximum complete interpretation of received information in specific situations of communication with speakers of another language and a different culture (Menuzane, 2021).

To avoid terminological confusion, based on the available literature, we consider the cross-cultural competence, intercultural competence and multicultural competence as complete synonyms without any difference in meanings. Out of these three, the basic concept for the current study is supposed to be cross-cultural communicative competence (CCC), though we rest on investigations of others devoted to either of these three.

In this study, we went beyond the existing experience and focused on the correlations of CCC development and acquiring a foreign language rather than enhancement of intercultural relations. In order to pursue this goal, we explored the impact CCC has on the overall level of the foreign language proficiency. Presumably, we expected the direct correlation between the two. Thus, to approach the problem we had to identify the key elements of CCC in order to devise the way it can be comprehensively developed. The study rests on the elaborations of the previous research, though it expands and reframes the existing concepts to structure CCC as a product of any intellectual activity thus underpinning the integrity of developing language capabilities and a communication competence in intercultural context. For this reason, an entirely new, praxeological, criterion was added to the widely-recognized CCC framework.

Thus, the research questions we address by our paper can be formulated as below:

(1) What are the components of cross-cultural communicative competence?

(2) Does teaching targeted at developing cross-cultural communicative competence lead to progress of all its elements?

(3) Can development of cross-cultural competence have an impact on overall academic performance in terms of learning a foreign language?

Hypothetically, it was assumed that targeted teaching students in the area of intercultural communication is most likely to lead to the overall increase as well as the progress in each separate element within it. Whether it can positively influence a student's foreign language capabilities was a much more controversial issue and we were quite tentative to state our hypothesis about its positive impact.

LITERATURE REVIEW

Research Focus

Developing cross-cultural competence has repeatedly been studied in scientific literature. Primarily, authors elaborate on the generalized and empirically derived definitions of the key concept(s) (Gierke et al., 2018; Liao & Thomas, 2020); actively discuss its constituents and build its frameworks and competency areas, mostly for exploratory and applicable purposes (Byram et al., 2001, Gay, 2001; Deardorff, 2006b, Rasmussen et al., 2015, Barzykowski et al., 2019) or focus on the role culture plays in building CCC (Marxsen, 2007; Kramsch, 2008; Schaetti et al., 2009; Piasecka, 2011).

The importance of the link between theory and practice, which contributes to the applied nature of the research into intercultural competence, is identified as one of the key directions. So, the investigation of the paradigms of intercultural communication and their possible application in the learning process are studied by Bleszynska (2008), who claims intercultural education is a necessary part of any modern school, and Aririguzoh (2022) whose elaborations

stem from the necessity to improve cross-cultural communications through building inter-cultural competences.

Above that, a number of authors focus on the ways CCC can be developed in different professional fields, which corresponds to the main goal of the given research aimed at students of a multidisciplinary University providing higher professional education (Andreyeva et al., 2015; Bennett, 2009; Hammer, 2009; Jackson, 2009; O'Brien, 2019; Sehlaoui, 2001). A lot of investigations summarize on the ways intercultural communicative competence is developed in instructors rather than learners (Bickley et al., 2014; Chernova, 2018).

Terminology Issues

Despite the extensive literature on general issues of intercultural communication and communicative competences as they are, much less attention is paid to the area of intercultural (cross-cultural) communicative competence (Bickley et al., 2014, p. 138). This must be the main reason why there does not appear to be consensus on what exactly cross-cultural competence is (Gierke, 2018). The given paper seeks to fill in this niche as the meaning of CCC needs be clarified before it is measured and further conclusions are made (Deardroff, 2006b). To avoid confusion, first of all, a line of demarcation should be drawn between intercultural competence and intercultural communicative competence. The former is applied to using a mother tongue in interactions with people from a different culture whereas the latter means doing so in a foreign language (Byram, 1997). Secondly, a decision should be taken whether the author borrows the predominant, fundamentally theoretical definition of the concept under investigation (Spitsrberg, 2009, Deardroff, 2006a) or accepts the empirically driven meaning (Gierke, 2018).

Considering the practical focus of this study, there seems to be more advantages to the meaning embedded in the empirically valid definitions. So, communicative competence can be understood as "the ability to function effectively in another culture" (Johnson, 2006, p. 527). Bickley *et al*. refer to Adult ESL Curriculum Framework, which defines it as an "ability to communicate effectively and appropriately in English within a culture diverse society" (Bickley et al., 2014, p.138). Tallen (2010) defines CCC as a systemic unity that integrates personal, cultural and professional characteristics and components.

An even more pragmatic approach is taken by the methodologists. Thus, Merrifield et al. (2008), studying cross-cultural communicative competence, consider it as an ability to act in a situation of uncertainty; Brown & Terrell (2000) regard it as the willingness of a specialist to be engaged in certain activities. According to Peart (2019), cross-cultural communicative competence is a combination of interconnected personal qualities (knowledge, abilities, skills, ways of working) defined in relation to certain intercultural processes, and necessary for high-quality academic professional activity.

CCC Inventory and Measurement

To be able to do a quantitative and qualitative assessment of the level CCC is developed in students, the full potential of its constituents should be revealed and described. One of the recognized structural frameworks of CCC was proposed by Byram and Zarate (Byram & Zarate, 1996) and adopted by the Common European Framework of Reference (CEFR 2001). According to this model design, in the frameworks of foreign language education, there can be found four dimensions known under the French titles: "saviors" – knowledge of Self and Others, "savoir comprendre" – skills of interpreting and relating, "savoir être" – intercultural beliefs and attitudes, "savoir faire / apprendre" – skills of discovery and interaction (Byram & Zarate, 1996, 11).

In later studies the Byrame's model was accepted as "prominently influential" (Munezane, 2021), though it has been subject to critical revisions and reconsiderations, e.g. Hoff introduced to it the new 'savoir s'engager' component derived from the postulate that "disharmony and conflict are recognized as potential aspects of the intercultural dialogue in Byram's model" (Hoff, 2014). Consequently, the study vector shifted to constructing more complex models by integrating multidimensional variables often verified by surveying experiment participants. Thus, the model suggested by Arasaratnam et al. (2011) identifies experience, listening skills, positive attitudes toward people from other cultures, motivation to interact with people from other cultures, and ability to empathize as "variables contributing to ICC". Munezane (2021) proposed the nine-factor theoretical model of ICC based on "openness to experience, interdependent self-construal, ethnocentrism, willingness to communicate, motivation to introduce one's culture, constructive conflict resolution, motivation to engage in intercultural communication, confidence in intercultural interaction". Zamroni et al. (2021) test 4 factors: cultural awareness, understanding of own and other cultures, cultural competence and cultural encounter. They collected data by questionnaires and apply MANOVA test to process and cross compare the findings.

The available evidence is enough to see the corresponding relations between the approaches taken, e.g. knowledge of Self and Others is alternatively treated as declarative knowledge while skills of discovery and interaction is understood to be an ability to learn. What these models lack is, however, the bridge between intercultural interactions and the foreign language learning environment. This paper seeks to consider CCC as a tool to enhance general linguistic competences, that is why we are going to expand the existing inventory and reframe it to adjust to our specific needs.

According to the cognitive, effective, behavioral etc. components integrated into the CCC model researchers deploy a wide variety of appropriate tools including surveys, interviews, questionnaires etc. to assess to what degree each or all of them are developed (Johnson, 2006; Rasmussen, 2015;

Munezane, 2021). For example, it can be a survey with questions roughly grouped according to the "cultural awareness, understanding of own and Others' cultures, cultural competence and cultural encounter" (Zamroni et al., 2021; Munezane, 2021). There are, however, papers which dismiss the possibility for administered evaluation sessions and insist on the continuous forms of measuring the level for intercultural communicative competence (Lussier et al., 2007), but commonly they are outnumbered by the quantitative studies.

The exploration of the previous research provided solid background for devising and further implementation of our own CCC structural framework based on the organization of the foreign language learning process and measuring the degree it is formed in students. Additionally, the study fills the gap in previous research by building a bridge between intercultural education and overall communicative competence. Correlations between language learning and intercultural competence promotion have been a subject of earlier research marked with a diversity in focus. For instance, Starkey's study explored the tension between persistent traditions of teaching English and language learning policies that promote intercultural competence (Starkey, 2007); Tran et al. (2018) pursued a goal most similar to the one stated by this research and built a model of integrating intercultural competence into teaching English in the Vietnamese context, though did not particularly focus on the direct correlation between the students' gains and developing an intercultural communicative competence in them. East et al. (2022) report on a series of empirical research to prioritize the development of intercultural communicative competence in foreign language education as a tool part of intercultural citizenship, but not vice versa. So, to our knowledge, no attempt has been made to immediately trace if the CCC development can be a good venue for enhancing foreign language competences, and the given research seeks to contribute to this area of investigation.

METHOD

Participants

The population of the research consisted of 87 first year students (49 female, 38 male, aged 18-21 y.o., the mean is 22,68 y.o.) of the Faculty of Economics of the National Research University "Higher School of Economics". According to the placement English test, 92 students were assessed to have B2 level of English mastery, 87 people out of them agreed to participate in the study, 5 refused for different reasons. Before the experiment they were explained the main goals and objectives of the experiment and divided into two groups: the experimental and control ones. For technical convenience, each of the groups was subdivided into two subgroups without any difference in the course content and its delivery. Their participation and the group choice was absolutely voluntary. 12,6 % of the participants - 11 people - were ready to join either of the groups, 43 students expressed their intention to join the experimental group and 33 preferred to study in the control group.

Data Collection and Analysis

Structural Frameworks- Pre-Active Experiment

Based on the earlier attempts to propose the structural frameworks of CCC, we identified four our own criteria. In order to do that, we followed the main principles of any general form of intellectual activity, which, as we suppose, may enrich our understanding of the skill set this type of communicative competence contains. Thus, in our view, the main criteria related to the development of intercultural communicative competence include: *epistemological criterion,* which helps to determine the level of communicatively valuable knowledge; *praxiological criterion*, as a manifestation of the complex of necessary skills and abilities, as well as the communicative qualities of the individual and the nature of communication; *motivational-axiological criterion*, which characterizes the formation of communicative motivation and value relations; and the *reflexive criterion*, which is represented in the level of students' interest in the issues of optimal communication, self-control, and the desire for self-improvement based on the analysis of their own outcomes. These criteria of the CCC formation are implemented through the relevant indicators presented in table 1.

Didactic Potential of CCC Components

Our hypothetical views suggested the necessity to simultaneously develop all the four aspects of the CCC identified. In our opinion, such an approach would allow to fully attain the high levels of the skills necessary to successfully communicate with people from other cultures. Let us describe in more detail each of the criteria.

As seen from Table 1, among the indicators of the epistemological criterion of CCC, are the level of demonstrating a set of communicatively valuable knowledge, as well as its consistency and completeness. The level of mastery of the norms and conventions can reveal through different activities.

The praxeological criterion revealed itself through such indicators as the use of communication means, the degree of mastery of intercultural communicative skills and strategies, the degree of their variability, the degree of manifestation of individual communicative qualities and the preferred communication style.

The need to identify the motivational-axiological criterion was due to the fact that, like others, the CCC is formed under the influence of the person's attitudes, motives and inter-

Table 1

Criteria and Indicators of Intercultural Communicative Competence Formation

Criteria	Indicators
Epistemological	the level acquisition of cross-cultural communicative knowledge; its consistency and completeness
Praxeological	the degree of mastery of CCC skills, strategies and means of communication; their range; the degree of CCC qualities manifestation in behavior; communication style.
Motivational and axiological analysis	the level of CCC motivation development; the level of formation of the attitude to communication as a value; the - degree of desire to master cross-cultural communicative competence; the level of cognitive interest.
Reflexive	the degree of involvement and initiative; sufficiently developed ability to conduct self-analysis and ability to exercise self-control; the level of eagerness to improve themselves.

ests. The indicators of this criterion are: the level and type of the attitude to CCC as a value, the degree of students ' desire to form the competence under study, the level of cognitive interest to intercultural interactions and the desire to be able to successfully communicate with representatives of different cultures.

The reflexive criterion is characterized by a student's attitude to the activities described by such indicators as degree of activity and initiative, both in the educational and intercultural communicative activity, its self-awareness, the ability to exercise self-control, awareness of the need for self-improvement. Accordingly, the above-mentioned indicators, such as the quality and degree of manifestation in their totality, give grounds to assign students to a particular group of the CCC formation in accordance with their level of development.

All levels have close interrelations, since each of them acts as a certain condition for the transition to the next stage, that is, the final result of mastering the previous level. All the outcomes gained this way make up "reference points" that allow to observe CCC formation in students and coordinate it, thereby managing the process under study.

Grouping Participants

When considering the formation of the CCC degree in learners, it is necessary to rely on several levels of activity development, namely: 1) the level of familiarity and awareness revealed through recognition of realities; 2) the level of reproduction, i.e. reproduction of information; 3) the level of productive activity; 4) the level of productive activity on any set of objects by independently constructing the program of activity. Thus, the degree of manifestation of the student's ability and readiness in this study, elementary, limited, sufficient and advanced levels of CCC formation were identified.

The basic level is characterized by the minimal volume of random knowledge, a low level of communicative skill formation, an aggressive style of communication, rigidity, lack of understanding practical significance of the generated competence in professional activities and in life in general, poor motivation, absence or shortage of reflexive attitude to the learning activities, unwillingness to take responsibility for the result of their own educational activities.

The limited level is characterized by the following indicators: the total minor amount of knowledge at a sufficiently low degree of consistency, weak demonstration of communicative skills, a limited set of communication strategies, low interest in developing intercultural communicative competence, as well as low awareness of its significance, insufficient need in self-improvement, poorly developed reflective skills and self-control.

The sufficient level is represented by the presence of a sufficient amount of solid knowledge, which is of a systematic nature, an optimal level of mastery of communication skills and strategies, a sufficient degree of their range, in most cases, the manifestation of communicatively valuable personal qualities, a sufficient level of positive motivation, the ability to control, analyze and regulate their own activities, the vividly expressed desire for self-improvement.

An advanced level of cross-cultural communicative competence can be determined in the presence of profound knowledge, self-determination, a clear focus on achieving high results in future professional activities, a confident command of communication skills and strategies, a high degree of their variety and their fluent usage in relevant communication situations, sorting more complex and professionally oriented communicative tasks, a cooperative style of communication, a strong attitude to self-realization and constant self-improvement, self-initiative in mastering communicative competence, with a high level of reflection and self-control. The final assessment was carried out by using observation, questioning, testing and experts' continuous assessment. The indicators are presented in Table 2 General Characteristics of CCC levels.

Procedure

Basically, the main goal of the study was achieved through pre-active (theoretical), active (experimental) and post-active (diagnostic) stages. Describing the first, theoretical, stage of the experiment, in order to create cross-cultural communicative competence, individually customized meth-

ods of targeted CCC development were devised and developed to be further incorporated into the course of English as a foreign language delivered in the experimental group of students. The course curriculum in this group included the following stages: general acquaintance with the culture of different countries of interest and language training (language learning; development of language skills through self-education; vocabulary acquisition; collection and study of information about the cultural identity of the different countries).

The second, experimental, stage of the study consisted in giving classes and administering tests and exams with participation of 87 second-year students of the National Research University "Higher School of Economics". The practical classes were supplemented with individual tutorials based on active and interactive learning as well as Internet projects and web-quests. The training process in the control group was based on conventional teaching in accordance with the requirements of nationally accepted professional training syllabus of higher education. At this stage of intercultural communicative competence formation, the tailor-made methodology was introduced in the experimental group.

The development of intercultural competence was implemented under the conditions of enhancing students' motivation for intercultural communication (possible thanks to role-playing and simulations, raising awareness of different culture values, modelling success situations in foreign language classes); promoting students' self-reflexive positions in intercultural communication (developing critical thinking, problem-solving, project tasks, handling problems, free choice, accessible presentation of the material, supervised self-assessment, the "portfolio" method).

At the third, diagnostic, stage of the experimental study, rigorous analysis was carried out, methodological recommendations were developed for the use of various practical exercises and tasks in order to form CCC in students. This stage ran in two phases: measuring the degree of the CCC level (including the overall degree and its particular elements based on the structural frameworks) and comparing it against academic performance.

Table 2

General Characteristics of CCC Levels

Levels / Criteria	Elementary	Limited	Professionally sufficient	Advanced
Epistemological	Minimal amount and lack of systematic knowledge. General awareness of speech etiquette and forms of professional communication. Indifferent attitude to incoming data.	A small amount of knowledge; some attempt to systemize information. Basic understanding of basic speech etiquette; lack of ability to elicit information from the message received.	A sufficient amount of systematically organized knowledge. Basic concepts of speech etiquette. Some awareness of professional communication norms.	A significant amount of systemized knowledge. Profound erudition. Good awareness of professional communication manners. Good schemata. Ability to evaluate incoming information.
Praxeological	Inefficient use of verbal and non-verbal communication strategies. Poor communication skills and a narrow range of communication strategies. Low level of tolerance, conflicting inclinations. Expressing ethnocentrism.	The intermediate proficiency in verbal and non-verbal strategies. A small range of communication strategies. Possible communication conflict inclinations. The communication style depends on the speaker's emotional state.	A wide range of developed communication skills. Extended range of strategies. Displaying tolerance, empathy, low conflict inclinations. Respect of values, conventions, and traditions of other cultures.	Confident possession of skills and abilities of effective interpersonal communication. An extended range of communicative strategies. Tolerant behavioral responses, empathy. Cooperative communication style.
Motivational and axiological analysis	Poor communication motivation. Lack of understanding the importance of communicative competence. Lack of desire to master it.	Insufficiently developed communicative motivation. Weak awareness of the communicative competence impact. An unstable desire to master it.	Developed communicative motivation. Engaged attitude towards CCC Sustainable pursuit of its development.	Well-developed motivation. Ability to demonstrate self-determination. Focus on achieving high results.
Reflexive	Passive position, lack of autonomy. Lack of self-control, self-reflection, self-regulation.	Communicative behavior matching the situation. Eagerness to exercise self-control, but insufficient development of reflexive skills.	Continuous mastering knowledge, skills and abilities. Strong analytical skills, self-regulation, self-control, reflexive attitude to learning outcomes.	Initiative to master the communicative competence. Developed self-regulation skills, a high level of reflection. Focus on personal and professional self-improvement.

Data Verification

All criteria were verified through statistical data processing. To summarize on the overall progress throughout the academic year, A.A.Rean's ranking method (Danilova *et al*., 2018) was used. On a special form, the parameters included in the expert assessment were reflected in a tabular form. The students taking part in the experiment ranked all the parameters, assigning them a specific number of points, according to the assumed degree of their importance, after which in the next column («I-concept ») these same parameters were ranked in relation to themselves. When analyzing the results obtained, the difference between the desired and real level of each parameter was made – (d), later squared (d^2). After that, the sum of squares (Xd^2) was calculated and the rank correlation coefficient was determined by the formula: $r= 1 - [6 \Sigma d^2/n (n^2-1)]$, where n is the number of indicators used for ranking. Thus, the closer the correlation coefficient to 1 (from 0,7 to 1,0), the higher is the level of self-esteem, adequate self-assessment corresponds to a ratio of 0,4 to 0,6.

RESULTS

Measuring the Communicative Competence

In order to ensure the correspondence of the test and the structural frameworks of the CCC, we designed the tasks in strict compliance with each criterion recognized. Thus, in order to identify the characteristics of the communicative competence level achieved by the students according to the epistemological criterion, it is necessary to determine the degree of their development of knowledge that plays a significant role in the formation of communicative abilities.

The control task offered to students of both the experimental (EG) and control (CG) groups, was targeted at checking the degree of cultural awareness and the level of knowledge in most common traditions and conventions across cultures, including business etiquette, mostly through questions, both in the closed (multiple choice) and open formats.

Thus, it was found that the students of the EG coped with the material with the following results: excellent (26%) and good (43%) level of knowledge (in total 69%), while the indicators of the students of the CG were slightly lower: «excellent» – 20%, «good» – 27%, «satisfactory» – 48%, 2 students received unsatisfactory grades (12%). Below we will report how particular CCC criteria were considered and approached within the experiment.

Paraxeological Criterion

Regarding the *paraxeological* criterion, sessions based on role plays and problem solving through case studies were employed during the learning process (upon each unit completion) and then during the assessment session, which, in turn, was video recorded to provide for later evaluation by three independent experts. The task rested on most common gaffes and cultural misunderstandings.

Grades 0-3 were given for each indicator: 1) communicative task achievement by resolving the conflict or sorting the issue; 2) responding to the interlocutor's input; 3) taking into account the interlocutor's culture. The grades contributed to the students' overall assessment and level allocation.

Epistemological and Praxeological Criteria

Besides, in order to identify the characteristics of students' CCC according to *epistemological* and *praxeological* criteria, an expert assessment with the identification of indices (Hovanov et al., 2009), was used. It allowed to determine the level of communicative-valuable knowledge formation, skills and behavioral characteristics of students in both groups. Each relevant parameter was offered in 5 options in order to determine the degree of its severity in the student (from 1 to 5 points). The index was determined by the formula: i = Ver / 4, where «Ver» is the average score that the student received for a specific parameter, while the indices were correlated with the levels of formation of communicative competence in the following way: elementary – from 0,25 to 0,5; limited – from 0,51 to 0,75; sufficient – from 0,76 to 1; advanced – from 1,01 to 1,25. Conducting an expert assessment meant analyzing the outcomes and classroom observation of students' the practical activities during all classes. When analyzing the results, it was found that, compared with the CG, the students of the EG to a greater extent, demonstrate confidence in solving communication problems. They definitely possess the ability to establish contacts, both with the interlocutor and with the audience as a whole, to argue and logically build their opinion, while selecting appropriate speech means and responding to questions.

However, it is necessary to emphasize the fact that the use of nonverbal and paraverbal strategies by students, as well as the ability to change communication tactics in accordance with the communicative situation, leaves much to be desired. The reason for this is understood to be in the lack of communication practice. In this regard, personal communication is based on the achieved results. For example, according to the results of the study, out of the total number of students in both groups, only 7 people, which is about 10%, were able to achieve a sufficiently high level of competence formation according to the paraxeological criterion.

Motivational and Axiological Criteria

Speaking of the motivational and axiological criteria of the CCC, the complex of assessment procedures also included a multiple choice questionnaire designed in accordance with the R. Cattell's factor personality questionnaire (Cattell&-Cattell, 1995), which was aimed at determining the motives and value orientations of students in the field under consid-

eration. The answer to the question posed in the questionnaire about the motives for CCC development among the students of the EG was «to become a professional specialist able to communicate with people from different countries», while in the CG this answer was chosen by only 28% of respondents. The dominant motives of EG students were: «to ensure a prosperous future in the field of professional activity» (83%), «to develop as a person» (70%). It is possible to conclude that EG students are focused on a personally significant result, which in turn indicates the coincidence of educational and cognitive activity with the pedagogical goal. When considering the prevailing motives of the students of the CG, the following answers were obtained: «meeting the learning requirements» and «successful study» (76% each), «avoiding unpleasant consequences» (64%), which indicate that the students of the CG are more focused on solving current problems and avoiding difficulties.

Almost all students who are members of the EG (92%) are aware of the importance of cross-cultural communication competence, both in their future profession and in later life in general, preferring tolerant behavior and a cooperative style of communication. The same opinion is shared by about 60% of the CG students.

The results obtained during the experimental part of the research allowed to note a slight change in the motivational-axiological component of intercultural communicative competence of students in the CG, a positive trend which was only 18%, while in the EG, this figure stood at 40%.

Analysis

In order to clarify the characteristics of the students ' communicative competence according to the reflexive criterion, it was necessary to check the results of the delayed subjective self-assessment in terms of cross-cultural literacy and evaluating students' "own perceptions of the value systems of different cultures" (Skopinskaja, 2009). Some of the proposed assessment tools here would be the so-called "culture logs" or portfolios where students record their progress in CCC at regular intervals and "portfolios" which are used to keep records (Lussier et al., 2007). Reflexive essays, personal observations, self-audio and video recordings were employed in both groups and assessed based on evaluation grids.

The results of the experiment showed that the majority of students of the EG (65%) had adequate self-assessment, for 5 students (29%) it was quite low and in 3 people (18%) – much too high. Among the students of the CG, 12 people (71%) rated themselves fairly adequately, 3 (18%) became the owners of low self-esteem, and for 10 (59%) it was overestimated. It should be emphasized that only an adequate self-assessment indicates that the student has a reflexive attitude of students to their own communicative behavior and the result of educational and cognitive activity. A comparison of the results obtained with the expert assessment showed that the most successful students, which is about 30% in the EG and 16% in the CG, demonstrate a demanding attitude towards themselves as a subject of communication, but this indicator also demonstrates an increase in their own competence of these respondents. Generalization, analysis and comparison of empirical data obtained during the entire experiment revealed the achieved level of each student's communicative competence. The comparison of the initial and final levels was the basis for identifying the dynamics of the CCC formation among students (table 4).

After all, the results of all the tests were thoroughly analyzed and the average statistics were derived for both the control and experimental groups. The dynamics of the outcomes are presented in Table 4 Dynamics of the cross-communicative competence formation among students.

Thus, the results of the final test revealed a significant positive trend in the distribution of EG students according to the level of formation of intercultural communicative competence. The overall growth of all indicators in the CG indicates that the delivery of the State Educational Standard makes it possible to form the communicative competence in students.

Summing up on the overall results of the experiment, a comparative analysis of the findings of the input and final diagnostics achieved in both the experimental and control groups, give the basis to conclude that the developed model of the intercultural communicative competence formation in students is effectively implemented in practice.

To address the research questions about the correlation of the CCC and overall command of English skills, we conducted a language test aimed at assessing students' linguistic progress. The results demonstrated by the students of both groups, united by didactic units, are presented in Table 3 - Results of testing the students' learning gains in the English course.

The tasks used in the test were arranged in such a way as to identify the degree of the material acquisition by students. At the same time, the difference was only in the strategy of working with it, for example, a significant emphasis was placed on students' self-control in the EG. When comparing the results of this test, it was found that the students of the EG were much better able to master a number of program sections compared to the students of the CG, which is especially valuable in the field of solving problems of future professional activity.

It should be noted that the average score derived from the test in the EG was significantly higher than in the CG (83,5 and 68.0, respectively), which also indicates a greater efficiency of the proposed technology.

Through classroom observations, it was possible to report that those students who had high scores possessed a better-formed ability to organize their own educational activities for mastering their communicative competence. Quite indicative was the fact that all the students of the EG coped with the task within the allocated time period, while nine students of the CG did not succeed, which once again emphasized the lack of self-organization skills. Thus, the targeted work carried out by students on developing self-control during the learning process and the learning outcomes have a positive effect on their overall academic success in the EFL course.

DISCUSSION

The paper sought to identify the key structural elements within the CCC frameworks in order to devise the curriculum so that it could focus on each of them. A number of similar models previously offered by other authors have been analyzed. So, borrowing the main ideas and principles of the CCC structural frameworks (Skopinskaja, 2009; COE, 2020; Zamroni et al., 2021) we proposed our own model which partially coincides with the previously offered ones but is still different in terms of the elements identified and their values recalibrated. Thus, the epistemological criterion we identified can be considered an equivalent of the "saviors" (Byram & Zarate, 1996), cultural awareness (Zamroni, 2021), motivational and cognitive criteria (Chernova, 2018). We, however, introduced the new, praxeological, criterion to be added to the CCC structural model driven by the same reason, i.e. to make it more transparent in terms of further measuring the level of CCC development in the classroom.

Besides, the study surveyed the way a training course may incorporate CCC development and investigate its correlation with the degree of general academic performance at mastering a foreign language. To our knowledge, it was the first attempt to trace the correlation between the level of CCC formed and fluency in a foreign language, the previous ones restricted themselves by evaluating just one of the CCC element (Arasaratnam & Doerfel, 2005; Arasaratnam & Banerjee, 2011).

It was revealed that the multi-sided formation of different CCC components yields quite good results. The findings obtained strongly support the hypothetical assumptions of the positive impact CCC development has on acquisition of a foreign language. It can be concluded that the conventional way of building CCC does not compare to the targeted development of its key criteria. So, both hypotheses were proved. To summarize, the customized course content and assessment tools were implemented and adjusted to the particular needs, that is, for evaluating individual aspects

Table 4

Dynamics of the Cross-Cultural Communicative Competence Development among Students, %

Groups	The control group			The experimental group		
Levels	The ascertaining stage	Control stage	Dynamics of the level	The ascertaining stage	Control stage	Dynamics of the level
1	2	3	4	5	6	7
Elementary	31	17	- 14	32	8	- 24
Limited	48	45	+ 3	44	33	-11
Sufficient	21	34	+ 13	24	42	+ 18
Advanced	0	4	+ 4	0	17	+ 17

Table 3

Results of Testing the Students' Learning Gains in the English Course, %

Group Didactic units	CG	EG	Results compared
Vocabulary	74	72	-2
Grammar	59	64	+5
Speech etiquette	68	91	+23
Writing	67	82	+15
Reading	58	89	+31
Listening comprehension	61	69	+8

which bring about students' higher performance, in particular, in speech etiquette, writing and reading.

The advantages of the proposed methodology are demonstrated by establishing the relationship between the development of students' abilities, increased curiosity desire and ability to learn, as well as fostering moral qualities and values necessary for effective interaction with representatives of other cultures and nationalities. The results gathered do not contradict those obtained from studies that have been done by the other authors (Cavalheiro, 2015; Collier, 2015, Tzheva, 2021). The study has also revealed the way CCC can be used as a tool to enhance the linguistic skills and abilities as it made an attempt to investigate the correlation between the two, whereas the previous papers primarily focused on solely measuring the overall degree of its development (Deardorff, 2006b; Skopinskaja, 2009; Zamroni, 2021) or examined one of its components (Arasaratnam & Doerfel, 2005; Arasaratnam & Banerjee, 2011).

Though the authors countered some of the limitations in the previous studies on both building the CCC (or its analogy) framework and assessing its level in students studying a foreign language, it definitely has a number of constraints as well. So, the size of the research population speaks for the very preliminary conclusions to be made and should be significantly increased in the following studies. Though claimed above, the study also lacked profound comparison of evaluation and self-evaluation matrices in order to provide a fuller picture of the participants' outcomes and find out lurking indicators. Mismatches between different experts' evaluations and lack of rigorous unified assessment criteria are also likely to result in biased conclusions.

CONCLUSION

The given research pursued the aim of identifying the CCC components (pre-active stage), adhering the curriculum in the experimental group to focus on targeting them all while teaching English as a foreign language (active stage) and analyzing the correlation between the level of its formation and students' overall communicative competence.

The research was conducted by using a number of different techniques and tools. Thus, having examined the ways intercultural communicative competence is formed and the impact it has on a foreign language acquisition, we have concluded what exactly contributes to developing cross-cultural communicative competence among students.

In turn, the developed system of criteria makes it possible to identify the degree of competency level. The principles of building cross-cultural awareness determine the success of teaching intercultural communication. Its key principles, directions, stages and means of implementation turn out to be effective in terms of obtaining the necessary knowledge by students, as well as mastering skills in the professional field.

Thus, cross-cultural communication forms special conditions for the development of not only each individual person, but also national-cultural groups that live in a diverse world. It should be recognized that CCC as a way of organizing society leads to the need to rethink the cultural integrity of its members, to abandon the notion that certain cultures can be stronger, dominant or less advanced. All this, undoubtedly, has a positive effect on the interaction of representatives of different ethnic backgrounds in all areas of human life.

IT should also be born in mind that, based on the provisional conclusion this study has arrived at, raising the CCC level can possibly be an effective tool in boosting the overall command of the foreign language the students are acquiring. Presumably, it can be explained by the intrinsic motivation growth, which is, in turn, caused by exposure to cultural reality of international communities, the so-called "feeling of authenticity".

The results of the study do not solve all the problematic issues of the formation of cross-cultural (intercultural) communicative competence in students, but they can serve as a basis for organizing further investigations. In the framework of future research, it is planned to identify correlations and their meanings between the established components of the intercultural communicative competence.

ACKNOWLEDGEMENT

This paper has been supported by the RUDN University Strategic Academic Leadership Programme.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Natalia Guskova: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing-review and editing.

Elena Golubovskaya: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing-review and editing.

REFERENCES

Allen, B. J. (2010). *Difference matters: Communicating social identity.* Waveland Press. https://doi.org/10.1108/09649420410529898

Andreyeva, O. A., Tuleubayeva, Sh. K., Ganyukova, A. A., & Tentekbayeva, Zh. M. (2015). Cross-cultural communication as a way of achievement of cross-cultural communicative competence. *European Researcher, 92*(3), 208-213. https://doi.org/10.13187/er.2015.92.208

Arasaratnam, L. A., & Banerjee, S. C. (2011) Sensation seeking and intercultural communication competence: A model test. *International Journal of Intercultural Relations,* 35, 226–233.

Arasaratnam, L. A., & Doerfel, M. L. (2005) Intercultural communication competence: Identifying key components from multicultural perspectives. *International Journal of Intercultural Relations, 29*, 137–163.https://doi.org/10.1016/j.ijintrel.2004.04.001

Aririguzoh, S. (2022). Communication competencies, culture and SDGs: Effective processes to cross-cultural communication. *Humanities and Social Sciences Communications, 9*(96). https://doi.org/10.1057/s41599-022-01109-4

Bagiyan, A., Shiryaeva, T., Tikhonova, E., & Mekeko, N. (2021). The real value of words: How target language linguistic modelling of foreign language teaching content shapes students' professional identity. *Heliyon, 7*(3), e06581. https://doi.org/10.1016/j.heliyon.2021.e06581

Barzykowski, K., Majda, A., Przyłęcki, P., & Szkupl, M. (2019). The cross-cultural competence inventory: Validity and psychometric properties of the Polish adaptation. *PLOS ONE 14*(3), e0212730. https://doi.org/10.1371/journal.pone.0212730

Bennett, M. J. (2009). Defining measuring and facilitating intercultural learning: A conceptual introduction to the intercultural education double supplement. *Intercultural Education, 20*(1), 1-13. https://doi.org/10.1080/14675980903370763

Bickley, C., Rossiter, M. J., & Abbott, M. L. (2014). Intercultural communicative competence: Beliefs and practices of adult English as a second language instructors. *Alberta Journal of Educational Research*, *60*(1), 135-160. https://doi/org.10.20414/edulangue.v4i1.3044

Bleszynska, K. M. (2008). Constructing intercultural education. *Intercultural Education, 19*(6), 537-545. https://doi.org/10.1080/14675980802568335

Brown, D. H., & Terrell, T. D. (2000). *Principles of language learning and teaching*. Pearson Education. https://doi.org/10.2307/414380

Byram, M., Nichols, A., & Stevens, D. (2001). developing intercultural competence in practice. *Languages for Intercultural Communication and Education* (pp. 282-284). Multilingual Matters. https://doi.org/10.21832/9781853595356-021

Byram, M., & Zarate, G. (1996). Defining and assessing intercultural competence: Some principles and proposals for the European context. *Language Teaching*, 29(4), 239-243. https://doi.org/10.1017/S0261444800008557

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Cattell, R. B., Cattell H. E. P. (1995) Personality structure and the new fifth edition of the 16 PF. *Educational and Psycological Measurement, 55*(6), 926-937. https://doi.org/10.1177/0013164495055006002

Chernova, N. I. (2018). Cross-cultural communicative competence of specialists at the innovative university cluster in the logic of their language training quality improvement. *Advances in Economics, Business and Management Research*, 47. *Proceedings of the International Scientific Conference "Far East Con" ISCFEC 2018*. Atlantis Press. https://doi.org/10.2991/iscfec-18.2019.97

Collier, M. J. (2015). Intercultural communication competence: Continuing challenges and critical directions. *International Journal of Intercultural Relations, 48*, 9-11. https://doi.org/10.1016/j.ijintrel.2015.03.003

Danilova, M.V., Rykman, L.V. (2018) Psychoemotional well-being and features of self-development of teenagers with different family status. *Psycological Science and Education, 23*(5), 40-50. https://doi.org/10.17759/pse.2018230505

Deardorff, D. K. (2006a). Policy paper on intercultural competence. Intercultural competence – The key competence in the 21st century? *Theses by the Bertelsmann Stiftung based on the models of intercultural competence by Dr. Darla Deardorff* (pp. 12-30). Bertelsmann Stiftung.

Deardorff, D. K. (2006b). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education, 10*(3), 241–266. https://doi.org/10.1177/1028315306287002

East, M., Tolosa, C., Howard J., Biebricher, C., & Scott, A. (2022). Journeys *towards intercultural capability in language classrooms*. Springer. https://doi.org/10.1007/978-981-19-0991-7

Gay, G. (2001). Curriculum theory and multicultural education. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural education* (pp. 25-43). Jossey-Bass Publishers. https://doi.org/10.4135/9781412958806.n316

Gierke, L., Binder, N., Heckmann, M, Odağ, O., Leiser, A., & Kedzior, K. K. (2018). *Definition of intercultural competence. Undergraduate students at a private university in the USA: A mixed-methods study. PLoS ONE, 13*(4), e0196531. https://doi.org/10.1371/journal. pone.0196531

Hoff, H. E. (2014) A critical discussion of Byram's model of intercultural communicative competence in the light of bildung theories. *Intercultural Education*, *25*(6), 508-517. https://doi.org/10.1080/14675986.2014.992112

Hovanov, N., Yudaeva, M., & Hovanov, K. (2009). Multicriteria estimation of probabilities on basis of expert non-numeric, non-exact and non-complete knowledge. *European Journal of Operational Research, 195*(3), 857–863. https://doi.org/10.1016/j.ejor.2007.11.018

Johnson, J., Lenartowicz, T. & Apud, S. (2006). Cross-cultural competence in international business: toward a definition and a model. *Journal of International Business Studies*, 37, 525–543. https://doi.org/10.1057/palgrave.jibs.8400205

Kramsch, C. (2008). Culture in foreign language teaching. *Iranian Journal of Language Teaching Research, 1*(1), 57-78. https://doi.org/10.2307/j.ctv62hgw3.4

Liao, Y., Thomas, D.C. (2020). Cross-Cultural competence. In S. A. Houghton & M. Porto (Eds.), *Cultural intelligence in the world of work. springer series in emerging cultural perspectives in work, organizational, and personnel studies*. Springer. https://doi.org/10.1007/978-3-030-18171-0_1

Lussier, D., Ivanus, D., Chavdarova-Kostova, S., Golubina, K., Skopinskaja, L., Wiesinger, S., & de la Maa Retamar, G. (2007). *Guidelines for the Assessment of Intercultural Communicative Competence*. In I Lazar, M. Hubert-Kriegler, D. Lussier, G. Matei, & C. Peck (Eds.), *Developing and assessing intercultural communicative competence: A guide for language teachers and teacher educators* (pp. 23-39). European Centre for Modern Languages and Council of Europe Publishing.

Mallia, J. (2015). Embedding grammar while developing communicative competence in English: Relevant cultural contexts and teaching approaches. *Arab World English Journal, 6*(1), 50-67. https://doi.org/10.24093/awej/vol6no1.4

Martin, J. N., Nakayama, T. K. (2015) Reconsidering intercultural (communication) competence in the workplace: A dialectical approach. *Language and Intercultural Communication*, *15*(1), 13-28, https://doi.org/10.1080/14708477.2014.985303

Marxsen, P. (2007). Intercultural student teaching: A bridge to global competence. *Education and Culture* (pp. 80-81). Rowman & Littleton Education. https://doi.org/10.1353/eac.0.0005

Merrifield, M., Lo, J., Sum, P., & Kasai, M. (2008). Worldmindedness: Taking off the blinders. *Journal of Curriculum and Instruction, 2*(1), 6-20. https://doi.org/10.3776/joci.2008.v2n1p6-20

Munezane, Y. (2021). A new model of intercultural communicative competence: Bridging language classrooms and intercultural communicative contexts. *Studies in Higher Education, 46*(2), 1-18. https://10.1080/03075079.2019.1698537

O'Brien, T. (2019). Bringing cross-cultural communication analysis into foreign language classrooms. *Intercultural Foreign Language Teaching and Learning in Higher Education Contexts* (pp. 71-97). Carleton University. https://doi.org/10.4018/978-1-5225-8128-4.ch004

Peart, T. (2019). Cross-cultural communication differences in online learning. In J. Keengwe & K. Kungu (Eds.), *Handbook of research on cross-cultural online learning in higher education* (pp. 159-171). Louisiana State University. https://doi.org/10.4018/978-1-5225-8286-1.ch009

Piasecka, L. (2011). Sensitizing foreign language learners to cultural diversity through developing intercultural communicative competence. In J. Arabski & D. Wojtaszek (Eds.), *Aspects of culture in second language acquisition and foreign language learning* (pp. 21-33). Yonkers. https://doi.org/10.1007/978-3-642-20201-8_3

Rasmussen, L.J., Sieck, W.R. (2015). Culture-general competence: Evidence from a cognitive field study of professionals who work in many cultures. *International Journal of Intercultural Relations, 14*(3), 75-90. https://doi.org/10.1016/j.ijintrel.2015.03.014

Schaetti, B. F., Ramsey, Sh. J., & Watanabe, G. C. (2009). From intercultural knowledge to intercultural competence: Developing an intercultural practice. In M. A. Moodian (Ed.), *Contemporary leadership and intercultural competence* (pp. 125-138). Sage Publications. https://doi.org/10.4135/9781452274942.n10

Sehlaoui, A. S. (2001). Developing cross-cultural communicative competence in pre-service ESL/EFL teachers: A critical perspective. *Language, Culture and Curriculum, 14*(1), 42-57. https://doi.org/10.1080/07908310108666611

Skopinskaja, L. (2009). Assessing intercultural communicative competence. *Synergies Pays Riverains de la Baltique*, (6), 135-144.

Spitzberg B. H., (2009). Changnon G. Conceptualizing intercultural competence. In D. Deardorff (Ed.), *The SAGE handbook of intercultural competence* (p. 2–52). SAGE.

Starkey, H. (2007) Language Education, identities and citizenship: Developing cosmopolitan perspectives. *Language and Intercultural Communication*, *7*(1), 56-71. http://doi.org/10.2167/laic197.0

Tazheva, Z., Sycheva, I., & Artyukhova, N. (2021). Formation of cross-cultural communicative competence among foreign students during pre-university training, *Propósitos y Representaciones, 9*(SPE2), e1024. https://doi.org/10.20511/pyr2021.v9nSPE2.1024

Tran, T. Q., & Duong, T. M. (2018). The effectiveness of the intercultural language communicative teaching model for EFL learners. *Asian-Pacific Journal of Second and Foreign Language Education, 3*(1). https://doi.org/10.1186/s40862-018-0048-0

Tikhonova, E. V., Kosycheva, M. A., & Efremova, G. I. (2021). Primary professionalization of foreign students: Barriers, stigmatization, adaptation. *Integratsiya obrazovaniya = Integration of Education*, 25(4), 608-628. https://doi.org/10.15507/1991-9468.105.025.202104.608-628

Zamroni, Dwiningrum, S. I. A., Hope, J., Kartowagiran, B., Sudartinah, T., Siteine, A., & Yao, Z. (2021). Cross-Cultural competence in multicultural education in Indonesian and New Zealand high schools. *International Journal of Instruction, 14*(3), 597-612. https://doi.org/10.29333/iji.2021.14335a

https://doi.org/10.17323/jle.2023.12359

Technology-Enabled Language Leaning: Mediating Role of Collaborative Learning

Divya Jyot Kaur [1], Niraja Saraswat [2], Irum Alvi [3]

[1] Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, Rajasthan, India

[2] Malaviya National Institute of Technology, Jaipur, India

[3] Rajasthan Technical University, Kota, India

ABSTRACT

Background. Technology-Enabled Language learning (TELL) encourages peer communication and collaboration through its innovative instructional methods. Collaborative student activities are recognised as an important component of the instructional approach of higher education, More recently, collaborative learning in conjunction with digital teaching tools has emerged as a preferred SLA pedagogical approach. Despite growing interest in TELL, research into the effects of collaborative learning on affective factors in SLA remains unexplored.

Purpose. The aim of the proposed study is to identify factors influencing the behavioral intention of students to use WhatsApp for second language acquisition. Constructs from previous models: performance expectancy, effort expectancy, and perceived relevance (UTAUT) are tested, along with the mediating role of a new variable 'collaborative learning'.

Method. Using the convenience sampling technique, the sample comprises 202 undergraduates studying in Institutes in Rajasthan, India. Data collected through Google forms was analyzed through IBM SPSS ver. 26 and Smart-PLS ver. 3.2.9, using structural equation modeling.

Results. A positive and significant relationship was established between all the selected constructs. The indirect effects were positive, yet less significant than the direct effects. Moreover, the partially mediating effect of collaborating learning was affirmed. Empirical data confirms that collaborative learning acts as a mediating variable enhancing the intention to use WhatsApp for SLA.

Conclusion. The present study makes an original and innovative contribution to language studies by analysing the relationship between the predictors. Such a systematic understanding of the topic can assist instructors in designing robust future pedagogical techniques.

KEYWORDS

UTAUT, technology, WhatsApp, second language acquisition, collaborative learning, mediation

Citation: Kaur D. J., Saraswat N., & Alvi I. (2023). Technology-enabled language leaning: Mediating role of collaborative learning. *Journal of Language and Education, 9*(1), 89-101. https://doi.org/10.17323/jle.2023.12359

Correspondence:
Divya Jyot Kaur,
divyajyotdjk@gmail.com

Received: May 01, 2021
Accepted: Mach 15, 2023
Published: March 31, 2023

INTRODUCTION

In the 21st century, the concept of technology has introduced new connotations, changing lives and transforming the way people communicate with one another. Due to COVID-19, online education superseded traditional classroom instruction as the preferred method of instruction, and this shift in the mode of education became one of the most significant academic changes (Sharma & Alvi, 2021). This has caused a spike in education programmes offered online and a need for better articulation of technology and pedagogy in higher education (Divjak et al., 2022).

Since learners today are digital natives, technology-enabled language learning allows them to learn as per their interest, and it also appeals to both the optical and the acoustic perceptions of learners (Kaur et al., 2021). Consequently, educators have worked to incorporate new technologies that engage students in active and creative learning rather than being passive knowledge beneficiaries (Chen Hsieh, 2022). Given the global situation during the COVID-19 pandemic, educators began to investigate methods

of collaborative work among students who were geographically dispersed from one another. Thus, the current educational emphasis has shifted significantly from individual to active collaborative learning (Dashkina et al., 2022)

Active collaborative learning is a continuous process in which students communicate and share views and ideas through social media Alismaiel et al. (2022). TELL encourages collaborative learning by encouraging students to interact with and learn from their peers, identifying their mistakes, increasing motivation, and being interesting (Zhang et al., 2022). Teachers with digital tools and devices employ different ways of practicing learning activities when compared with those in traditional classrooms. They can use a variety of strategies and practices to successfully implement collaborative learning (Pietarinen et al., 2021).

Easy access to mobile devices, social networking sites (SNSs) which can meet the needs of people, have become essential in our everyday lives. With the rising number of users, some SNSs have grown in popularity and have become common contact networks. As these platforms are synchronous, they allow users to receive as well as provide instructional support in real time (Miguel & Carney, 2022). The preferred app is WhatsApp, referred to as 'a simple social network' (Fischer, 2013). WhatsApp facilitates better communication and collaboration with others in your phone book, either personally or in groups. It has begun to be studied academically because of its ability to affect teaching and learning processes by extending the spatial and temporal borders of the classrooms and allowing teachers to communicate with students anywhere (Fouz-González, 2017). In terms of adaptability, it allows students to easily access online learning, if they have internet access and a mobile device, even when they live in a remote area.

Digital initiatives have now amplified and contributed enormously to pedagogy and the learning ability of students using hybrid pedagogies. They have also enhanced their online competencies. The creation of such planned blended pedagogies by educators has led to a resolution of the uncertainties of virtual learning. However, in Asian English language classrooms, the teacher-centered approach and rote learning are routine, with less focus on interactive student-centered learning in the language-learning process (Fischer & Yang, 2022). Therefore, the current research focuses on students' intentions to adopt TELL with the mediating effect of collaborative learning, in order to better prepare practitioners to use digital tools for collaborative learning in English language classrooms.

Purpose and Research Questions

Given the complex and multifaceted challenges of the transformation to digital learning in India, it was fairly obvious that the need for new solutions to optimise educational activities has increased. With the growing importance of higher inclusive education in today's world and the scarcity of studies on adoption and intent to use TELL, the findings of this study can help with the implementation of strategic pedagogical planning in this direction at the national and regional levels. The proposed study is one of few to modify and extend the constructs of UTAUT (Venkatesh *et al*., 2003) which assess the mediating role of collaborative learning in acceptance of technology (WhatsApp) for second language acquisition. The study is further distinguished by the addition of a new variable: collaborative learning. Collaborative learning is an emerging topic for educational studies. The concept is premised in the idea that learning occurs as a result of interactions between people and their surroundings. This social process serves as the foundation for collaborative learning. In this regard, the online world, as an interaction space, is a natural and dynamic learning domain (Herrera-Pavo, 2021). The research questions for this study are as follows:

(1) Do the selected constructs have a direct and affirmative impact on collaborative learning as a mediating variable in acceptance of WhatsApp for second language acquisition?

(2) Do the selected constructs have an indirect and affirmative effect on behavioral intention towards the use of WhatsApp for second language acquisition?

(3) Do the selected constructs have an indirect and affirmative effect on user intention towards the use of WhatsApp for second language acquisition?

LITERATURE REVIEW

Technology-Enabled Language Learning

The teaching of the English language itself has experienced a remarkable transformation in the last decade in particular (Mofareh, 2019). In the wake of the pandemic, there have been extraordinary transformations across various fields, so it is indispensable that English language teaching keeps pace with the global technological revolution, in order to continue training in a systematic and advanced way (Ko & Lim, 2022). The diverse nature of educational technology makes TELL a versatile discipline. It supersedes conventional concepts & rules and provides a creative learning environment, reflected in the practices that extend ELT pedagogy (Sharifi et al., 2018). Today's students are tech-savvy and multi-taskers, to be more open to innovative and stimulating teaching styles that incorporate cutting-edge technology in the classroom (Fischer & Yang, 2022). Traditional lecturing models, for example, which attempt to cram dozens of students into one class despite the economic gains, are no longer acceptable to digital-age Gen Z students (Sixto-García & Duarte-Melo, 2020).

Several studies have highlighted the advantage of technology for English language learning. The purpose is not only to keep learners engaged but also to encourage the motivation of learners and learner-centered education (Zhang et al., 2022). The learner is able to direct their own educational plan, construct meaning, and monitor and improve their own work (Dashkina et al., 2022). Research conducted (Sun & Gao, 2019) to explore the relationships between motivation, technology adoption factors, and behavioral intention indicate that adequate instructional design compatible with and promoting the mission of language learning is necessary, in order to enhance students' behavioral intention to make use of technology for language learning.

Social Networking Sites (SNS)

The more educational content developed through already widely used SNSs, the greater their potential to influence education. As a result, they need to be incorporated into education, in order to realise their full potential. These platforms are synchronous, allowing participants to receive and provide instruction and feedback in real time (Miguel & Carney, 2022). Since online education during the pandemic has become the 'new normal', students are bored of traditional forms of learning and are open to new and innovative learning styles. Therefore, SNSs offer new possibilities for the modern requirements of the new generation (Toyama & Yamazaki, 2021).

Several studies have shown that SNSs promote educational programs by facilitating engagement, teamwork, constructive involvement, knowledge and data sharing, as well as critical thinking (Çetinkaya & Sütçü, 2018). The evolutionary interest of social networking and social interaction is an important aspect of digital technology. People now communicate globally in synchronous and asynchronous environments via mobile applications. This would have been just a vision a few years ago. Social networks are referred to as 'affinity space' in which individuals learn both social and communicative skills (Mellati et al., 2018).

WhatsApp and Collaborative Learning

WhatsApp was created for smartphones and tablets and seeks to facilitate general connectivity, provide image, audio, and video sharing, and a host of helpful features such as community formation (Cetinkaya & Sutcu, 2019). In particular, their great effect on young people's social development makes it important to assess their effect on scholastic development and their aspirations (Kaur et al., 2021). WhatsApp has become the most commonly used platform to strengthen and promote accessibility, enable collaboration, and motivate learners to engage actively in classroom activities (Soria et al., 2020). According to the findings of researches on WhatsApp for educational purposes, it has been shown to promote learning at anytime and anywhere, raise engagement and inspiration, provide teamwork, facilitate learning, boost learner interaction and student-teacher interaction (Smit, 2012; Rambe & Chipunza, 2013; Plana et al., 2013; Church & De Oliveira, 2013; Jafari & Chalak, 2016; Rambe & Bere, 2013; Mistar & Embi, 2016; Güler, 2017).

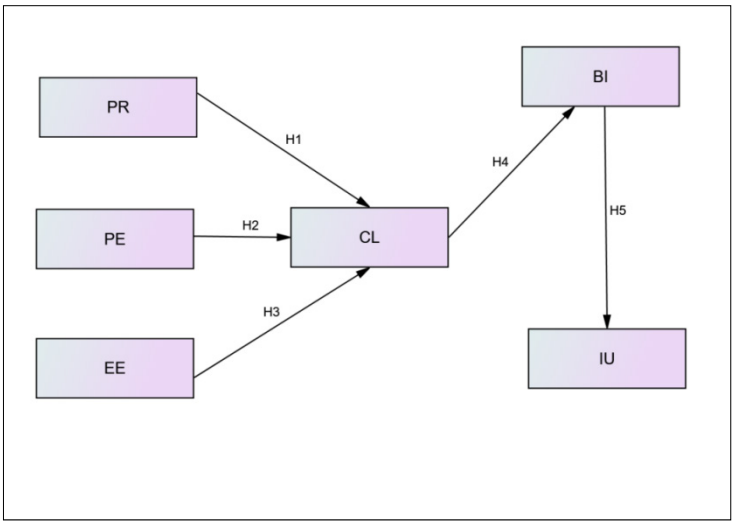
Bui et al. (2021) provides the broadest definition of active collaborative learning, defining it as a situation in which two or more individuals study or seek to learn something new together. The core aspects of collaborative learning are interpersonal trust, individual responsibility, promotional engagement, interpersonal abilities, and group interaction (Lee & Yang, 2020). Collaborative learning with the use of technology has seen a lot of exposure in recent years. The ability to gather learners together using diverse platforms has created a lot of opportunities for events that take advantage of these technologies (James et al., 2022). Learning through social collaboration has been shown to improve students' academic performance and to reduce students' feelings of isolation by creating a more positive interactive online learning environment (Stöhr et al., 2020). With the advancement of online technology, educators are beginning to incorporate newer technological tools, in order to enhance more effective peer-to-peer/online synchronous/collaborative learning activities to ensure that students diligently acquire knowledge (Fischer & Yang, 2022).

Understanding the factors that promote student collaboration is critical to understanding how this approach to learning can be used more effectively in higher education online courses (Haugland et al., 2022). The research aims to establish a succinct conceptual model for evaluating the mediating role of collaborative learning in the acceptance of technology (WhatsApp) for second language acquisition (SLA).

The UTAUT Model

The Unified Theory of Acceptance & Use of Technology (UTAUT) analyses and integrates eight models of technology adoption, explaining users' behavioral intention towards the use of technology. The UTAUT includes TAM, theory of reasoned action, social cognitive theory, theory of planned behavior (TPB), motivational model, innovation diffusion theory, model of personal computer utilisation, and TAM and TPB combined (Botero et al., 2018). UTAUT initially aimed to explain the indicators affecting the adoption of employee information technology. Nevertheless, several researchers have applied it to the sense of education (Escobar-Rodríguez et al., 2013). The model is used to calculate the acceptance of blogs and wikis (Yueh *et al*., 2015), Facebook (Escobar-Rodríguez et al., 2014), and podcasts (Avci & Askar, 2012). In the educational context, UTAUT is also used for assessing mobile learning (Abu-Al-Aish & Love, 2013) and assess m-learning in emergent countries (Iqbal & Qureshi, 2012). In order to consider the intervention from an ethnographic standpoint, it is important to know whether or not people use the device and how they interact with the new technologies.

Figure 1
Conceptual Model



As shown in figure 1, the conceptual model is based on the UTAUT model by Venkatesh et al. (2003). It has been modified to meet the purpose of the present study, and a new moderating variable namely, collaborative learning, was introduced into the model.

Perceived Relevance (PR)

PR is «the degree to which consumers perceive an object to be self-related or in some way instrumental to achieving their personal goals and values» (Celsi & Olson, 1988, p. 2011). It is the level to which one assumes that a method can help execute a job differently, quicker, and with precision, effectiveness, and reliability (Kaur et al., 2021). Perceived relevance has piqued the interest of researchers from a variety of technical areas (González-Ibáñez et al., 2016). Accordingly, these hypotheses were anticipated:

H1: PR has a direct affirmative effect on CL.

H1a: PR has an indirect and affirmative effect on BI.

H1b: PR has an indirect and affirmative effect on the IU.

Performance Expectancy (PE)

PE is the magnitude to which people think that using a certain device will help to improve their job outcome (Venkatesh et al., 2003). Perceived expectancy from TAM and TAM 2, motivational model's extrinsic incentive, job fit via the model of personal computer use, relative benefit from the innovation diffusion theory, and social cognitive theory's outcome expectation: all contributing to performance expectancy. Including the fact that a few studies have found no connection between behavioral intention and performance expectancy (Iqbal & Qureshi, 2012; Yueh et al., 2015; Acharya & Ganesan, 2019), many studies identify the significance of performance expectancy for assessment of technology for education (Martinho et al., 2018; Scherer et al., 2019). The strongest predictor of intention is performance expectancy (Alismaiel et al., 2022). Accordingly, these hypotheses were proposed:

H2: PE has a direct and affirmative effect on CL.

H2a: PE has an indirect and affirmative effect on BI.

H2b: PE has an indirect and affirmative effect on the IU.

Effort Expectancy (EE)

EE is the "degree of ease associated with the use of the system" (Botero et al., 2018, p. 5). According to Venkatesh et al. (2003), effort expectancy is important near the beginning stages of prolonged and continuous technology use, and becomes negligible over time. The initial technology acceptance model has this expanded definition of perceived user-friendliness. Rahmi & Birgoren (2020) believe EE is imperative in the milieu of digital learning, and its ease of use determines its acceptance. So, the anticipated hypotheses were:

H3: EE has a direct and affirmative effect on CL.

H3a: EE has an indirect and affirmative effect on BI.

H3b: EE has an indirect and affirmative effect on IU.

Collaborative Learning (CL)

CL consists of student partnerships that help complete learning activities (Zhampeissova et al., 2020). Students collaborate in groups of two or more, looking for under-

standing, answers, and definitions, or making products in a collaborative study (Yueh et al., 2015). Learning, according to Lave and Wenger (1994), is not only a single person's approach to acquiring experience but an interconnected practice that occurs within a group of people. Consequently, the following hypotheses were proposed:

H4: CL has a direct and affirmative effect on BI.

H4a: CL has an indirect and affirmative effect on IU.

Behavioral Intention (BI) & Intention to Use (IU)

This is a metric used to determine how strong a person's desire to carry out a specific action is (Dwivedi et al., 2019). Intention to use is the desire to perform a particular action (Bagozzi, 1981). Theoretically, BI has a substantial affirmative effect on technology use. This connection has been proven in studies involving a variety of educational technology. Therefore, the following hypotheses were proposed:

H5: BI has a direct and affirmative effect on IU.

METHOD

Participants

By using convenience sampling, the study sample comprises two hundred and two (N=202) undergraduate students studying in Institutes in Rajasthan, since college-going students make the maximum use of mobile phones. It is necessary to gauge their acceptance of technology for second language acquisition at this initial stage, in order to ensure better learning opportunities.

Instrument

The research instrument used for the survey is the questionnaire developed by the researcher. The questionnaire comprises two parts. Part A dealt with the demographic information such as student's age, gender, the field of study, etc. Part B comprised questions adopted from the modified UTAUT model. The UTAUT questions in Part B were subdivided into six subcategories. They were: effort expectancy, perceived relevance, performance expectancy, collaborative learning, behavioral intention, and intent to use. Responses were elicited from the respondents with the aid of a 5- point Likert scale.

Procedure

A pilot study included a sample of 30, in order to confirm that the instrument was consistent. Feedback was sought from the respondents, and any difficulties in interpretation of the items were removed. The respondents were informed that there was no obligation, and there was no correct or incorrect answer to any of the questions. They were also conversant with the purpose of the study through the description of the Google form used for data collection.

Data was collected online through Google forms during February 2021. The SPSS application (IBM SPSS version 26) was utilised for data analysis. In order to empirically evaluate the proposed model, structural equation modeling was done via Smart-PLS (v.3.2.9). The reliability of data was confirmed through Cronbach's Alpha, AVE, and composite reliability. Discriminant validity was established through the Fornell Larker criterion (1981), cross-loadings, and Hensler criterion.

RESULTS

Demographics

The gender report of the students showed that the maximum number of participants were male undergraduates totaling 146 (72.3%), while female participants were 54 (26.7%). The analysis depicted that 160 (79.2%) of the total 202 participants were of the 18-20 age group. The remaining were above the age of 20. Among the participants, 189 (93.6%) were Bachelor of Technology students. The majority of the respondents totaling 130 (64.4%) claimed their daily mobile usage to be more than 4 hours a day, 26% of them marked it between 2 to 4 hours, and the remaining 10% reported it to be less than 2 hours a day.

Measurement Model

In the current study, PLS-SEM was performed in dual steps. In the first phase the measurement model was evaluated, and the structural model was assessed in the second step. A measurement model evaluates the competency of a scale used for research objectives as depicted in table 1. Cronbach's alpha is the coefficient of internal consistency. It asserts that the instrument's components pertaining to an aspect are internally coherent and can be utilised for the measurement of aspects. An α rating in the range of 0.6 to 0.7 denotes a sufficient level of reliability, whereas a value of 0.8 or larger indicates excellent reliability (Naveed et al., 2020). The values for the subscales ranged from 0.837 to 1.00, indicating that each variable had acceptable internal consistency. Composite reliability must exceed the usual benchmark of 0.70 as per Henseler et al. (2015); for the given constructs, it is above 0.8. AVE is equal to or greater than 0.50 for all constructs which meets the criteria given by Fornell and Larcker (2012). A VIF less than 10 suggests no multicollinearity, and that the expected regression model is accurate and meets the goodness-of-fit criteria (Deraman et al., 2019). Besides, another gauge, rho A (Dijkstra-Henseler's rho), establishes a precise estimate of data consisten-

Table 1
Reliability Measurements

Variable	Items	Variance Inflation Factor	Factor Loadings	Cronbach's alpha	rho_A	Composite Reliability	Average Variance Extracted
BI	BI2	2.075	0.81				
	BI3	2.075	0.888	0.837	0.842	0.839	0.723
CL	CL1	2.188	0.723				
	CL2	2.884	0.827				
	CL3	2.791	0.891				
	CL4	2.611	0.876	0.899	0.905	0.899	0.692
PE	PE1	2.118	0.725				
	PE2	2.746	0.844				
	PE3	2.955	0.738				
	PE4	2.300	0.796				
	PE5	2.239	0.895	0.9	0.904	0.9	0.643
EE	EE1	1	1	1	1	1	1
PR	PR2	1	1	1	1	1	1
IU	IU1	1	1	1	1	1	1

Note. BI=behavioral intention; CL=collaborative learning; EE=effort expectancy; IU=intention to use; PE=performance expectancy; PR=perceived relevance.

cy. Values greater than 0.7 are considered fit for the study (Ramírez & Palos-Sánchez, 2018). The convergent validity of the scale was verified because the results met the stipulated requirements.

As per the Fornell-Larcker criterion, the cross-loadings, and the Heterotrait-monotrait ratio, the discriminant validity is deliberated: (1) The AVE value, according to Fornell and Larcker (1981), must be higher than the correlation measurements of the variables. This criterion was met as the Fornell-Larcker test shown in table 2 depicted that when compared with the other values, the diagonal values are utmost for a definite construct (Naveed et al., 2020). (2) Next, cross-loadings were checked to ensure that each item loads maximum on its related construct. (3) As shown in table 3, HTMT below 0.80 confirmed the validity, as per the set criteria (Henseler et al., 2015). Hence, the scale was deemed reliable and valid for further study.

As observed in Table 4, a Standardised Root Mean Square Residual (SRMR) below 0.1 depicts a satisfactory fit. "This measures the difference between the observed correlation matrix and the correlation matrix implied by the model" (Ramírez & Palos-Sánchez, 2018, p. 13). An NFI value close to one is deemed suitable, and table 4 shows NFI of 0.88 (Botero et al., 2018). The endogenous underlying variables' coefficient of determination (R^2) is essential for interpretation. The R^2 was above 0.2 for CL which is considered moderate, while that for BI and IU was small to moderate (Ramírez & Palos-Sánchez, 2018). The Stone-Geisser test (Q^2) determined that the model was predictive as Q^2 was above 0, confirming that the model's endogenous components have a good extrapolative significance (Geisser, 1974).

The authors employed bootstrapping on 5000 samples using Smart-PLS ver.3.2.9, in order to assess the direct and mediation effects. The bootstrapping protocol is a non-parametric inferential procedure which randomly selects many subsamples (for example, 5,000) from the initial data set and replaces them with new data. It is essential to gather data on the population distribution, which will serve as the foundation for hypothesis testing. Bootstrapping technique is used for drawing conjectures about indirect effects, even when the mediators in the model are complicated (Preacher & Hayes, 2008).

Structural Model

The relationships amongst the constructs were estimated using structural equation modeling. It offers precise estimates of these error variance dimensions, in contrast to standard multivariate methods, incapable of measuring or

Table 2

Fornell–Larcker Discriminant Validity

Variable	BI	CL	EE	IU	PE	PR
BI	0.85					
CL	0.429	0.832				
EE	0.356	0.554	1			
IU	0.370	0.232	0.087	1		
PE	0.366	0.622	0.655	0.107	0.802	
PR	0.331	0.352	0.165	0.167	0.197	1

Note. BI= behavioral intention; CL=collaborative learning; EE= effort expectancy; IU= intention to use; PE= performance expectancy; PR= perceived relevance.

Table 3

HTMT

Variable	BI	CL	EE	IU	PE	PR
BI						
CL	0.431					
EE	0.355	0.552				
IU	0.372	0.237	0.087			
PE	0.364	0.618	0.654	0.105		
PR	0.332	0.351	0.165	0.167	0.194	

Note. BI= behavioral intention; CL=collaborative learning; EE= effort expectancy; IU= intention to use; PE= performance expectancy; PR= perceived relevance.

Table 4

Model Fit

	Saturated Model	Estimated Model	
SRMR	0.039	0.054	
NFI	0.896	0.887	
	R Square	**R Square Adjusted**	**Q^2 (=1-SSE/SSO)**
BI	0.184	0.180	0.110
CL	0.475	0.467	0.305
IU	0.137	0.133	0.112

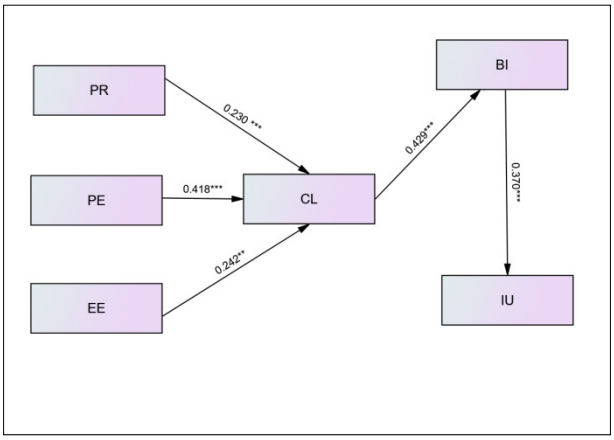
Note. BI=behavioral intention; CL=collaborative learning; IU=intention to use.

accounting for error. In this analysis, direct and indirect effects were calculated. Figure 2 displays the research model.

The study included the dependent variable, IU; the mediating variable, CL; and the three previously mentioned independent variables: PE, PR, and EE. The outcomes of direct, indirect, and total effects are clearly shown in Table 5. Let us suppose the value of t is between –1.96 and +1.96, the relationship between factors is negligible at a 95% confidence level, and if t > 1.96 and þ < 0.05, the relationship between factors is significant at a confidence level of 95%.

The significance of all hypothesis paths in the research model was assessed. Therefore, none of them were dropped. The results prove that PR (β=0.230, p-value=0.000), PE (β=0.418, p-value=0.000), EE (β=0.242, p-value=0.007) were positively and significantly associated with CL. Hence H1, H2, and H3 were accepted. CL had a direct and considerable impact on BI (β=0.429, p=0.000) to use WhatsApp for language acquisition, supporting H4. This implies that CL partially mediated the effects of PR, EE, and PE. As H5 states, BI directly impacted IU positively (β=0.370, p-value=0.000). Therefore, the hypothesis was accepted.

Figure 2
Research Model



CL was found to play a momentous part in the adoption of TELL. The indirect relationships of PR (β=0.099, p=0.005), PE (β=0.180, p=0.001), and EE (β=0.104, p=0.028) were considerably affirmative with BI, implying that H1a, H2a, and H3a were accepted. The indirect relationships between PR (β=0.037, p=0.019), PE (β=0.066, p=0.008), and EE (β=0.038, p=0.042) with IU were also significant, which confirms H1b, H2b, and H3b.

All the indirect effects were substantial, but their effects were less significant than direct ones. Therefore, they were considered as partially mediated (Kale *et al*., 2019). This shows that the mediating role of collaborative learning enhanced the user intention to use technology for SLA. Table 5 explicitly reveals that PR, PE, and EE influence CL. At the same time, CL has a substantial direct effect on BI and the most decisive indirect impact on IU.

DISCUSSION

The aim of the study was to examine the impact of the identified factors on the intent to use WhatsApp for SLA, emphasising the mediating role of collaborative learning. The researchers proposed an adapted model of exploring-IT acceptance and use amongst students. The researchers hypothesised the exogenous variables which influence CL in the theoretical model. This is supported by the results. According to López-Nicolás et al. (2008), PR is an essential aspect in shaping individuals' IT acceptibility. Likewise, the current study also asserted that PR had a positive and considerable, direct consequence on collaborative learning. This implies that an individual's desire towards adopting technology for collaborative learning is shaped by its academic relevance. As for its significance in technology acceptance, it has caught the attention of many researchers over the years like González-Ibáñez et al. (2016). Thus, the present study indicated that WhatsApp is a useful tool for students, something which could supplement their language learning needs.

As determined in the prior study by Venkatesh et al. (2003), the finest determinant of intention is PE, and in the present study, it had a strong and affirmative effect on CL; although a few studies suggest otherwise (Iqbal & Qureshi, 2012; Yueh et al., 2015; Acharya & Ganesan, 2019). This is perhaps not surprising since an essential factor in determining students' behavior is the extent to which technology can be beneficial in educational settings (Lin et al., 2013; Tan, 2013; Martinho et al., 2018; Scherer et al., 2019). Similarly, EE also had a considerably affirmative impact on CL which is similar to the findings of Wang et al. (2009) and Botero et al. (2018). It was found that students were drawn towards tools that are easy to handle, yet informative and time-saving (Rahmi & Birgoren, 2020). The easier the device, the more likely they are to utilise it for regular academic purposes.

Collaborative learning, according to Zhampeissova et al. (2020), promotes student partnerships which help complete learning activities. Therefore, it is an essential construct in SLA. Social networking tools help enhance collaborative tasks (Heflin et al., 2017). WhatsApp notably has become a commonly used platform by young people, promoting accessibility and encouraging collaborative activities (Mistar & Embi, 2016; Güler, 2017; Soria et al., 2020). Hence, it is believed that CL will have a considerable direct impact on BI and will mediate the influence of PR, EE, and PE on BI. The current study notes that CL had a significantly affirmative impact on using WhatsApp for SLA, and it partially mediates the influence of the exogenous variables. PR, PE, and EE exert more substantial direct effects on CL than on BI. Therefore, the study validates and extends the research by Güler (2017), Gašević et al. (2019), and Soria et al. (2020).

Finally, it was anticipated that BI would have a substantial direct impact on IU. BI is a strong predictor of the desire to engage in a particular task (Dwivedi et al., 2019), Therefore, it had a direct and considerably positive impact on users' intent to use WhatsApp for SLA. This means that students who have higher BI to use WhatsApp for SLA are positively influenced towards actually using it. This finding goes in line

Table 5

Direct, Indirect & Total Effects

	Paths	*β*	Standard Deviation	*T* Statistics	*P* Values
DIRECT EFFECTS	BI → IU	0.370	0.073	5.075	0.000
	CL → BI	0.429	0.088	4.902	0.000
	EE → CL	0.242	0.090	2.695	0.007
	PE → CL	0.418	0.096	4.336	0.000
	PR → CL	0.230	0.064	3.616	0.000
	CL → IU	0.159	0.047	3.418	0.001
INDIRECT EFFECTS	EE → BI	0.104	0.047	2.194	0.028
	EE → IU	0.038	0.019	2.035	0.042
	PE → BI	0.180	0.055	3.283	0.001
	PE → IU	0.066	0.025	2.657	0.008
	PR → BI	0.099	0.036	2.778	0.005
	PR → IU	0.037	0.016	2.344	0.019
	BI → IU	0.370	0.073	5.075	0.000
	CL → BI	0.429	0.088	4.902	0.000
	CL → IU	0.159	0.047	3.418	0.001
	EE → BI	0.104	0.047	2.194	0.028
TOTAL EFFECTS	EE → CL	0.242	0.090	2.695	0.007
	EE → IU	0.038	0.019	2.035	0.042
	PE → BI	0.180	0.055	3.283	0.001
	PE → CL	0.418	0.096	4.336	0.000
	PE → IU	0.066	0.025	2.657	0.008
	PR → BI	0.099	0.036	2.778	0.005
	PR → CL	0.230	0.064	3.616	0.000
	PR → IU	0.037	0.016	2.344	0.019

Note. BI=behavioral intention; CL=collaborative learning; EE=effort expectancy; IU=intention to use; PE=performance expectancy; PR=perceived relevance.

with the study of Dwivedi et al. (2019). The empirical testing signifies that collaborative learning as a mediating variable enhanced the intention to use WhatsApp for SLA.

The present study has both practical and theoretical ramifications for academicians, as well as researchers. The investigation uses suitable statements, in order to understand the use and acceptance of technology for SLA. This is very important, since as technology is easily available to everyone in the form of android devices. Based on our findings, the researchers propose CL as an essential part of TELL in future research. Furthermore, the current analysis established connections that were not present in the previous technology acceptance models. This is a consequence of introducing a new construct (CL). These paths include PR → CL, EE → CL, PE → CL, CL → BI, and CL → IU. These offer new insights and maybe important considerations in IT acceptance and use among individuals.

The findings show that CL plays an essential role in determining students' behavioral intention towards TELL. Mainly, it had a significant effect on BI, which suggests that practitioners or curriculum designers may find it (CL) essential in influencing the behaviors and intention of students to adopt TELL. Since all technology attributes (PE, EE, and PR) have a significant effect on CL and BI, educationists should concentrate on improving the system's usefulness, its ease of utility, and the use of innovations relevant to their academic pur-

poses. The results of the study may assist in overcoming the existing problems and designing more robust pedagogical language learning practices through TELL.

CONCLUSION

Based on a systematic literature review, a theoretical model was developed. Hypothetically, three constructs were described to be the most contributory to enhancing university students' use of TELL: PR, PE, and EE, with the mediating role of CL to use WhatsApp for SLA. In order to empirically test the model, a study was conducted with 202 undergraduate students studying in institutes in Rajasthan, using the survey method. Empirical testing signifies that the mediator, collaborative learning, enhanced the user intention to use WhatsApp for second language acquisition. The indirect effects were positive yet less significant than the direct effects. If these aspects are taken into account, the successful integration of TELL in ELT classrooms could be facilitated. There are three limitations to this research. Firstly, the results of the survey cannot be generalised due to the sample size, location, sampling method, etc. Secondly, the study relied solely on student self-assessment, which can differ significantly as the nature of the respondents varies. Thirdly, the respondents are all internet users. So, the output of this study could be biased. The intention to adopt TELL may be superior when compared with non-users or occasional users of the internet. Therefore, future studies should involve the acceptance behavior of less active internet users and compare them to internet users, in order to arrive at more precise conclusions. Future research could attend to the effects of some more new and emergent moderating variables to strengthen the variance described by the predictors, since there are differences in the research setting.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Divya Jyot Kaur: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Niraja Saraswat: Conceptualization; Data curation; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Irum Alvi: Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing – review & editing; Writing – original draft; Writing – review & editing.

REFERENCES

Abu-Al-Aish, A., & Love, S. (2013). Factors influencing students' acceptance of m-learning: An investigation in higher education. *International Review of Research in Open and Distance Learning*, *14*(5), 82–107. https://doi.org/10.19173/irrodl.v14i5.1631

Acharya, S., and Ganesan, P. (2019). Factors that affect students' intention and use of technology: An assessment of UTAUT2 in the context of YouTube video forwarding behaviour. *International Journal of Business and Emerging Markets*, *11*, 348–367. https://doi.org/10.1504/ijbem.2019.105231

Alismaiel, O. A., Cifuentes-Faura, J., & Al-Rahmi, W. M. (2022). Social media technologies used for education: An empirical study on TAM model during the COVID-19 pandemic. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.882831

Avci, U., & Askar, P. (2012). The comparison of the opinions of the university students on the usage of blog and wiki for their courses. *Educational Technology and Society*, *15*(2), 194–205.

Bagozzi, R. P. (1981). Attitudes, intentions, and behavior: A test of some key hypotheses. *Journal of Personality and Social Psychology*, *41*(4), 607–627. https://doi.org/10.1037/0022-3514.41.4.607

Botero, G. G., Questier, F., Cincinnato, S., He, T., & Zhu, C. (2018). Acceptance and usage of mobile assisted language learning by higher education students. *Journal of Computing in Higher Education*, *30*(3), 426–451. https://doi.org/10.1007/s12528-018-9177-1

Bui, T. X. T., Ha, Y. N., Nguyen, T. B. U., Nguyen, V. U. T., & Ngo, T. C. T. (2021). A study on collaborative online learning among EFL students in van lang university (VLU). *AsiaCALL*. *12*, 9–21.

Celsi, R. L., & Olson, J. C. (1988). The role of involvement in attention and comprehension processes. *Journal of Consumer Research*, *15*(2), 210. https://doi.org/10.1086/209158

Cetinkaya, L., & Sutcu, S. S. (2019). Students' success in English vocabulary acquisition through multimedia annotations sent via whatsapp. *Turkish Online Journal of Distance Education*, *20*(4), 85–98. https://doi.org/10.17718/TOJDE.640517

Çetinkaya, L., & Sütçü, S. S. (2018). The effects of Facebook and WhatsApp on success in English vocabulary instruction. *Journal of Computer Assisted Learning*, *34*(5), 504–514. https://doi.org/10.1111/jcal.12255

Chen Hsieh, J. (2022). Multimodal digital storytelling presentations among middle-school learners of English as a foreign language: Emotions, grit and perceptions. *RELC Journal*, 500. https://doi.org/10.1177/00336882221102233

Church, K., & De Oliveira, R. (2013). What's up with WhatsApp? Comparing mobile instant messaging behaviors with traditional SMS. *MobileHCI 2013 - Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services* (352–361). Germany-Association for Computing Machinery. https://doi.org/10.1145/2493190.2493225

Dashkina, A., Kobicheva, A., Lazovskaya, T., Tokareva, E., Tarkhov, D., & Guselnikova, I. (2022). Examining the effectiveness of computer-supported collaborative learning for language proficiency purposes. *Sustainability*, *14*(10), 5908. https://doi.org/10.3390/su14105908

Deraman, R., Wang, C., Yap, J. B. H., Li, H., & Mohd-Rahim, F. A. (2019). Developing internet online procurement frameworks for construction firms. *Future Internet*, *11*(6), 1–22. https://doi.org/10.3390/FI11060136

Divjak, B., Rienties, B., Iniesto, F., Vondra, P., & Žižak, M. (2022). Flipped classrooms in higher education during the COVID-19 pandemic: Findings and future research recommendations. *International Journal of Educational Technology in Higher Education*, *19*(1). https://doi.org/10.1186/s41239-021-00316-4

Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-Examining the Unified Theory of Acceptance and Use of Technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers*, *21*(3), 719–734. https://doi.org/10.1007/s10796-017-9774-y

Escobar-Rodríguez, T., Carvajal-Trujillo, E., & Monge-Lozano, P. (2014). Factors that drive students to use facebook as a learning tool. An extension of the UTAUT. *IX Jornada De Docencia En Contabilidad Buenas Practicas Docentes En Materia Contable Dentro Del Eees* (pp. 135–152). Copicentro S. L.

Escobar-Rodríguez, T., Carvajal-Trujillo, E., & Monge-Lozano, P. (2014). Factors that influence the perceived advantages and relevance of Facebook as a learning tool: An extension of the UTAUT. *Australasian Journal of Educational Technology*, *30*(2), 136–151. https://doi.org/10.14742/ajet.585

Fischer, I. D., & Yang, J. C. (2022). Flipping the flipped class: using online collaboration to enhance EFL students' oral learning skills. *International Journal of Educational Technology in Higher Education*, *19*(1). https://doi.org/10.1186/s41239-022-00320-2

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Fouz-González, J. (2017). Pronunciation instruction through Twitter: The case of commonly mispronounced words. *Computer Assisted Language Learning*, *30*(7), 631–663. https://doi.org/10.1080/09588221.2017.1340309

Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika*, *61*(1), 101–107. https://doi.org/10.2307/2334290

González-Ibáñez, R., Escobar-Macaya, M., & Manriquez, M. (2016). Using low-cost electroencephalography (EEG) sensor to identify perceived relevance on web search. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–5. https://doi.org/10.1002/pra2.2016.14505301146

Güler, Ç. (2017). Use of WhatsApp in higher education: What's up with assessing peers anonymously? *Journal of Educational Computing Research*, *55*(2), 272–289. https://doi.org/10.1177/0735633116667359

Haugland, M. J., Rosenberg, I., & Aasekjær, K. (2022). Collaborative learning in small groups in an online course – A case study. *BMC Medical Education*, *22*(1), 1–10. https://doi.org/10.1186/s12909-022-03232-x

Heflin, H., Shewmaker, J., & Nguyen, J. (2017). Impact of mobile technology on student attitudes, engagement, and learning. *Computers and Education*, *107*, 91–99. https://doi.org/10.1016/j.compedu.2017.01.006

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Herrera-Pavo, M. Á. (2021). Collaborative learning for virtual higher education. *Learning, Culture and Social Interaction*, *28*, 100437. https://doi.org/10.1016/j.lcsi.2020.100437

Iqbal, S., & Qureshi, I. A. (2012). M-learning adoption: A perspective from a developing country. *International Review of Research in Open and Distance Learning*, *13*(3), 147–164. https://doi.org/10.19173/irrodl.v13i3.1152

Jafari, S., & Chalak, A. (2016). The role of WhatsApp in teaching vocabulary to Iranian EFL learners at junior high school. *English Language Teaching*, *9*(8), 85. https://doi.org/10.5539/elt.v9n8p85

James, M., Baptista, A. M. T., Barnabas, D., Sadza, A., Smith, S., Usmani, O., & John, C. (2022). Collaborative case-based learning with programmatic team-based assessment: A novel methodology for developing advanced skills in early-years medical students. *BMC Medical Education*, *22*(1), 1–12. https://doi.org/10.1186/s12909-022-03111-5

Kale, E., Aknar, A., & Başar, Ö. (2019). Absorptive capacity and firm performance: The mediating role of strategic agility. *International Journal of Hospitality Management*, *78*, 276–283. https://doi.org/10.1016/j.ijhm.2018.09.010

Kaur, D. J., Saraswat, N., & Alvi, I. (2021). Exploring the effects of blended learning using WhatsApp on language learners' lexical competence. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, *13*(4), 1–17. https://doi.org/10.21659/rupkatha.v13n4.60

Ko, E. G., & Lim, K. Y. (2022). Promoting English learning in secondary schools: Design-Based research to develop a mobile application for collaborative learning. *Asia-Pacific Education Researcher*, *31*(3), 307–319. https://doi.org/10.1007/s40299-021-00562-0

Lave, J., & Wenger, E. (1994). Situated learning: Legitimate peripheral participation. *American Ethnologist*, *21*(4), 918–919. https://doi.org/10.1525/ae.1994.21.4.02a00340

Lee, W. W. S., & Yang, M. (2020). Effective collaborative learning from Chinese students' perspective: A qualitative study in a teacher-training course. *Teaching in Higher Education*, *24*, 1–17. https://doi.org/10.1080/13562517.2020.1790517

Lin, S., Zimmer, J. C., & Lee, V. (2013). Podcasting acceptance on campus: The differing perspectives of teachers and students. *Computers and Education*, *68*, 416–428. https://doi.org/10.1016/j.compedu.2013.06.003

Martinho, D. S., Santos, E. M., Miguel, M. I., & Cordeiro, D. S. (2018). Factors that influence the adoption of postgraduate online courses. *International Journal of Emerging Technologies in Learning, 13*, 123–141. https://doi.org/10.3991/ijet.v13i12.8864

Mellati, M., Khademi, M., & Abolhassani, M. (2018). Creative interaction in social networks: Multi-synchronous language learning environments. *Education and Information Technologies*, *23*(5), 2053–2071. https://doi.org/10.1007/s10639-018-9703-9

Miguel, E., & Carney, W. (2022). Foreign language acquisition anxiety and project-based learning in collaborative L2 instruction: A case study. *Theory and Practice in Language Studies*, *12*(1), 1–6. https://doi.org/10.17507/tpls.1201.01

Mistar, I., & Embi, M. A. (2016). Students' perception on the use of Whatsapp as a learning tool in ESL classroom. *Journal of Education and Social Sciences*, *4*, 96–104.

Mofareh, A. (2019). The use of technology in English language teaching. *Frontiers in Education Technology*, *2*(3), 168. https://doi.org/10.22158/fet.v2n3p168

Naveed, Q. N., Alam, M. M., & Tairan, N. (2020). Structural Equation Modeling for mobile learning acceptance by university students: An empirical study. *Sustainability*, *12*(20), 8618. https://doi.org/10.3390/su12208618

Pietarinen, T., Palonen, T., & Vauras, M. (2021). Guidance in computer-supported collaborative inquiry learning: Capturing aspects of affect and teacher support in science classrooms. *International Journal of Computer-Supported Collaborative Learning*, 1–27. https://doi.org/10.1007/s11412-021-09347-5

Plana, M. G., Hopkins, J. E., Plana, M. G., & Appel, C. (2013). Improving learners ' reading skills through instant short messages : A sample study using WhatsApp. *IV World CALL Conference on Global Perspectives on Computer-Assisted Language Learning* (pp. 10–13). WorldCALL.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Rahmi, B. A. K. I, and Birgören, B. (2020). Analysis of the studies on e-learning acceptance of learners in the Middle East and the proposal of an extended technology acceptance model. *Kastamonu E˘gitim Dergisi*, *28*, 1977–1986. https://doi.org/10.24106/kefdergi.4169

Rambe, P., & Bere, A. (2013). Using mobile instant messaging to leverage learner participation and transform pedagogy at a South African University of Technology. *British Journal of Educational Technology*, *44*(4), 544–561. https://doi.org/10.1111/bjet.12057

Rambe, P., & Chipunza, C. (2013). Using mobile devices to leverage student access to collaboratively-generated resources: A case of WhatsApp instant messaging at a South African University. *International Conference on Advanced Information and Communication Technology for Education (ICAICTE)*. Atlantis Press. https://doi.org/10.2991/icaicte.2013.66

Ramírez, R. R., & Palos-Sánchez, P. R. (2018). Environmental firms' better attitude towards nature in the context of corporate compliance. *Sustainability (Switzerland)*, *10*(9). https://doi.org/10.3390/su10093321

Scherer, R., Siddiq, F., and Tondeur, J. (2019). The technology acceptance model (TAM): a meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers and Education*, *128*, 13–35. https://doi.org/10.1016/j.compedu.2018.0 9.009

Sharifi, M., Rostami AbuSaeedi, A. A., Jafarigohar, M., & Zandi, B. (2018). Retrospect and prospect of computer assisted English language learning: a meta-analysis of the empirical literature. *Computer Assisted Language Learning*, *31*(4), 413–436. https://doi.org/10.1080/09588221.2017.1412325

Sharma, A., & Alvi, I. (2021). Evaluating pre and post COVID 19 learning : An empirical study of learners ' perception in higher education. *Education and Information Technologies*. https://doi.org/10.1007/s10639-021-10521-3

Sixto-García, J., & Duarte-Melo, A. (2020). Self-destructive content in university teaching: New challenge in the digital competence of educators. *Communication & Society*, *33*(3), 187–199. https://doi.org/10.15581/003.33.3.187-199

Soria, S., Gutiérrez-Colón, M., & Frumuselu, A. D. (2020). Feedback and mobile instant messaging: Using whatsapp as a feedback tool in EFL. *International Journal of Instruction*, *13*(1), 797–812. https://doi.org/10.29333/iji.2020.13151a

Stöhr, C., Demazière, C., & Adawi, T. (2020). The polarizing effect of the online flipped classroom. *Computers & Education*, *147*, 103789. https://doi.org/10.1016/j.compedu.2019.103789

Sun, Y., & Gao, F. (2019). An investigation of the influence of intrinsic motivation on students' intention to use mobile devices in language learning. *Educational Technology Research and Development*, *68*(3), 1181–1198. https://doi.org/10.1007/s11423-019-09733-9

Tan, P. J. B. (2013). Applying the UTAUT to understand factors affecting the use of English e-learning websites in Taiwan. *SAGE Open*, *3*(4). https://doi.org/10.1177/2158244013503837

Toyama, M., & Yamazaki, Y. (2021). Classroom interventions and foreign language anxiety: A systematic review with narrative approach. *Frontiers in Psychology*, *12*(80), 13-25. https://doi.org/10.1080/09588221.2018.1557692

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly: Management Information Systems*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Wang, Y. S., Wu, M. C., & Wang, H. Y. (2009). Investigating the determinants and age and gender differences in the acceptance of mobile learning. *British Journal of Educational Technology*, *40*(1), 92–118. https://doi.org/10.1111/j.1467-8535.2007.00809.x

Webb, M., & Doman, E. (2020). Impacts of flipped classrooms on learner attitudes towards technology-enhanced language learning. *Computer Assisted Language Learning*, *33*(3), 240–274. https://doi.org/10.1080/09588221.2018.1557692

Yueh, H. P., Huang, J. Y., & Chang, C. (2015). Exploring factors affecting students' continued Wiki use for individual and collaborative learning: An extended UTAUT perspective. *Australasian Journal of Educational Technology*, *31*(1), 16–31. https://doi.org/10.14742/ajet.170

Zhampeissova, K., Kosareva, I., & Borisova, U. (2020). Collaborative mobile learning with smartphones in higher education. *International Journal of Interactive Mobile Technologies (IJIM)*, *14*(21), 4. https://doi.org/10.3991/ijim.v14i21.18461

Zhang, R., Zou, D., Cheng, G., & Xie, H. (2022). Implementing technology-enhanced collaborative writing in second and foreign language learning: A review of practices, technology and challenges. *Education and Information Technologies, 27,* 8041-8069. https://doi.org/10.1007/s10639-022-10941-9

https://doi.org/10.17323/jle.2023.10923

Where do Critical Pedagogy and Language Needs Analysis Meet? English as an Additional Language for Adult Refugees and Migrants in Greece: A Case Study

Christina Maligkoudi [1,2], **Anna Mouti** [2,3], **Eleni Triantafyllou** [2]

[1] Democritus University of Thrace, Alexandroupolis, Greece
[2] Hellenic Open University, Thessaloniki, Greece
[3] Aristotle University of Thessaloniki, Thessaloniki, Greece

ABSTRACT

Background. Language classes organized for adult refugees and migrants are heterogeneous. Students in these educational settings differ across a number of various aspects, including language competences, educational background and levels of literacy. Seen through the Critical Pedagogy lens language is considered not simply as a means to express or communicate, but as a product constructed by the ways language learners recognise themselves, their social surroundings, their histories, and their potentialities for the future.

Purpose. The purpose of our study is to unfold and identify the language needs of a specific group of migrants learners learning English as an additional language in Greece, where English is not the dominant language. We will try to focus and analyse language needs through the critical pedagogy lens and thus make the whole procedure an empowerment tools for the adult refugees and migrants.

Method. As a case study, this study follows a qualitative research design. This small-scale study focuses on a specific target group of language learners and their needs and attitudes towards learning. Class observations, field notes, interviews with the participants and questionnaires with open-ended questions were used as main methodological tools.

Results. The present article examines the needs of a group of immigrant adult learners attending English language classes at a non-formal educational setting located in Greece. The participants come from diverse ethnic and linguistic backgrounds and the majority of them speak Greek fluently since they have resided in Greece for a long time. A focal point throughout the process was students' greater involvement in the learning procedure and decision-making processes regarding the content and the presentation of the educational material.

Conclusion. Although the systematic needs analysis revealed that the reasons for participation and competence levels among participants varied a lot, a common goal for everyone was achieving oral fluency in the target language. Moreover, the results of this attempt were expressed n terms of learners' contributions, willingness to share their stories, even to talk about difficulties they met and caring about their classmates' stories. Thus, we suggest that the incorporation of personal experience in the learning process, not only functions as a link between students and language but also a process for team bonding and motivation.

KEYWORDS

needs analysis, adult refugees, migrants, critical pedagogy, L2 English, Greece

Citation: Maligkoudi C., Mouti A., & Triantafyllou E. (2023). Where do critical pedagogy and language needs analysis meet? English as an additional language for adult refugees and migrants in Greece: A case ыtudy. *Journal of Language and Education, 9*(1), 102-111. https://doi.org/10.17323/jle.2023.10923

Correspondence:
Anna Mouti,
mouti@itl.auth.gr

Received: July 06, 2020
Accepted: November 03, 2022
Published: March 31, 2023

INTRODUCTION

When it comes to language education for refugees and migrants, there is a vital need for courses to be relevant to the needs and lives of specific groups of learners (Malicka, Gilabert Guerrero & Norris, 2019) and for language programmes to take into consideration learners' present knowledge, lacks and

available resources (Nation & Macalister, 2009). These parameters should be taken into careful account, if language courses addressed to migrant populations are aimed at supporting learners' integration (Little, 2008). A starting point to meet this purpose in language classes is a design which considers three main variables: learners' needs analysis; environment analysis; and the application of language curriculum/ material design principles (Nation & Macalister, 2009). A successful needs analysis will ensure that the course will be appropriately designed for the participants (Serafini, Lake, & Long, 2015), will not be inefficient or inadequate (Long, 2005) and that learners' motivation will increase (Van Avermaet & Gysen, 2008). A needs analysis design for this study follows the Common European Framework of Reference for Languages (Council of Europe, 2001) and focuses on the domains where language is used or/and is desired to be used by the participants (Little, 2008). This is in an attempt to actively engage learners' in the decision-making procedure. Moreover, in an educational context shaped by critical literacy pedagogy, teachers must constantly teach a dual curriculum: a curriculum that empowers students to make sense of their everyday life; and a curriculum which enables students to obtain the tools for mobility valued in the dominant culture (Macalister & Nation, 2019).

There is limited research in the Greek context regarding the language needs analysis of adult refugees and migrants, especially after the 2015 refugee crisis (Androulakis et al., 2017; Mouti et al., 2021). Furthermore, the multilingual needs of migrants who are either residing temporarily or permanently settled in Greece, and use either lingua francas (mainly English) or their mother tongue to communicate with other non-Greeks (Mouti et al., 2021, p. 232), identified also in the Italian context by Bianco and Ortiz Cobo (2019, p. 12), seem to explain the migrants interest and desire to learn English as an additional language.

In the Greek context, as presented in the Eurobarometer Special Surveys (2014)[3], English is considered the most widely known language (L2). Among the three most widely known languages in Greece, English comes first (51% while in EU27 this percentage is much lower at 38%). More details on the Greek EFL context can be found in Angouri et al. (2010) but also in Sifakis (2009) for ELF (English as Lingua Franca). Tsagari (2016) offers the assessment orientations of EFL teachers in Greece and Cyprus, whileMouti et al (2019) offers information on test-taking strategies in L2 language assessment in Greece and Cyprus. In an exploratory study which attempted to look into the language education offered to refugees and migrants in Greece, Kantzou et al (2017) mentioned that there is an identified call for courses in languages other than Modern Greek ,at least amongs the population awaiting resettlement. English seems to be one of the languages offered in the framework of formal and non-formal language education for refugees and migrants in Greece.

The purpose of our study is to unfold and identify the language needs of a specific group of migrants learners learning English as an additional language in Greece, where English is not the dominant language. We will try to focus and analyse language needs through the critical pedagogy lens and thus make the whole procedure an empowerment tools for the adult refugees and migrants.

The rationale for this study is based mainly upon the question of the quality and content of English language programmes for migrants in formal or non-formal settings. In a context where motivation cannot be characterised as instrumental but as integrative, since English is an additional language for the majority of migrants residing in Greece, and they have already experienced the host language learning, parameters such as diverse linguistic repertoires and previous knowledge, identity and critical consciousness awareness tend to be neglected for the sake of grammar and vocabulary instruction leading courses to traditional ESL standards. On the other hand, as Mouti et al. (2022) state "the language needs of the refugees have to do not only with the host country language but also with the language of a destination host country (e.g. English or German)", especially in Greece and Italy who share a double role both as host and transition countries.

THEORETICAL BACKGROUND

Needs Analysis and English as an Additional Language

Language teaching has to reflect learners' experiences and language biographies in particular. In this way, learners are encouraged to build up on their own language learning skills and they are not treated as inexperienced learners (Corder, 2012). Cook (2001) states that in adult learner classes, very little is under the teacher's control. Learners make deliberate choices, they follow idiosyncratic strategies. They add things, in order to suit their needs creating complex meanings even with a little grammar. The purpose is to support learners to draw on their existing competences and experience but at the same to further develop these competences and to support their becoming autonomous learners. In other words, the aim is their being able to manage their own learning (Little, 2008). Another fact about adult learner groups is that, by nature, they present high levels of heterogeneity, simply because of the fact that they are adults and they share several characteristics, i.e. age, gender, educational background, profession etc. (Corder, 2012). Regarding linguistic heterogeneity, it is much higher among

[3] http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf

immigrant learners because they have extremely different linguistic biographies, depending on the status of their first language(s) in the country of origin, the other languages they have used during their migration and the language contacts in the host society (Krumm & Plutzar, 2008). Therefore, language programmes that affirm the value of all languages, all cultures and all ethnicities devoid of the implication that the language of the host community or any other language is superior to learners' mother tongue are of utmost importance (Little, 2008). On the contrary, the wealth of participants' linguistic repertoires can be seen as a useful tool for learners and educators alike which can promote effective learning and empowerment. For this reason, it is generally suggested that small groups and differentiated course systems are more effective than standardised programmes given the heterogeneity of target groups (Krumm & Plutzar, 2008).

It cannot be assumed that the completion of a language programme will lead to participants' complete integration in a host country as this is a long-lasting process. However, learning an additional language, English, in the case of this study, is believed to move beyond integration and give access to everyday interactions with locals, and access to equal opportunities. English in Greece although it is not the dominant language, is a language of power/ a lingua franca and there are a great number of Greeks or people residing in Greece who speak English fluently. Parks (2010, as cited in Warriner, 2016) states that improving one's English is ideologically identified with maximising one's capital, competing to become a valued worker in the new economy and it is clear that linguistic practicality, communicative efficiency, social mobility and economic advancement have all become increasingly associated with large languages, thus interfering with the maintenance of smaller ones (Edwards, 2002).

Moreover, it is often found that neoliberal discourse equates English language learning to improved educational, social and employment opportunities (Warriner, 2016). Graddol (2006, as cited in Wilton, 2009, p. 46) describes English as one of the basic skills, stating that "its function and place in the curriculum is no longer that of 'foreign language' and this is bringing about profound changes in the person who is learning English, their motives for learning it and their needs as learners". Therefore, English language acquisition is regarded as an unquestionable, qualification in modern societies for everybody. It is widely believed and argued that the user of English can, through effort and hard work, be transformed into a better form of human capital through increasing his/her formal or measurable competence in English. This view dominates not only the English language teaching universe, but it also "circulates in public portrayals of what kinds of skills, competencies, and trajectories immigrants need, want, and should develop for themselves" (Warriner, 2016, p. 496). Therefore, the choice of adult migrants learning English cannot be examined separately from its implications on their identity. The concept of international posture thus considerably broadens the external reference group from a specific geographic and ethnolinguistic community to a non-specific global community of English language users (Ushioda & Dörnyei, 2009).

Critical Pedagogy and Language Learning

L2 learners will master L2 communication, if they are stimulated and prepared to think critically towards L2 cultural values, in order to be exposed to deeper layers of culture and linguistic knowledge, in such a way that social and cultural empathy can be developed. In this sense, working along with learners', the educator's role is not to fill learners with knowledge but to engage them in meaningful dialogue and to develop a relationship with them, in order to provide both parts with a new way of thinking and thus, acting (Freire, 1973). Taking into account that neither theoretical knowledge nor the educator's authority is the core of any evolving dialogue (Pessoa & Freitas, 2012), discussing critical issues becomes a meaningful and empowering procedure. A critical pedagogical approach can provide learners with skills to learn and communicate with success in a L2 (Forcelini, 2016). According to the critical pedagogy approach, the construction of the self within or against mainstream conceptual views is achieved through critical pedagogic practices aimed at building up tolerance towards distinct cultural views, developing social and political awareness, and cultivating moral practices aligned with social and political integrity. According to Hinchey (2004), the focus of the critical classroom is not on rote learning but on encouraging learners to question, realise and examine the existing conditions in the world around them and their positioning in them ,as well as to develop a skeptical position on their surroundings and the relations of power underlying them. That is why modern SLA (Second Language Acquisition) teaching techniques (i.e. Task-Based Teaching, Negotiation for Meaning etc.) have been criticized as being rationalist activities since they are devoted to the transfer of information without any reference to the social context (Okazaki, 2005).

By implementing critical pedagogic practices, educators and learners become partners, as learners are encouraged to express their ideas and, thus, channels are created where teachers and students can learn from each other's ideas and experiences and develop conclusions together (Forcelini, 2016). Language is not just a means of communication but it is connected "with considerations of equal and full participation and access to symbolic, material and cultural resources" (Gounari, 2014, p. 262). A critical approach to pedagogy needs to "aim at transformation, a way of shifting pedagogical relations to give students more curricular control, and ways of engaging with difference not merely in terms of inclusivity and issues but also at the level of desire" (Pennycook, 1999, p. 341). The basic point in the theory of critical pedagogy is that by gaining one's voice and resisting unjust reproduction in their own self-interest, students start to become active agents for social change (Okazaki, 2005).

As yet there is not much empirical research that explicitly examines language learners' evolving perspectives during a critical content-based course.

This paper documents how the actual needs of migrant students could be identified and recorded, in order for them to be used to design tailor-made educational material for English as an additional language. Language is not simply a means of expression or communication but also a practice that constructs, and is constructed by, the ways in which language learners understand themselves, their social surroundings, their histories, and their possibilities for the future (Norton & Toohey, 2004). In this vein, the students who participated in our study were able to work in groups and reveal and express their needs lacks, wants and preferences with the aim of contributing to material development, to integrate their personal stories and beliefs and to gain the power of deciding what they will be taught and how. The special element of our study in the Greek context is the fact that English is not the language of the host country, and that this is a case study implemented in Greece about English as an additional language for migrants, setting the role of Greece as a transition country more strongly than its role as a destination country for refugees and migrants.

METHOD

Research

The aim was to explore the meaning which individuals or groups of people give to social realities (Creswell, 2014). As a case study through action-research, this study follows a qualitative research design. Qualitative research strategies and tools were adopted, in order to collect specific data.

More specifically, the following research tools were implemented: (1) Class observations, (2) Field notes, (3) Interviews with the participants and, finally, (4) Questionnaires with open-ended questions, in order to gain access to participants' demographic data. To this end, this small-scale study focuses on a specific target group of language learners and their needs and attitudes towards learning.

The Research Site and Sample

The study took place from November 2018 to January 2019, in a solidarity school located in Thessaloniki. The foundation and beginning of the action of the school dates back to 1997, when a group of volunteer teachers started delivering Greek language lessons to migrants, refugees, asylum seekers and repatriated adults for free. It is a non-formal educational setting aiming not only to provide language lessons but also to help students develop a critical attitude towards social inequality, to counter racism and discrimination, to empower and facilitate integration in the host country and to provide students with a space where they can exchange experiences and their creativity. The school also stands by,and associates with a number of antiracist collectivities in the city sharing the same goals.

The average number of research participants is 17-20 migrants attending A2 - B1 English language classes at a solidarity school in Thessaloniki, for two hours twice a week, during the school year 2018-2019. Given the participants' irregular attendance in the classes, the number of learners participating in the needs analysis process and the interventions is specified accordingly in the needs analysis section, The attendance of any lesson was recorded. The participants came from diverse ethnic backgrounds with the majority of them coming from Albania and Russia, and some of them from Senegal, Romania, Georgia and Iran. Two of the students came from Greece. All participants gave their consent to the researcher-educator.

Data Analysis

Participants' Profile

The demographic information of the participants, as collected by the questionnaire, is presented in the table below. In order to preserve confidentiality, alias are used in lieu of participants' real names.

Regarding the demographic data of the research, it has to be clarified that the target group was highly heterogenous in almost any variable examined. Older participants tended to participate less in the process than younger, most of the times feeling insecure or resorting to Greek in order to make meaning. However, no other significant differences were found regarding the age variable. The needs analysis was conducted by the researcher-teacher herself. She performed some diagnostic assessment test, in order to measure their performance in Greek. The time of residence in Greece was not found to affect learners' performance, apart from the case of Haleh and Arash who, enrolled in the lessons later but, very quickly, became active participants in the lessons and began communicating with their classmates. In the case of the two men, both residing in Greece only for a short period (5 and 3 months respectively), the fact that they did not speak Greek nor attended Greek language lessons but were able to create meaning in English, had a significant impact on the way the lessons were conducted. It should be noted here that actual changes may occur in language and situation needs when new students enter the programme (Brown, 1995). The two students altered group dynamics in a positive way, since the rest of the participants stopped using Greek and tried to make meaning only in the target language, in order to ensure that their two classmates understood what was happening in the classroom. This was a habit that had been gradually built from the beginning of the lessons but it was fully embraced by everyone when the need for a common code of communication for all students

Table 1

Participants' Profile (N=17)

Name (Gender)	Age	Country of origin	Years in Greece	Spoken languages	Educational background	Previous English language attendance	English language learning setting
Belina (f)	28	Albania	6	Albanian, Greek, French	University graduate	3 months	Private language courses
Kevin (m)	20	Armenia	3	Armenia, Russian, Greek	University student	6 years	School/ internet
Enid (m)	38	Albania	20	Albanian, Greek, Italian	High School graduate	8 years	School
Salliou (m)	30	Senegal	10	Senegalese, French, Greek	High School graduate	3 years	School
Dimitra (f)	62	Greece	Since birth	Greek	High School graduate	2 years	Solidarity school
Adania (f)	35	Albania	20	Albanian, Greek	High School graduate	3 months	Private language courses/ internet
Festim (m)	45	Albania	20	Albanian, Greek	University graduate	1 year	Private language courses
Dazim (m)	25	Albania	9	Albanian, Greek	High School graduate	3 years	School
Samir (m)	48	Egypt	18	Arabic, Greek	University graduate	5 years	School
Stephane (m)	55	Uzbekistan	16	Russian, Greek	University graduate	4 years	School/ Solidarity school
Samuil (m)	36	Georgia	25	Georgian, Russian, Greek	High school graduate	3 years	School
Crista (f)	40	Romania	25	Romanian, Greek	University graduate	1 year	Private language courses
Popi (f)	60	Greece	Since birth	Greek, French	High School graduate	1 year	Solidarity school
Vladimir (m)	34	Georgia	13	Georgian, Russian, Greek	High School graduate	2 years	Solidarity school
Milona (f)	30	Albania	8	Albanian, Greek	High School graduate	1 year	Solidarity school
Haleh (m)	30	Iran	5 moths	Farsi, Turkish, French, Deri	Postgraduate studies	2 years	Private language courses
Arash (m)	28	Iran	3 months	Farsi, Polish	University graduate	1 year	Private language courses

was created by the arrival of the two new students. Moreover, it had been observed that the use of mother tongue between the two men in order to negotiate meaning and reach understanding, in some cases, was noticed by other participants who gradually felt more confident to do the same with classmates of the same background.

The linguistic repertoires of most participants included more than two languages, one being their mother tongue and the second Greek. Here, it was found that in speaking and listening activities, Dimitra and Popi, the two Greek women, were less flexible and hesitant than their classmates. French was also used sometimes to give explanations both between Salliou and Haleh and by the teacher to make comparisons which proved helpful for the two participants. The variable of the time of previous formal or non formal instruction in English revealed interesting results in comparison to group observation. Although in the case of some students, the time of language learning reflected their performance, as in the cases of Kevin, Enid and Samir, the cases of Belina, Adania and Christa were a surprise. All of them actively participated and communicated using the target language, making minor mistakes and trying to overcome areas of struggle by paraphrasing or asking for help, always in English.

Regarding language use in everyday situations, the majority of the participants do not use English, while few of them who use it, practice only receptive and not productive skills.

More specifically, nine students answered that they only practice language during the lessons, and three of them answered that although they do not speak English, they use it daily to browse and find information on the internet. Two students use English daily for their studies and work and three of them use English daily: Salliou, when he does not understand Greek, and Haleh and Arash to communicate with other people since they do not speak Greek. The students' future plans varied regarding migrating to an English-speaking country Two of the students, Dazim and Kevin, the youngest in the class, stated that they plan to move to another country, namely Canada and the USA. Five of the students expressed a possibility, depending on professional and financial opportunities that may arise. The rest of the class answered that they have no intention of moving to another country.

The last data collected by the first part of the questionnaire was relevant to learners' attitudes towards English. All students expressed their positive attitudes towards English, the international prevalence of the language and its popularity were mentioned by all, stressing that it is the only way to communicate with people from other countries. Kevin identified speaking English with speaking the language of future writing "it's global, it's the language of computers and the future", a statement closely related to his everyday need for using English at school and home. Festim answered that "you can go wherever you want with English" and Adania confessed that "Sometimes I feel embarrassed when I can't speak [English]".

Needs Analysis

The self-assessment grid of the Common European Framework of Reference for Languages (Council of Europe, 2001, p. 24-26) was used by the researcher-teacher as a guide, in order to identify students' level during observation and to design the second part of the questionnaire. The questionnaire included questions about student's learning preferences, in order to guide the material design.

More analytically, seven students could effectively understand and use familiar everyday expressions and basic phrases, introduce themselves and others ask and answer questions about personal details. They could interact in simple ways provided that the interaction was clear and slow. They could write short, simple texts and fill in their personal details. Six students presented a higher level of competence, being able to understand and produce more complex sentences, interact fluently in dialogues requiring exchange of information and read and understand different types of texts. Four of the students were capable of interacting in simple conversations but it was observed that they could easily manage texts of a higher difficulty with unknown vocabulary and complex grammar structures without assistance.

As a result, the target levels of the class would be A2 and B1. By considering this, the second part of the questionnaire was designed to include competencies and skills present in the two target levels according to the CEFR (Council of Europe, 2001). The selection of the variables was based upon these suggestions. On a Likert scale from 1 (least important) to 5 (most important), students were asked to rate the skills which were more important to them. A brief analysis of students' answers in each section follows.

Regarding the selection of their preferable skill, all students answered that speaking was the most important. Reading and listening followed in importance by the majority of students. Writing was ranked as important and four students replied that grammar skills are more important to them (Table 2).

Table 2

Most Important/Preferable Skill Selection - Ranking

Skill	Mean
Speaking	4.8
Reading	3
Listening	3.3
Writing	1.2
Grammar skills	0.6

The most popular writing exercise among the target group were to write about yourself/describe experience and write about familiar topics. Note-taking followed. Accuracy also received a high rate of answers, with seven students answering that it is important and, two very important and one most important. Learners' answers about reading activities were hard to analyse. Reading timetables and stories received the most important answers, however, reading simple texts and articles on the internet were also high in importance (Table 3).

Table 3

Writing and Reading Activities-Preference Ranking

Writing	Mean
Write about yourself/describe experience	4.2
Write about familiar topics	3.8
Note-taking	3.3
Use correct grammar and vocabulary	2.8
Write simple letters/emails	3.1
Reading	Mean
Reading timetables	4.4
Reading stories	3.4
Reading simple texts and articles on the internet	4
Reading simple everyday material	3.1

As far as speaking skills are concerned, the most popular skills were simple everyday dialogues about oneself and dealing with situations while travelling. Describing background and education was ranked as important by five participants. Expressing hopes and ambitions was selected by only two participants and ranked as less important, while formal dialogues received the greatest number of least important answers. Most answers about listening activities included listening to dialogues and four students ranked listening to songs as important (Table 4).

Table 4

Speaking and Listening Activities-Preference Ranking

Speaking	Mean
simple everyday dialogues about oneself	4.5
deal with situations while travelling	4.3
Expressing hopes and ambitions	3
formal dialogues give reasons for my choices	2.5

Listening	Mean
listening to dialogues	4
listening to songs	3.7
listening to simple information	2
listening to the news	1.4

Regarding grammar instruction, explicit presentation of grammar rules was chosen as most important by only two students and, very important by three. Seeing grammar rules in text was the most popular answer and learn grammar rules by heart the least (Table 5).

Table 5

Grammar Instruction-Preference Ranking

Grammar Instruction	Mean
Clear presentation of grammar rules/theory	1.6
Grammar exercises	2.3
See grammar in texts	4.1
Learn grammar rules by heart	0.4

At the end of the questionnaire, a grid was provided so that learners could complete their own ideas on content selection. Although their answers varied, the majority of learners answered travelling and attending job interviews (which presented some inconsistency with their previous answers regarding formal interactions with people they do not know). Other answers included shopping for clothes, talking about sports, serving customers in a shop and going to the doctor. All suggested categories were covered throughout the school year, as tailor-made educational material was developed to address their needs and preferences.

The interviews conducted with the participants, provided similar results as those elicited through the questionnaire. Specifically, Stephane attributes lack of communication to lack of target vocabulary, understands what he already possesses, and asks for more conversations in the classroom:

> "to talk to each other more, we understand a bit but it is quite difficult to speak"

Dimitra's answer was also based on what she believes that she lacks and what she wants:

> "It is easy when I read slowly from a text, I understand, but I can't manage oral skills, I can't make meaning in this fast pace [...] when you speak, I understand. I can't understand my classmates because they have a different accent"

Other students also mentioned improvement of oral skills, becoming specific to what exactly they want:

> "To speak dialogues" (Festim),

> "I would like more dialogues to be able to speak with others, we need conversation" (Enid).

Participants also have added other topics as well, mainly focused on everyday interactions. Enid said:

> "talking about everyday things, what we do, where we work.... There are a lot of things but mainly these...".

Adania became more specific and expressed her desire regarding content and functions stating that she wants to learn how to book a room or order food at a restaurant. She also made her intentions clear, "since I want to sit English language exams I want more grammar".

DISCUSSION

The main goal of this study was to explore learners' needs and interpret them into those choices that would effectively guide the curriculum design and provide learners with the best learning outcomes. Before the Needs Analysis process, however, two main characteristics were already known about the participants: the first being that they are adults; and the second that they are migrants. As a result, the design of the educational material would be insufficient, if these two characteristics were not taken into consideration. According to Berwick (1989), material designs can be based on different educational values that underpin language programme planning. They can be based on an organised body of knowledge, on specific competencies, on social activities and finally, on cognitive or learning processes, on feelings and attitudes and on the needs and interests of the learners. Moreover, although the design of this study focuses mainly

on the latter, involving the systematic assessment of learners' language needs and their participation in the planning process, it should also be mentioned that the design was also an attempt to include learners' feelings and attitudes, while adopting a humanistic approach in language learning with a view to bringing development of the person through language (Berwick, 1989).

It is widely accepted that people enroll in language course with the aspiration of being able to communicate using the target language in various interactions (Nunan, 2004, Richards, 2001, Spratt, 1999), and that students needs and interests are reported to be communicative and functional in nature (Alalou, 2001). The target group of this study was no exception. Students, from the very beginning of the research, expressed their desire to acquire language for communicative purposes. Although the systematic needs analysis revealed that the reasons for participation and participants' competence levels varied a lot, achieving oral fluency in the target language was a common goal for everyone.

Regarding content, travelling and job seeking were the most popular topics suggested by the participants throughout not only in the Needs Analysis process but also during discussions inside and outside classroom. Harlow et al. (1980) and Alalou and Chamberlain (1999) (as cited in Alalou, 2001), found that many students view travel as the most likely opportunity to practice language skills. This correlates with their need for acquiring both speaking and listening skills. In the same way, job-related vocabulary and interactions constituted a chance for communication in the target language with a functional and practical value for participants.

It can be argued that learners' involvement in the designing and decision-making process was achieved, albeit through a lot of effort, negotiation and time spent in the learning setting. It was unrealistic to expect learners to be able to express their needs and preferences from the beginning of the lessons was unrealistic. However, after explaining the need for their involvement and experiencing the realization of their suggestions in practice, a lot of the participants felt free to express their preferences, their difficulties and what they would like to do next in the lessons.

An important factor which contributed to this end, was the gradual building of a relationship of trust between the participants and the teacher, but also among the participants. It was observed that after the completion of the first month of lessons, learners began to express their difficulties and inquiries in an assertive way, rather than insecurely as they had done in the beginning. To this end, it can be stated that learner autonomy for this particular group was achieved and the start was made during the needs analysis procedure.

Furthermore, placing learners' personal experience in the centre of the design proved to be a practice which radically changed classroom and group dynamics. Albeit hesitant in the beginning, since participants' response to such an initiative was unpredictable, it was found that it encouraged participation. Frequent times sharing with others even challenging experiences, had a positive effect on participants. It is believed that during this process, learners' diversity and wealth of experience were used as a resource and not as an obstacle to overcome and learning took place in a more even way (Arnold, 2011).

Last but not least, non-formal educational settings can be considered as more suitable contexts for the implementation of innovative teaching methods in comparison with formal educational settings, where the educators probably are more oriented towards a specific syllabus. In this vein, such educational contexts can be considered as 'safe spaces', where the learners can navigate through their multilingual and multicultural identities (Mattheoudakis, Chatzidaki and Maligkoudi, 2020).

CONCLUSION

The results of this attempt were learners' contribution and willingness to share their stories, even talk about difficulties they encountered and to care about their classmates' stories. To this end, the incorporation of personal experience in the learning process, not only functioned as a link between individuals and language but also resulted in team bonding and motivation. Moreover, it was vital, from the beginning of the research that possible gaps in the assessment and interpretation of learners' needs, created by diverse perceptions or inadequate interpretation of learners' subjected needs, would be filled by building a strong rapport with the participants. Although not measurable, considerable effort and time was spent in order to create a supportive learning environment, based on mutual trust where learners would feel free to express their likes and dislikes regarding the process, share their difficulties with their classmates and the teacher and make their own suggestions. In the end, it was found that learners were willing to participate in this process and set their own goals along with the teacher and the educational setting could be considered as a site of negotiation.

The entire Language Needs Analysis field as a procedure implemented in the language classroom is probably a prerequisite or an essential element in the attempt of applying transformative pedagogy principles in the foreign/second language classroom. Language Needs Analysis brings the learners' personal needs and identity profiles to the fore. A learner-centered framework leads to the development

of tailor-made courses, meeting the expectations of the relevant groups. Adopting the design and development of the materials and teaching approach to the specific group's needs and providing the opportunities for cooperative language learning through autonomy is where language learning and transformative pedagogy could meet and evolve.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Christina Maligkoudi: Conceptualization; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Anna Mouti: Conceptualization; Methodology; Resources; Supervision; Validation; Writing – original draft; Writing – review & editing.

Eleni Triantafyllou: Conceptualization; Data curation; Formal analysis; Funding acquisition.

REFERENCES

Alalou, A. (2001). Reevaluating curricular objectives using students' perceived needs: The case of three language programs. *Foreign Language Annals, 34*(5), 453-469. https://doi.org/10.1111/j.1944-9720.2001.tb02085.x

Alalou, A., & Chamberlain, E. (1999). Using Student expectations and perceived needs to rethink pedagogy and curriculum: A case study. *Foreign Language Annals, 32*, 27-35.

Angouri, J., Mattheoudakis, M., & Zigrika, M. (2010). Then how will they get 'the much-wanted paper'? A multifaceted study of English as a foreign language in Greece. In A. Psaltou-Joycey & M. Mattheoudakis (Eds.), *Advances in research on language acquisition and teaching* (pp. 179-194). Greek Applied Linguistics Association.

Arnold, J. (2011). Attention to affect in language learning. *Anglistik. International Journal of English Studies, 22*(1), 11-22.

Berwick, R. (1989). Needs assessment in language programming: From theory to practice. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 48-62). Cambridge University Press.

Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. Heinle & Heinle Publishers.

Cook, V. (2009). Developing links between second language acquisition research and language teaching. *Handbook of Foreign Language Communication and Learning, 6*, 139-161. https://doi.org/10.1515/9783110214246.1.139

Corder, N. (2012). *Learning to teach adults: An introductio*n. Routledge.

Creswell, J. W. (2014). *Research design. Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications, Inc.

Edwards, J. (2002). *Multilingualism*. Routledge.

Forcelini, J. (2016). Critical pedagogy and language acquisition: Benefiting from a country's crisis to improve Second Language Instruction. *Journal of Global Initiatives: Policy, Pedagogy, Perspectives, 11*(1), 123-136.

Freire, P. (1973). *Education for critical consciousness* (vol. 1). Bloomsbury Publishing.

Gounari, P. (2014). *Rethinking heritage language in a critical pedagogy framework. Rethinking heritage language education* (pp. 254-268). Cambridge University Press.

Hinchey, P. (2004). *Becoming a critical educator. Defining a classroom identity, designing a critical pedagogy*. Peter Lang.

Kantzou, V., Manoli, P., Mouti, A., & Papadopoulou, M. (2017). Language education for refugees and migrants: Multiple case studies from the Greek context. Διάλογοι. *Theory and Praxis in Education, 3*, 18-34. https://doi.org/10.12681/dial.15000

Krumm, H. J., & Plutzar, V. (2008). *Tailoring language provision and requirements to the needs and capacities of adult migrants*. Council of Europe.

Macalister, J., & Nation, I. P. (2019). *Language curriculum design*. Routledge.

Malicka, A., Gilabert Guerrero, R., & Norris, J. M. (2019). From needs analysis to task design: Insights from an English for specific purposes context. *Language Teaching Research, 23*(1), 78-106.

Mattheoudakis, M., Chatzidaki, A., & Maligkoudi, C. (2020). Heritage language classes and bilingual competence: The case of Albanian immigrant children in Greece. *International Journal of Bilingual Education and Bilingualism, 23*(8), 1019-1035. https://doi.org/10.1080/13670050.2017.1384447

Mouti, A., Maligkoudi, C., & Gogonas, N. (2022). Language needs analysis of adult refugees and migrants through the CoE-LIAM Toolkit: The context of language use in tailor-made L2 material design. *Selected Papers on Theoretical and Applied Linguistics, 24*, 600-617. http://dx.doi.org/10.26262/istal.v24i0.9210

Mouti, A., Rousoulioti, Th., & Ypsilandis, S. G. (2019). Test-Taking strategies in L2 language assessment: Perceptions of L2 learners in Greece and Cyprus. In *Situating strategy use in the Greek setting. Series language's parallels: Theory and teaching practice* (vol. 1). Saita Publications.

Nation, I. S., & Macalister, J. (2009). *Language curriculum design*. Routledge.

Norton, B. & Toohey, K. (2004). *Critical pedagogies and language learning*. Cambridge University Press.

Nunan, D. (2004). *Task-Based language teaching*. Cambridge University Press.

Okazaki, T. (2005). Critical consciousness and critical language teaching. *Second Language Studies, 23*(2), 174-202.

Pennycook, A. (1999). Introduction: Critical approaches to TESOL. *Tesol Quarterly, 33*(3), 329-348.

Pessoa, R. R. & Freitas, M. T. U. (2012). Challenges in Critical Language Teaching. *TESOL Quarterly, 46*(4), 753-776. http://dx.doi.org/10.1002/tesq.46

Richards, J. C. (2001). *Curriculum development in language teaching*. Ernst Klett Sprachen.

Serafini, E. J., Lake, J. B., & Long, M. H. (2015). Needs analysis for specialized learner populations: Essential methodological improvements. *English for Specific Purposes, 40*, 11-26. https://doi.org/10.1016/j.esp.2015.05.002

Sifakis, N. C. (2009). Challenges in teaching ELF in the periphery: The Greek context. *ELT Journal, 63*, 230–237. https://doi.org/10.1093/elt/ccn057

Spratt, M. (1999). How good are we at knowing what learners like? *System, 27*(2), 141-155. https://doi.org/10.1016/S0346-251X(99)00013-5

Tsagari, D. (2016). Assessment orientations of state primary EFL teachers in two mediterranean countries. *CEPS Journal, 6*, 1, 9-30. https://doi.org/10.26529/cepsj.102

Ushioda, E. & Dörnyei, Z. (2009). Motivation, language identities and the L2 self: A theoretical overview. In Z. Dörnyei & E. Ushioda (Ed.), *Motivation, language identity and the L2 self* (pp. 1-8). Multilingual Matters. https://doi.org/10.21832/9781847691293-002

Van Avermaet, P., & Gysen, S. (2008). Language learning, teaching and assessment and t he integration of adult immigrants. The importance of needs analysis. *Thematic studies: The linguistic integration of adult migrants* (pp. 59-70). Council of Europe.

Warriner, D. S. (2016). 'Here, without English, you are dead': Ideologies of language and discourses of neoliberalism in adult English language learning. *Journal of Multilingual and Multicultural Development, 37*(5), 495-508. http://dx.doi.org/10.1080/01434632.2015.1071827

Wilton, A. (2009). Multilingualism and foreign language teaching. In K. Knapp & B. Seidhofer (Eds.), *Handbook of foreign language communication and learning HAL 6* (pp. 45-78). Mouton de Gruyter.

https://doi.org/10.17323/jle.2023.13801

An Output-Oriented Approach to the Impact of Online Written Languaging on Form-Focused Writing Tasks

Natasha Pourdana

Islamic Azad University, Karaj Branch, Karaj, Iran

ABSTRACT

Background. Despite the growing interest of second language acquisition (SLA) researchers in the languaging process, a few studies have been done on its dynamic attributes in various writing tasks.

Purpose. This study investigated how online written languaging (WL) might impact English-as-a-Foreign-Language (EFL) learners' performance on form-focused writing tasks with production-based and comprehension-based output orientation in Google Docs, and how the output orientation of form-focused writing tasks could determine the WL attributes of quantity and focus.

Method. To do so, 112 Iranian EFL university students were selected and assigned to four parallel groups. In an eight-week experiment, two groups worked on gap-filling tasks (production-based) and two groups on error-identification tasks (comprehension-based) in parallel ±WL conditions.

Results. Statistical analysis indicated a significant interaction between task output orientation and WL production. So, on both production-based and comprehension-based tasks, the +WL groups outperformed the -WL groups. Moreover, in the +WL condition, the task output orientation determined the quantity of WL episodes, but not their focus on grammar (G-WL) and lexis (L-WL) in production-based and comprehension-based tasks. As such, the +WL group who completed the production-based tasks produced much more WL episodes than the +WL group who completed the comprehension-based tasks. Yet, both groups equally produced more L-WL episodes than G-WL episodes.

Conclusion. The study had several implications for language teachers to maximize learning opportunities by teaching *how to language* in various writing tasks on online platforms. The L2 teachers are also recommended to adopt an alternative approach to translation as a form-focused writing task.

KEYWORDS

comprehension-based, google docs, output orientation, production-based, written languaging

Citation: Pourdana N. (2023). An output-oriented approach to the impact of online written languaging on form-focused writing tasks. *Journal of Language and Education, 9*(1), 112-127. https://doi.org/10.17323/jle.2023.13801

Correspondence:
Natasha Pourdana,
natasha.qale@gmail.com

Received: January 28, 2021
Accepted: September 14, 2022
Published: March 31, 2023

INTRODUCTION

Grounded in Vygotsky's sociocultural theory of mind (SCT, 1987), languaging is conceptualized as "an action – a dynamic, never-ending process of using language to make meaning" (Swain, 2006, p. 96). Underpinning the Comprehensible Output Hypothesis, Swain (2006) paralleled the languaging output with metalinguistic output, by stating that both might function as the "tools of the mind, mediating the cognition and recognition of experience and knowledge" (p. 106).

It has been argued that L2 learners' languaging can facilitate second/foreign language (L2) learning when they are working on form-focused writing tasks (Behbahani et al., 2011; Nguyen, 2020; Pourdana et al., 2011; 2021; Swain, 2006). To reciprocate, the form-focused writing tasks can substantially induce languaging episodes which are deemed to enhance the depth of processing in writing (Pourdana, 2022) and mediate the feedback uptake (Bataineh et al., 2017; Ellis, 2001; Jang et al., 2020; Nour et al., 2021). By the same token, Suzuki and Itagaki (2009) speculated that the attributes of

written languaging episodes are highly task-dependent. In other words, the quality and quantity of WL episodes can notably change while L2 learners perform different types of writing tasks with comprehension-based (e.g., error-recognition) and/or production-based (e.g., gap-filling) output orientation (Keshanchi et al., 2022).

Despite the growing interest of second language acquisition (SLA) researchers in the constructive role of languaging in L2 learning, they have more focused on the attributes of oral languaging (Ammar & Hassan, 2018; Azkarai & Kopinska, 2020; Heidari et al., 2019; Lavasani et al., 2021) with less interest in languaging in written modality (Yilmaz, 2016). Moreover, SLA researchers narrowed their scope to study the languaging production on content-focused writing tasks such as picture description or data commentary (Kazemi et al., 2022; Falhasiri, 2021; Pourdana & Asghari, 2021), with minimal attention to form-focused writing tasks such as translation (Keshanchi et al., 2022). Similarly, adopting an output-oriented approach to WL on production-based tasks or comprehension-based tasks has been the subject of scrutiny in only a few studies (Pourdana et al., 2012; Storch, 2013; Suzuki & Itagaki, 2009; Zhang, 2021), with scant attempts to cross-examine the impact of WL on writing tasks performance with production-based and comprehension-based output orientation. Such interdependency among the WL attributes and the output orientation of the L2 writing tasks is the untaken road that the current study has pursued.

LITERATURE REVIEW

Languaging in Theory and Practice

The notion of *languaging* was initially developed by Swain in 2006. Subsequently, languaging and its self-explaining effect have widely been examined in various non-linguistic fields of science (e.g., Chi, 2000; Karpov, 2013). In Vygotsky's SCT framework, human cognition is optimally constructed by the language function as a semiotic device. Accordingly, L2 learners can shape and sharpen their thoughts as they talk with themselves (i.e., self-languaging) or with others (i.e., peer/pair languaging) (Lantolf et al., 2015), or in collaborative dialogues (Rafi & Pourdana, 2023). Recently, SLA researchers incorporated the languaging process as a platform for studying various aspects of L2 development (Falhasiri, 2021; Manchón et al., 2020). It has been argued that languaging enhances the memory span of L2 learners in performing cognitively complex tasks (Esfandiari & Noor, 2018), and pushes their attention to the work-in-progress (Suzuki & Storch, 2020).

Languaging is not always confined to the oral mode. Written languaging (hereafter, WL) was introduced by Suzuki (2009) and claimed as an "equivalent of private speech, but in writing" (p. 4). He further acknowledged three important advantages of WL over oral languaging, in terms of (1) serving the L2 learners with extra time to process the target language form, (2) freeing them from immediate language processing, and (3) offering them an extra memory space (Suzuki, 2017).

Several studies on WL have focused on the association of WL and L2 learner proficiency (Ishikawa, 2018), written corrective feedback (WCF, Nicolás-Conesa et al., 2019), face-to-face vs. computer-mediated modalities (CMC) (Shekary & Tahririan, 2006), and task types (Suzuki & Itagaki, 2009). In an experimental study, for instance, Ishikawa (2018) examined the impact of WL production on 83 Japanese EFL learners' grammar achievement. She assigned them to four groups at higher/lower proficiency levels to complete one of the fill-in-the-blank production or multiple-choice recognition grammar tasks with and without generating WL episodes. Ishikawa (2018) concluded that only the +WL groups who completed the production grammar tasks had significant improvement in learning grammatical structures. Moreover, the +WL group at the lower level benefitted more from producing WL episodes than the higher-level group. Her findings supported the impact of producing WL on L2 grammar learning.

On the same track, in a pretest-posttest experimental study, Nicolás–Conesa et al. (2019) examined the complex interaction among the type of WCF (direct vs. indirect), type of committed errors (grammatical vs. non-grammatical), and the WCF perspective (accuracy vs. acquisition). The researchers recruited 46 Spanish EFL university students and assigned them to two experimental groups that were required to produce WL episodes after they received direct and indirect WCF and a control group that produced WL and revised their written assignment without receiving WCF. They concluded that (1) both direct/indirect WCF groups equally corrected more errors than the control group, (2) the direct WCF group outperformed the indirect WCF group in revising grammatical and non-grammatical errors, and (3) all experimental and control groups reduced the percentage of committed errors from the pretest to the posttest. The authors emphasized the positive and integrated impacts of WCF and WL on L2 writing improvement. In another study, Shekary and Tahririan (2006) required a group of 16 Iranian EFL learners at mixed language proficiency levels to complete a dictogloss, a jigsaw, and a collaborative writing task in a chatroom environment. The researchers attributed the students' large number of WL episodes, successful resolution of errors, and in-depth engagement in the text-based medium of collaboration to the hybrid nature of the CMC environment which enhanced the process of noticing in L2 learners. Finally, Suzuki and Itagaki (2009) promoted languaging as a task-induced verbalization that actively engages L2 learners in the production of the target language. They examined 114 Japanese EFL learners who were required to produce WL while they were completing one of the two comprehension

and production tasks of translation. By analyzing the WL episodes and students' task outcomes, the authors reported the mediating role of WL and a significant interaction between the task types and the participants' level of language proficiency.

To sum up, SLA researchers have lately shown a growing interest in the constructive role of WL in L2 learning improvement. Accordingly, they have promoted the mediating role of WL and a significant interaction between the WL and L2 language proficiency, corrective feedback, communication modality, and task types.

Output Orientation in Form-Focused Writing Tasks

To emphasize the role of language output as a socially-constructed cognitive tool, Swain (1985) proposed her Comprehensible Output Hypothesis and argued that producing target forms can serve as "a trigger that forces the learner to pay attention to the means of expression to successfully convey his or her intended meaning" (p. 249), and "the act of producing language constituents, under certain circumstances, as part of the focusing process in language learning" (2005, p. 471). Swain (1995) moved forward with three major functions of language output in L2 learning, namely, noticing, hypothesis testing, and metalinguistic reflection. While producing language output, L2 learners notice the gap in their interlanguage and try to amend it (Pourdana & Rad, 2017). They also attempt to generate the L2 forms as hypothetically accurate statements which can raise their focus on form (Schmidt, 2001; Wang, 2019). Moreover, language output engages them in metalinguistic reflections to negotiate the meaning and the form.

Language output plays a more critical role in performing form-focused tasks. In form-focused tasks, L2 learners are required to (1) notice the form-meaning mapping in the L2, (2) integrate the language processing into producing the L2 form, and (3) juxtapose their interlanguage forms with the target language model (Izumi & Bigelow, 2000).

Translation as a Form-Focused Writing Task

A controversial form-focused pedagogical task, translation seems indispensable in most bilingual educational contexts (Cerezo *et al*., 2019). A translation task is known as a type of content-controlled writing task (Ishikawa, 2018) that intensely demands the L2 learners to focus on the linguistic aspects of their writing output more than the content of the message they convey (Nord, 2005; Pym, 2003).

Translation tasks have been recommended by several SLA researchers (e.g., Canagarajah, 2013; Moradian et al., 2017; Pennycook, 2008) in L2 learning environments where the teachers and students share the same L1. In 2005, Nord introduced the functional model of translation and argued that contrastive analysis of the structural similarities and differences between the source language and the target language can become a source of metalinguistic awareness which improves task engagement and L2 learning. In other words, the L2 learners' mother tongue can play the critical role of a metacognitive mediator (Pourdana et al., 2014) and a regulating device, because the "utterances in L1 mediate the cognitive processes that L2 learners need in general problem-solving tasks" such as translation (Antón & DiCamilla, 1999, p. 238).

By adopting Nord's functional approach, SLA researchers responded to the L2 practitioners' growing interest in translation tasks while recommending the translation of "carefully selected, authentic texts with a clear context and purpose" (Károly, 2014, p. 90). As a result, the dynamic code-switching strategies in translation tasks can have the potential to draw the L2 learners' focus on form, make a decision, foster self and collective scaffolding, and externalize their metatalk through languaging (Kazemi et al., 2022; Keshanchi et al., 2022; Rodrick Beiler & Dewilde, 2020).

Computed-Mediated Written Languaging

Computer-mediated communication (CMC) advanced technologies have changed the face of education by creating a learner-centered environment that eliminates the apprehension, embarrassment, and stress that students usually feel when participating in real classroom discussions (Pourdana, 2022; Pourdana & Tavassoli, 2022; Rafi et al., 2022). The CMC can serve L2 learners with a platform to share comments, generate content, and exchange feedback with peers and teachers, either synchronously or asynchronously (Abe, 2020). Moreover, the CMC can offer multimedia modalities in e-learning contexts in terms of downloadable texts and audio/video files.

Faster than expected, e-learning has turned into a reality in developing countries despite serious challenges of outdated infrastructure, the digital divide, teachers' low expertise, and learners' inadequate computer literacy (Engerer, 2020; Lin, 2020). The situation became even worse with the COVID-19 pandemic crisis which pushed many students to distance education. In this adverse situation, some user-friendly CMC interfaces such as Google Docs have reached a peak in popularity.

A CMC Web 2.0-word processor, Google Docs is an interactive online context that allows L2 learners to generate, edit and share their Word processing documents. The potential of Google Docs includes being associated with other Google tools such as Google Translate and Google Sheets, browsing document folders in Google Drive, managing documents by

sharing them, and free-of-charge accessibility (Ebadi & Rahimi, 2017; Yamashita, 2021).

In the research literature on languaging, Kazemi et al. (2022) are among the few who focused on computer-mediated languaging in L2 writing tasks. To compare the impacts of collaborative writing and mediation modalities in determining the WL attributes in Google Docs, the researchers selected 68 EFL learners. The participants were paired and divided into two advanced groups who collaborated on form-focused and content-focused writing tasks while simultaneously producing WL episodes. Both groups received teacher-made and automated mediation by Google Docs on their task performance. The findings indicated that the group that worked on the form-focused writing task generated more WL episodes than the other, while both groups focused more on lexis than grammar. The researchers also promoted the role of Google Docs' automated mediation in causing a more successful resolution of WL episodes than teacher-led mediation.

The Current Study

The languaging literature lacks enough research on the interaction between the attributes of WL episodes and the output orientation of the online writing tasks. This study addressed this large gap by examining (1) the extent to which the online production of WL episodes and the output orientation of the form-focused writing tasks (i.e., production-based vs. comprehension-based) might interact to facilitate the EFL learners' writing improvement, and (2) the extent to which the attributes of WL episodes (i.e., quantity and focus) might be determined by the output orientation of writing tasks.

This study was led by the following research questions:

(1) Is there any potential interaction between producing WL episodes and the output orientation of the form-focused writing tasks to affect the EFL learners' task outcomes?

(2) How does the output orientation of the form-focused writing tasks affect the attributes of WL episodes in terms of quantity and focus?

METHOD

Context and Participants

This research was conducted in the middle of the COVID-19 pandemic in Iran. A group of 130 university students majoring in English Translation Studies volunteered to partake in this experiment. They were Persian-speaking (L1) EFL learners whose ages ranged from 20 to 25 (M = 22.08, SD = .31) and the length of their formal exposure to English was 12 years on average.

The participants received bilingual instructions in English and Persian in mandatory university courses such as academic writing, translation of journalistic texts, and oral interpretation. The researcher's criteria for participation in the research included signing a consent form, owning a smartphone, accessing the Internet, and registering with Google Docs. Also, participants were required not to use dictionaries to complete the assigned tasks.

The experiment was announced as an extra-curricular university program which received a huge response from the volunteers at the university campus. The convenience sampling method (Best & Kahn, 2006) was used to select the participants who were available and willing to participate (N = 112, 100 females, 89.28%, and 12 males, 10.71%). To control the confounding effect of the participants' English proficiency level on the quantity of WL episodes (Ishikawa & Suzuki, 2016), it was planned to select a uniform sample of students at the advanced level of general English. To do so, the Oxford Placement Test (OPT, Version 1.1, 2001) was administered. Due to the restrictions imposed by the COVID-19 lockdown, the 60 items of the test were converted into online Google Forms.

The volunteers whose proficiency level was determined as advanced by their OPT scores were selected (48-54, C1 on OPT rating scale) (M = 52.00, SD = .601, Cronbach's α = .860, representing strong inter-item reliability). Next to administrating the OPT and excluding 18 volunteers whose scores were below the threshold, the 112 remaining participants were randomly assigned to two groups who worked on production-based tasks with and without producing WL (hereafter, +WLPT and -WLPT) and two groups who worked on comprehension-based tasks with and without producing WL (hereafter, +WLCT and -WLCT). After assigning the participants to their respective groups, the OPT mean scores of the four groups were compared and no significant inter-group differences were found (F (3, 109) = 1.39, p = .79; Partial η^2 = .006, interpreted as a weak effect size).

Assessments and Measures

Assessment Tasks

Three translation tasks were developed which functioned as the pretest, a production-based posttest, and a comprehension-based posttest. The task content was paralleled by the situationally-relevant topic of COVID-19 to balance their cognitive loads. The pretest was a Persian-to-English translation of a descriptive passage prompted with a bar graph about the marriage rate in the last decade in Iran (counted words = 92, M = 23 per sentence). The participants were

required to translate the passage in 30 minutes on Google Docs without using dictionaries.

The production-based posttest was a gap-filling translation task, and the comprehension-based posttest included an error-identification task of translation. The logic behind developing gap-filling and error-identification tasks was their efficiency in representing the contextualized form-focused writing tasks with comprehension-oriented and production-oriented outputs, respectively (Brown & Abeywickrama, 2018; Pica et al, 2006; Purpura, 2014).

The content of the posttests was adopted from a medical report on the susceptibility of ABO blood types to COVID-19 infection by ALKhikani (2020). The readability index of the adopted text was measured as 57.30 (i.e., fairly difficult to read) on the Flesch Reading Ease Scoring system. The pretest and the posttests were pilot-studied with a random group of 18 undergraduate students and co-rated by the researcher and her two research assistants (the inter-rater agreement being 95.40%). The responses were measured dichotomously (1 for a correct response, 0 for an incorrect response). Grammatical accuracy was the only standard for the correctness of the responses to the pretest and posttests.

Treatment Tasks

Eight authentic Persian texts were extracted from www.mehrnews.com, an official news network headquartered in Tehran, Iran. They were incorporated into eight parallel tasks of gap-filling (production-based) and error-identification (comprehension-based) as the treatment tasks in this study. The topics such as marriage, unemployment, national festivities, COVID-19, and Persian arts were selected with a 95.08% familiarity index after a topic familiarity questionnaire was administered to the participants. The treatment tasks were piloted with a total of 49 undergraduate students similar to the main sample of participants.

The error-identification tasks consisted of a translated passage with some selected grammatical (e.g., verb tense, preposition) (N = 6), and lexical choices (e.g., adverb of manner, adjective) (N = 6). The choices were numbered and the participants had to decide on their grammatical accuracy. For example, the initial sentence in the error identification task # 5 was:

> *This picture [1] reports that the ABO blood types plays [2] roles in susceptibility [3] to COVID-19.*

[1] Correct/Incorrect	*[2] Correct/Incorrect*	*[3] Correct/Incorrect*

in which the participants had to highlight one of the correct/incorrect options for each item on Google Docs. Only the +WLCT group had to complete the task while producing WL episodes. The WL episodes could be written on the left or right margin of the text layout (Appendix A). The selected responses were scored collaboratively by the researcher and her assistants (Cronbach's α = .982, representing strong inter-item reliability).

The parallel gap-filling tasks were spotted with 12 blanks that corresponded to similar items in the error identification tasks. The six grammatical and six lexical missing words were numbered and the participants were required to fill all the gaps on Google Docs. While the +WLPT group was producing WL episodes and completing the gap-filling task, the –WLPT group had to work on the task without languaging. The first sentence in the gap-filling task # 5 was:

> *This .....[1] reports that the ABO blood types .....[2] roles in .....[3] to COVID-19.*

In rating the produced responses to the gap-filling task, the appropriate-word scoring method was adopted to give credit to a wider range of responses that fitted the blanks (Greene, 2001). The misspellings and typographies were not penalized unless they were illegible (N = 4). The written responses were also co-rated by the researcher and her assistants (Cronbach's α = .982, representing strong inter-rater agreement) (Appendix B). The disagreements in the rating procedure were resolved case-wise reaching a full consensus.

Procedure

Informed by Kim and McDonough (2011) who provided evidence for the mediating role of the pretask modeling, the researcher believed that the tutorial was necessary for the participants who were unfamiliar with the notion of virtual WL production. Therefore, the study commenced with a 90-minute webinar on *Virtual Written Languaging on Google Docs* in Week 1. Upon the announcement in June 2020, 130 volunteers signed up for the Zoom video conferencing platform and used their Google email (Gmail) accounts to sign in to Google Docs.

In the Zoom meeting sessions, the researcher began to share the screen with the audience into a document file generated on Google Docs and live-streamed the virtual WL performance on various types of translation tasks. The instructions were also provided to the rating system. The participants were suggested to notice the automated error detection by Google Docs and try to apply the comments to resolve their errors. Through waved underlining, for example, Google Docs provided feedback on grammatical or lexical errors, misspellings, and abbreviated or non-English words (e.g., *Goole* for Google or *COVID*). Next to the researcher's tutorial, four volunteers screen-shared their WL production on translation tasks and received the audience's feedback.

After the OPT administration and grouping in Week 2, the participants completed the pretest in 30 minutes in the fol-

lowing week. For the next eight weeks (Weeks 4 to 11), the +WLCT and +WLPT groups completed treatment tasks while producing online WL episodes, and the –WLCT and –WLPT groups performed similar tasks without producing WL. All assignments were carried out on Google Docs. The online weekly sessions were 30 minutes long, in which samples of the participants' task outputs were rated by the researcher and discussed with the group. The participants in the +WLCT and +WLPT groups were frequently encouraged to produce WL episodes in English. Nonetheless, Persian WL episodes were accepted and encoded similarly to the English WL episodes. In Week 12, the WLCT groups completed a 30-minute error identification task, in parallel to the WLPT groups who completed a 30-minute gap-filling task on Google Docs. The whole procedure was illustrated in Figure 1.

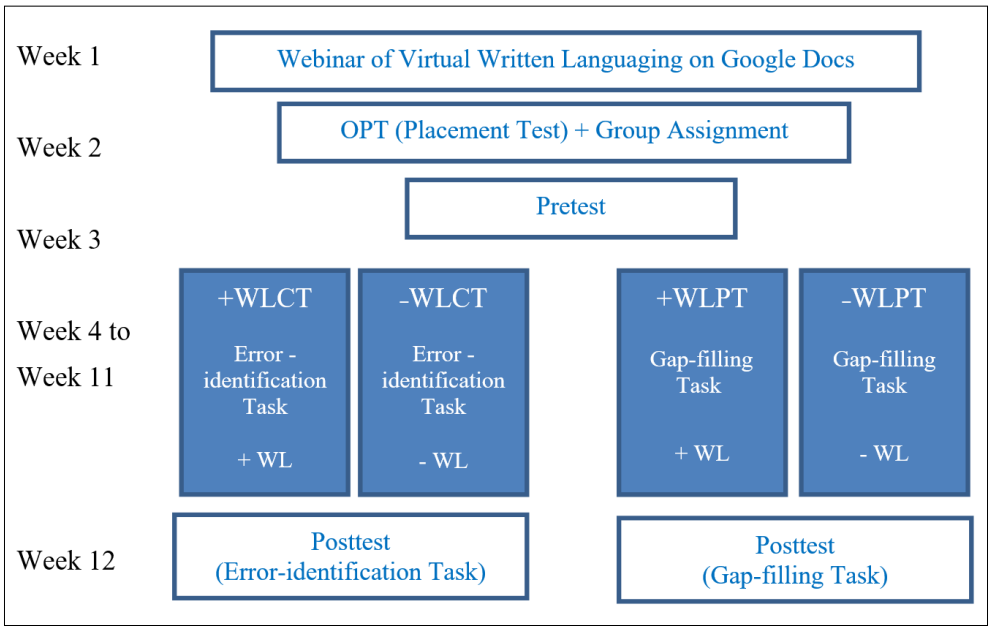
Data Analysis

A quantitative scheme of translation assessment was adopted in this study in terms of measuring the correct response percentages on all the treatment tasks, pretest, and posttests. Accordingly, the error detection was limited to the percentages of committed errors on grammatical structures and lexical choices to maintain objectivity in data collection and reliability of the rating system.

Moreover, the participants' production of WL episodes by the +WLCT and +WLPT groups was analyzed in terms of the quantity and the focus of the WL episodes. To encode and quantify the WL episodes, Swain's operational definition of languaging was adopted as "any segment of a dialogue where students [self-]talk about the language they are producing, question their language use, or other- or self-correct their language production" (Swain, 2000, p. 287).

Figure 1

The Flow of the Procedure



Further, the WL episodes were analyzed for their type of focus by using the languaging classification in Yang (2016): (1) Grammar-based WL (G-WL), including the WL segments dealing with aspects of morphology or syntax, such as articles, tense, or subject-verb agreement; (2) Lexis-based WL (L-WL), including the WL segments dealing with lexical choices, or equivalents

RESULTS

In a pretest-treatment-posttest design, this study addressed whether (1) producing online WL episodes and the output orientation of the translation tasks would compete to affect the EFL learners' task performance and whether (2) the output orientation of the translation tasks would determine the attributes of the WL episodes. The collected data to answer the first research question included the ±WLCT and ±WLPT groups' pretest and posttest scores, while the data for the second research question were collected through the content analysis of the WL episodes produced by only +WLPT and +WLCT groups.

Impact of WL and Output Orientation on Task Outcomes

The pretest and posttest scores of the four groups were inserted into Statistical Package for Social Sciences (SPSS) version 25 for running the tests of normality and descriptive statistics. The level of significance was set at α = .05 for all tests (Table 1).

As Table 1 indicates, the ratios of skewness and kurtosis for all test results were inside the ± 1.96 intervals to retain

the normality of the data (Byrne, 2010). The descriptive statistical analysis of the pretest and the posttests is summarized in Table 2. Accordingly, the highest mean score on the posttest belonged to the +WLPT group (M = 10.85, SD = .80), followed by the +WLCT group (M = 8.92, SD = .81). The other two groups (–WLCT and –WLPT) also showed improvement on the posttests, yet their improvements were not as noticeable as in the +WL counterparts.

To move forward to the inferential statistics, an analysis of covariance (ANCOVA) was run with the mean scores of the four groups on the posttests, when the condition of ±WL (with and without producing WL) was set as the between-group variable (Table 3). The logic behind running the ANCOVA was to control the initial within-group differences in the participants of the four groups which were measured by the pretest (covariate).

The assumptions of ANCOVA were met in this study, including the homogeneity of variances for the groups (indicated by the ratios of skewness and kurtosis in Table 1), the non-significant interaction between the covariate and the dependent variable (i.e., the posttests scores) (F (3, 104) = .780, p = .010, Partial η^2 = .016, interpreting as a weak effect size), and the homogeneity of variances measured by Levene's Test of Equality of Variances (F (3, 108) = .241, p = .867).

According to Table 3, the test of ANCOVA produced the main effect of the ±WL condition which caused significant differences among the mean scores on the posttests (F (3, 107) = 221.583, p = .000, partial η^2 = .861, representing a large effect size), after controlling for the covariate effect of the pretest.

To further explore the between-group differences, a series of Sheffé post hoc tests were carried out to examine the extent to which the posttests gain scores were different across groups. The results reported significant differences, with the largest gap between +WLPT and –WLPT groups (Mean Difference = 5.28, p < .005, 95% CI [4.69 – 5.78]), and the smallest gap between +WLCT and –WLCT (Mean Difference = 1.35, p < .000, 95% CI [.76 – 1.94].

As Figure 2 illustrates, the +WLPT and +WLCT groups improved their writing performance noticeably from the pretest to the posttest (5.10 to 10.85, and 5.49 to 8.92, respectively). Regarding the –WL groups, while the –WLCT group

Table 1

Normality of the Pretest and Posttest Scores (Four Groups)

			Skewness			Kurtosis		
Groups	**Test**	**WL**	**Statistics**	**Std. Error**	**Ratio**	**Statistics**	**Std. Error**	**Ratio**
WLCT	Pre	+	-.693	.441	-1.57	1.261	.858	1.46
	Post	+	.581	.441	1.31	-.012	.858	.01
WLPT	Pre	+	.482	.441	1.09	1.101	.858	1.28
	Post	+	.273	.441	.61	-.379	.858	-.44
WLCT	Pre	-	-.121	.441	-.27	1.198	.858	1.39
	Post	-	-.610	.441	-.07	.191	.858	.22
WLPT	Pre	-	.377	.441	.85	-.304	.858	-.35
	Post	-	.560	.441	1.26	1.196	.858	1.39

Table 2

Descriptive Statistics of the Test Scores

Groups	**Test**	**WL**	**Mean**	**Std. Deviation**	**95% CI**
WLCT	Pre	+	5.49	.17	[5.42 – 5.55]
	Post	+	8.92	.81	[8.61 – 9.24]
WLPT	Pre	+	5.10	.21	[5.02 – 5.19]
	Post	+	10.85	.80	[10.54 – 11.16]
WLCT	Pre	-	5.67	.14	[5.62 – 5.73]
	Post	-	7.57	.69	[7.30 – 7.83]
WLPT	Pre	-	5.49	.21	[5.41 – 5.57]
	Post	-	5.77	.79	[5.26 – 5.87]

Table 3

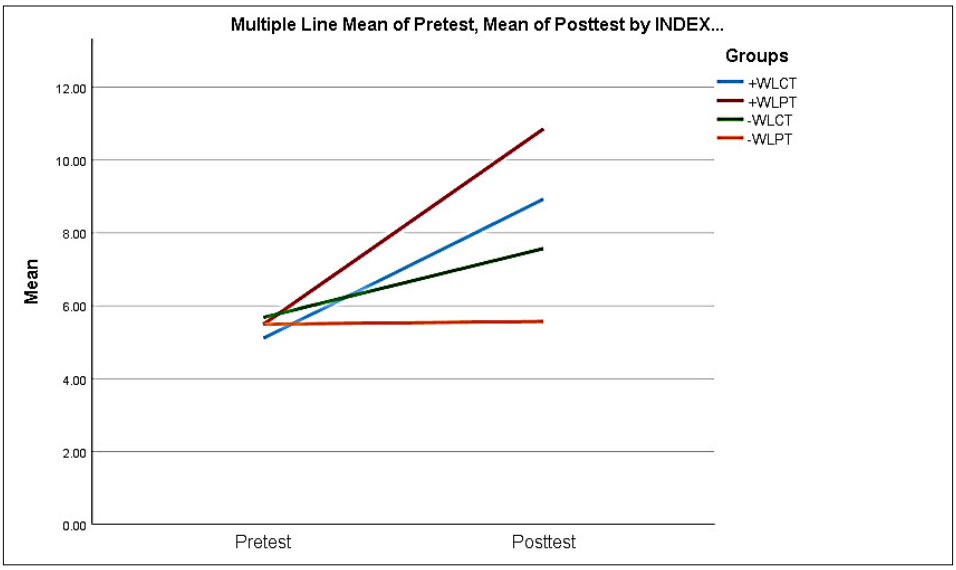
Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial η^2
Corrected Model	417.008[a]	4	104.252	171.729	.000	.865
Intercept	7.714	1	7.714	12.708	.001	.106
Pretest	.043	1	.043	.071	.790	.001
Groups	403.550	3	134.517	221.583	.000	.861
Error	64.957	107	.607			
Total	8072.000	112				
Corrected Total	481.964	111				

Note: R Squared = .865 (Adjusted R Squared = .860)

Figure 2

Estimated Marginal Means



had a moderate improvement, the performance of the –WLPT group showed a small progress from the pretest to the posttest (5.67 to 7.57 and 5.49 to 5.77, respectively).

Impact of Output Orientation of the Tasks on WL Attributes

To address the second research question, the quantity of WL episodes and their focus on grammar (G-WL episodes) and lexis (L-WL episodes) produced by the +WLPT and +WLCT groups were summed up.

Every occurrence of the G-WL and L-WL episodes was treated as one test score. Accordingly, the total number of WL episodes produced by the +WLPT group (N = 847) was more than doubled in size of the WL episodes produced by the +WLCT group (N = 405). Moreover, in both +WLCT and +WLPT groups, the tallies of the L-WL episodes (501, 54.75% to 300, 72.28%) were much higher than those of the G-WL episodes (346, 45.25% to 115, 27.71%). The descriptive statistics for G-WL and L-WL episodes are presented in Table 4.

As Table 4 reports, the ratios of skewness and kurtosis were within the ± 1.96 intervals as the indication of normal distribution. Moreover, the average number of L-WL episodes produced by both +WLCT and +WLPT groups (M = 10.71, SD = 1.30, and M = 17.89, SD = 1.10, respectively) showed a larger distribution of L-WL episodes than G-WL episodes (M = 4.10, SD = 1.42, and M = 12.35, SD = 1.41, respectively) in both groups. To further explore the differential impacts of the output orientation of the treatment tasks on the focus of WL episodes, a one-way analysis of variance (ANOVA) was carried out (Table 5).

As Table 5 indicates, the main effect of the output orientation on the average number of G-WL and L-WL episodes was reported as significant (F (3, 108) = 508.78, p = .000, partial η^2 = .320, representing a large effect size). In other words, the difference between the proportions of G-WL and L-WL episodes on gap-filling and error identification tasks was determined by their output orientation.

Table 4

Descriptive Statistics for G-WL and L-WL Episodes

				Skewness			Kurtosis		
Group	**WL**	**M**	**SD**	**Statistics**	**Std. Error**	**Ratio**	**Statistics**	**Std. Error**	**Ratio**
+WLCT	G-WL	4.10	1.42	-.285	.441	-.641	-1.198	.858	-1.396
	L-WL	10.71	1.30	-.379	.441	-.859	1.012	.858	1.179
+WLPT	G-WL	12.35	1.41	.377	.441	.854	-.304	.858	-.354
	L-WL	17.89	1.10	.046	.441	.104	-.478	.858	-.557

Table 5

Mean Comparison of WL Episodes

	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**	**Partial η^2**
Between Groups	2706.464	3	902.155	508.787	.000	.320
Within Groups	191.500	108	1.773			
Total	***2897.964***	***111***				

DISCUSSION

The first research question in this study queried whether producing WL and the task output orientation would have any interface to affect the L2 learners' performance on form-focused writing tasks. The findings were affirmative and in favor of the impact of WL production. In other words, the L2 learners who produced WL had better performance on both production-based gap-filling tasks and comprehension-based error-identification tasks, than those whose task output was without WL production.

The discussion of the first research question is anchored in Vygotsky's SCT argument and cognitive psychology of the mind. In the SCT framework, the distance between the L2 learners' potential and actual level of development is bridged by employing collective scaffolding of the teacher (the expert) to the learner (the novice) (Vygotsky, 1987) or the learners' self-scaffolding (Swain et al., 2009). In other words, the social or private speech exchanged over the form-focused tasks might be internalized and turned into metacognitive (i.e, languaging) output (Storch, 2013). Therefore, engagement of the L2 learners in producing languaging can mediate them to deepen their knowledge of linguistic form and "relocate the scaffolding agency from the expert to the learners to provide occasions for successful learning" (Knouzi et al., 2010, p. 26).

Moreover, the ANCOVA statistical results indicated that the participants in +WL groups who produced responses to the gap-filling tasks outperformed those who selected the correct responses to the error identification tasks. From the cognitive psychology perspective, while the error identification task can challenge only the L2 learners' existing grammatical and lexical knowledge, the gap-filling task demands both their existing knowledge and their ability to produce correct responses (Purpura, 2014). Therefore, the production-based gap-filling tasks could trigger more WL episodes than the comprehension-based error identification tasks and eventually higher task achievement (Falhasiri, 2021; Van Patten, 2015a).

Several studies have echoed the mediating role of WL in L2 learners' outperformance on various types of tasks similar to the results of the study (Manchón et al., 2020; Pica et al., 2006; Storch, 2013; Suzuki & Itagaki, 2009; Swain et al., 2009). The findings in this study partially contradict Ishikawa (2013) who required 18 Japanese EFL learners to complete a Japanese-to-English translation task with/without languaging. In the face-to-face classroom setting, Ishikawa assigned the participants to the treatment and control groups. It was reported that the languaging participants produced more L-WL than G-WL, while no differences were observed between their posttest scores and those in the control group. Such unexpected results were arguably due to several possible reasons, such as the small size of participants in each group, and the non-experimental nature of the study.

The second research question addressed the potential impact of the output orientation of the form-focused writing tasks on the quantity and focus of the WL episodes. The results indicated that the output orientation of translation tasks could only determine the quantity of the WL episodes in favor of production-based gap-filling tasks. In other words, the L2 learners who performed production-based gap-filling tasks produced more WL than those who performed comprehension-based error-identification tasks. Yet both groups focused more on lexis (L-WL) than grammar (G-WL) while they were producing WL and performing writing tasks.

The discussion of the second question is two-fold. On one hand, the results are supported by the fundamental argument in Swain's Comprehensible Output Hypothesis which accounts for producing a larger body of WL episodes by L2 learners who are working on production-based tasks such as gap-filling (Benati, 2017; Swain, 2005). On the other hand, Swain anticipated that L2 learners would generate more grammar-focused than lexis-focused language output when they work on form-focused writing tasks which seemed contradictory to the findings in this study. Despite the soundness of this argument, the *languaging output* may not necessarily function as the *language output* which is the core subject matter in Swain's Output Hypothesis. In other words, the gap-filling tasks might still require the L2 learners' languaging over the lexical points (L-WL) more than grammatical points (G-WL).

Several SLA researchers compared the impacts of different task types on the amount of WL episodes (García Mayo, 2002; Niu, 2009) and supported the potential of the production-based tasks to induce a larger body of languaging episodes. The findings in this study are consistent with Suzuki (2012), but partially endorse Yang (2016). To examine the impact of receiving form-focused CF on learning English grammar, Suzuki (2012) examined 24 Japanese EFL learners who worked on a production-based writing task and received CF exclusively on their grammatical errors. The participants were required to produce WL on the received CF and revise their writings. Consistent with the findings in the current study, Suzuki reported a huge number of students' WL episodes as a result of working on form-focused production-based writing tasks and the mediating role of CF in their writing task achievement. On the other hand, Yang (2016) explored the role of WL to facilitate eight Chinese EFL learners' performance on a story re-writing task (i.e., a production-based oral task). The participants' engagement in languaging was later analyzed in terms of the rates of L-WL and G-WL episodes. Similar to the findings of the current study, Yang reported that when the participants engaged in the re-telling stage, they produced a large number of L-WL episodes. Yet, in the comparing and revising stages of the story re-writing task, they switched to generating more G-WL episodes.

The relative advantage of the groups whose task performance concurred with producing WL episodes (+WLCT and +WLPT groups) can be further discussed in light of the learner-centered and user-friendly potential of the Google Docs platform. As a web-based Word processor, Google Docs mediated all participants through the visual display of their writing task outputs. Therefore, it simultaneously provoked a large amount of WL episodes in the two +WL groups who could visualize their WL episodes as well. As a result, Google Docs could benefit the +WL groups with more opportunities to notice their linguistic errors, self-repair, and deliberate (Falhasiri, 2021; Kazemi et al., 2022; Yamashita, 2021).

CONCLUSION

This study provided strong evidence for the interaction between the WL production and task output orientation and their potential to improve EFL learners' performance on form-focused writing tasks (i.e., translation). It was found that producing WL episodes could mediate the L2 learners' performance on both production-based and comprehension-based writing tasks. Moreover, the WL was provoked more rigorously by working on production-based than comprehension-based writing tasks, while the focus of WL episodes was more directed to lexis than grammar on both types of form-focused writing tasks.

In consequence, several pedagogical implications can be proposed. The L2 teachers are recommended to create opportunities for languaging by training the students how to language in oral and written modes. For more benefits, teacher-imposed languaging might be incorporated as an independent task rather than the by-products of a pedagogical task.

The L2 teachers are also recommended to adopt an alternative approach to translation as a constructive form-focused writing task. They can make the best use of translation as a semanticizer in teaching new forms in the target language, and as a forum to raise more awareness of the complex linguistic structures. More importantly, translation can generate a large amount of languaging and mutually become improved through languaging. Therefore, in the L2 contexts where students and teachers share similar L1, such as in Japan or Iran, translation can play an effective pedagogical role. Finally, by blending languaging and a variety of form-focused tasks on CMC platforms such as Google Docs, Google Meet, and Zoom, both L2 teachers and learners can enjoy the synchronicity, simplicity, and practicality of their invaluable resources.

The arguments in this research are still inconclusive due to several limitations. The COVID-19 pandemic was the major obstruction that caused countless adjustments to the procedures of selecting participants, collecting data, and setting follow-up discussions with the research team. The second limitation was the non-random method of sampling. The participants were English-major EFL university students at high levels of language proficiency who volunteered to collaborate. The researcher speculated that the participants' high level of task engagement and enthusiasm, advanced language proficiency, and educational background in English Translation Studies could have confounded the findings. Therefore, the generalizability of the results should be done with precautions. From the academic research design perspective, the researcher did not plan to examine the sustainability of the impact made by WL production on writing task performance by running delayed posttests. Neither did she isolate the moderating effect of the translation tasks by in-

cluding comparison groups who could complete other types of form-focused writing tasks. Beyond the limitations in this study, the reported findings reasonably suggest that if WL is carefully designed in online modalities and/or face-to-face context of language learning, it can provide optimal opportunities for L2 writing improvement.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

Abe, M. (2020). Interactional practices for online collaborative writing. *Journal of Second Language Writing, 49*, 100752. https://doi.org/10.1016/j.jslw.2020.100752

AL-Khikani, F. (2020). Surveillance 2019 novel coronavirus (COVID-19) spreading: Is a terrifying pandemic outbreak is soon? *Biomedical and Biotechnology Research Journal, 4*(1), 81–82. https://doi.org/10.4103/bbrj.bbrj_16_20

Ammar, A., & Hassan, R. M. (2018). Talking it through: Collaborative dialogue and second language learning. *Language Learning, 68*(1), 46-82. https://doi.org/10.1111/lang.12254

Antón, M., & DiCamilla, F. (1999). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. *Canadian Modern Language Review, 54*, 314-342. https://doi.org/10.1111/0026-7902.00018

Azkarai, A., & Kopinska, M. (2020). Young EFL learners and collaborative writing: A study on patterns of interaction, engagement in LREs, and task motivation. *System, 94*, 102338. https://doi.org/10.1016/j.system.2020.102338

Bataineh, R. F., Al-Qeyam, F. R. & Smadi, O. M. (2017). Does form-focused instruction really make a difference? Potential effectiveness in Jordanian EFL learners' linguistic and pragmatic knowledge acquisition. *Asian-Pacific Journal of Second and Foreign Language Education, 17*(2). https://doi.org/10.1186/s40862-017-0040-0

Behbahani, S. M. K., Pourdana, N., Maleki, M., & Javanbakht, Z. (2011). EFL task-induced involvement and incidental vocabulary learning: succeeded or surrounded. *In International Conference on Languages, Literature and Linguistics. IPEDR Proceedings, 26*, 323–325.

Benati, A. (2017). The role of input and output tasks in grammar instruction: Theoretical, empirical and pedagogical considerations. *Studies in Second Language Learning and Teaching, 7*(3), 377-396. https://doi.org/10.14746/ssllt.2017.7.3.2

Best, J. W. & Kahn, J. V. (2006). *Research in education (3rd ed.).* Pearson Education Inc.

Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge Taylor & Francis Group.

Canagarajah, S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations.* Routledge.

Cerezo, L., González-Cruz, B. & Mercader, J. (2019). 11. English as a foreign language writing teacher education and development in Spain: The relevance of a focus on second language writing as a tool for second language development. In L. Seloni & S. Henderson Lee (Eds.), *Second Language Writing Instruction in global contexts: English language teacher preparation and development* (pp. 222-249). Multilingual Matters. https://doi.org/10.21832/9781788925877-015

Chi, M. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 161-238). Erlbaum.

Ebadi, S., & Rahimi, M. (2017). Exploring the impact of online peer-editing using Google Docs on EFL learners' academic writing skills: A mixed-methods study. *Computer Assisted Language Learning, 5.* https://doi.org/10.1080/09588221.2017.1363056

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning, 51*(1), 1–46. https://doi.org/10.1111/j.1467-1770.2001.tb00013.x

Engerer, V. P. (2020). Temporality revisited: Dynamicity issues in collaborative digital writing research. *Education and Information Technologies*, 339–370. https://doi.org/10.1007/s10639-020-10262-9

Esfandiari, R., & Noor, P. (2018). Iranian EFL raters' cognitive processes in rating IELTS speaking tasks: The effect of expertise. *Journal of Modern Research in English Language Studies, 5*(2), 41-76. https://doi.org/10.30479/jmrels.2019.9383.1248

Falhasiri, M. (2021). Written languaging and engagement with written corrective feedback: The results of reflective teaching. *TESL Canada Journal, 38*(1), https://doi.org/10.18806/tesl.v38i1.1348

García Mayo, M. P. (2002). The effectiveness of two form-focused tasks in advanced ESL pedagogy. *International Journal of Applied Linguistics, 12*(2), 156 - 175.

Greene Jr., B. B. (2001). Testing reading comprehension of theoretical discourse with cloze. *Journal of Research in Reading, 24*(1), 82-98. https://doi.org/10.1111/1467-9817.00134

Heidari, S., Pourdana, N., & Khalili, G. F. (2019). Retrospective and introspective think-aloud protocols in translation quality assessment: A qual-quan mixed-method research. *Language and Translation*, *9*(4), 1-16.

Ishikawa, M. (2013). Examining the effect of written languaging: The role of metanotes as a mediator of second language learning. *Language Awareness, 22*(3), 220-233. https://doi.org/10.1080/09658416.2012.683435

Ishikawa, M. (2018). Written languaging, learners' proficiency levels and L2 grammar learning. *System, 74,* 50-61. https://doi.org/10.1016/j.system.2018.02.017

Ishikawa, M., & Suzuki, W. (2016). The effect of written languaging on learning the hypothetical conditional in English. *System, 58*, 97-111. https://doi.org/10.1016/j.system.2016.02.008

Izumi, S. & Bigelow, M. (2000). Does output promote noticing in second language acquisition? *TESOL Quarterly, 34*, 239-278. https://doi.org/10.2307/3587952

Jang, Y., Lee, H., Kim, Y., & Min, K. (2020). The relationship between metacognitive ability and metacognitive accuracy. *Metacognition Learning, 15*, 411–434. https://doi.org/10.1007/s11409-020-09232-w

Károly, K. (2014). Translation in foreign language teaching: A case study from a functional perspective. *Linguistics and Education, 25,* 90-107. https://doi.org/10.1016/j.linged.2013.09.009

Karpov, V. (2013). A way to implement the neo-Vygotskian theoretical learning approach in the schools. *International Journal of Pedagogical Innovations, 1*(1), 25-35. https://doi.org/10.12785/ijpi/010104

Kazemi, P., Pourdana, N., Khalili, G. F. (2022). Microgenetic analysis of written languaging attributes on form-focused and content-focused e-collaborative writing tasks in Google Docs. *Education and Information Technologies, 27,* 10681–10704*.* https://doi.org/10.1007/s10639-022-11039-y

Keshanchi, E., Pourdana, N. & Khalili, G. F. (2022). Differential impacts of pair and self-dynamics on written languaging attributes and translation task performance in EFL context. *English Teaching & Learning*. https://doi.org/10.1007/s42321-022-00117-6

Kim, Y., & McDonough, K. (2011). Using pretask modeling to encourage collaborative language learning opportunities. *Language Teaching Research, 15*(2), 183–199. https://doi.org/10.1177/1362168810388711

Knouzi, I., Swain, M., Lapkin, S., & Brooks, L. (2010). Self-scaffolding mediated by languaging: Microgenetic analysis of high and low performers. *International Journal of Applied Linguistics, 20*(1), 23-49. https://doi.org/10.1111/j.1473-4192.2009.00227.x

Lantolf, J., Thorne, S. L., & Poehner, M. (2015). Sociocultural Theory and Second Language Development. In B. Van Patten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 207-226). Routledge.

Lavasani, M., Latif, F., & Rahimi, A. (2021). Languaging in L1 as a mediating factor in EFL listening skill enhancement: Private vs. collaborative form. *Research in English Language Pedagogy, 9*(1), 215-234. https://doi.org/10.30486/RELP.2021.1909754.1222

Manchón, F., Nicolás-Conesa, F., Cerezo, L., & Criado, R. (2020). L2 writers' processing of written corrective feedback: Depth of processing via written languaging. In W. Suzuki & N. Storch (Eds.), *Languaging in language learning: A collection of empirical studies.* (pp. 242 – 265). John Benjamins.

Moradian, M. R., Miri, M., & Nasab, M. H. (2017). Contribution of written languaging to enhancing the efficiency of written corrective feedback. *International Journal of Applied Linguistics, 27*(2), 406-421. https://doi.org/10.1111/ijal.12138

Nguyen, L. (2020). A case study of Vietnamese EFL Teachers' conception of language output and interaction. *Journal of Language and Education*, *6*(1), 55-71. https://doi.org/10.17323/jle.2020.9777

Nicolás–Conesa, F., Manchón, R. M., & Cerezo, L. (2019). The effect of unfocused direct and indirect written corrective feedback on rewritten texts and new texts: Looking into feedback for accuracy and feedback for acquisition. *The Modern Language Journal, 103*(4). https://doi.org/10.1111/modl.12592

Niu, R. (2009). Effect of task-inherent production modes on EFL learners' focus on form. *Language Awareness, 18*(3-4), 384–402. https://doi.org/10.1080/09658410903197256

Nour, P., Esfandiari, R., & Zarei, A.A. (2021). Development and validation of a metamemory maturity questionnaire in the context of English as a foreign language. *Language Testing in Asia, 11*, 24. https://doi.org/10.1186/s40468-021-00141-6

Nord, C. (2005). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis* (2nd ed.). Rodopi.

Pennycook, A. (2008). English as a language always in translation. *European Journal of English Studies, 12*(1), 33–47. https://doi.org/10.1080/13825570801900521

Pica, T., Kang, H. S., & Sauro, S. (2006). Information gap tasks: Their multiple roles and contributions to interaction research methodology. *Studies in Second Language Acquisition, 28*(2), 301-338. https://doi.org/10.1017/S027226310606013X

Pym, A. (2003). Redefining translation competence in an electronic age: In defense of minimalist approach. *Meta: Translators' Journal, 48*(4), 481–497. https://doi.org/10.7202/008533ar

Pourdana, N. (2022). Impacts of computer-assisted diagnostic assessment on sustainability of L2 learners' collaborative writing improvement and their engagement modes. *Asian-Pacific Journal of Second and Foreign Language Education, 7*, 11. https://doi.org/10.1186/s40862-022-00139-4

Pourdana, N., Behbahani, S. M. K., & Safdari, M. (2011). The impact of task types on aspects of Iranian EFL learners' writing performance: Accuracy, fluency, and complexity. In *Proceedings of International Conference on Humanities, Society and Culture*. IACSIT Press.

Pourdana, N., Bornaki, F., Moayedi Fard, Z., & Sarkhosh, S. Z. (2012). Test-taking strategies and performance on reading comprehension tests by Iranian EFL learners. *International Journal of Applied Linguistics & English Literature, 1*(2),138–153. http://dx.doi.org/10.7575/ijalel.v.1n.2p.138

Pourdana, N., & Behbahani, S. M. K. (2013). Task types in EFL context: Accuracy, fluency, and complexity in assessing writing performance. *International Journal of Social Science and Humanity, 2*(1). https://doi.org/10.7763/IJSSH.2012.V2.73.

Pourdana, N., Sahebalzamani, S., & Rajeski, J. S. (2014). Metaphorical awareness: A new horizon in vocabulary retention by Asian EFL learners. *International Journal of Applied Linguistics and English Literature, 3*(4), 213–220. https://doi.org/10.7575/aiac.ijalel.v.3n.4p.213

Pourdana, N., & Rad, S. M. (2017). Differentiated instructions: Implementing tiered listening tasks in mixed-ability EFL context. *Journal of Modern Research in English Language Studies*, *4*(1), 87-69. https://doi.org/10.30479/JMRELS.2017.1566

Pourdana, N., & Asghari, S. (2021). Different dimensions of teacher and peer assessment of EFL learners' writing: Descriptive and narrative genres in focus. *Language Testing in Asia, 11*(6). https://doi.org/10.1186/s40468-021-00122-9

Pourdana, N., Nour, P., & Yousefi, F. (2021). Investigating metalinguistic written corrective feedback focused on EFL learners' discourse markers accuracy in mobile-mediated context. *Asian-Pacific Journal of Second and Foreign Language Education, 6*(7). https://doi.org/10.1186/s40862-021-00111-8.

Pourdana, N., & Tavassoli, K. (2022). Differential impacts of e-portfolio assessment on language learners' engagement modes and genre-based writing improvement. *Language Testing in Asia*, *12*(7), 1-19. https://doi.org/10.1186/s40468-022-00156-7

Purpura, J. E. (2014). Assessing grammar. In A.J. Kunnan (Ed.), *The companion to language assessment* (vol. 1, pp. 100-124). John Wiley & sons.

Rafi, F., Pourdana, N., & Ghaemi, F. (2022). Computer-mediated diagnostic assessment of mixed-ability EFL learners' performance on tiered tasks: Differentiating mediation on Google Meet™. *Journal of Modern Research in English Language Studies, 9*(2), 1-26. http://dx.doi.org/10.30479/jmrels.2021.16118.1950

Rafi, F., & Pourdana, N. (2023). E-diagnostic assessment of collaborative and individual oral tiered task performance in differentiated second language instruction framework. *Language Testing in Asia*, *13*(1), 6. https://doi.org/10.1186/s40468-023-00223-7

Rodrick Beiler, I. & Dewilde, J. (2020). Translation as translingual writing practice in English as an additional language. *The Modern Language Journal, 104*(3), 533-549. https://doi.org/10.1111/modl.12660

Saadatmandi, M., Modarres, K. S., & Pourdana, N. (2018). Teaching English pragmatic features in EFL context: A focus on request speech acts. *Theory and Practice in Language Studies, 8*(7), 829–835. https://doi.org/10.17507/tpls.0807.14.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Shekary, M., & Tahririan, M. H. (2006). Negotiation of meaning and noticing in text-based online chat. *The Modern Language Journal, 90*(4), 557–573. https://doi.org/10.1111/j.1540-4781.2006.00504.x

Storch, N. (2013). *Collaborative writing in L2 classrooms.* Multilingual Matters.

Suzuki, W. (2009). Improving Japanese university students' second language writing accuracy: Effects of languaging. *Annual Review of English Language Education in Japan, 20*, 81-90. https://doi.org/10.1016/j.sbspro.2016.10.057

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning, 62*(4), 1110-1133. https://doi.org/10.1111/j.1467-9922.2012.00720.x

Suzuki, W. (2017). The effect of quality of written languaging on second language learning. *Writing & Pedagogy, 8*(3), 461-482. https://doi.org/10.1558/wap.27291

Suzuki, W., & Storch, N. (2020). *Languaging in language learning and teaching: A collection of empirical studies.* John Benjamins.

Suzuki, W., & Itagaki, N. (2009). Languaging in grammar exercises by Japanese EFL learners of differing proficiency. *System, 37*, 217-225. https://doi.org/10.1016/j.system.2008.10.001

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seildlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125–144). Oxford University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *The handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In B. Heidi (ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95-108). Continuum.

Swain, M., Lapkin, S., Knouzi, I., Suzuki, W., & Brooks, L. (2009). Languaging: University students learn the grammatical concept of voice in French. *The Modern Language Journal, 93*, 5-29. http://dx.doi.org/10.1111/j.1540-4781.2009.00825.x

Van Patten, B. (2015a). Input processing in adult SLA. In B. Van Patten & J. Williams (Eds.), *Theories in second language acquisition* (2nd ed.) (pp. 113-135). Routledge.

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber, & A. S. Carton (eds.), *The collected works of L. S. Vygotsky: Vol. 1: Problems of general psychology* (pp. 39-285). Plenum.

Wang, Y. (2019). Calibration in multiple text use. *Metacognition Learning, 14***,** 131–166. https://doi.org/10.1007/s11409-019-09201-y

Yamashita, T. (2021). Corrective feedback in computer-mediated collaborative writing and revision contributions. *Language Learning & Technology*, *25*(2), 75–93.

Yang, L. (2016). Languaging in story rewriting tasks by Chinese EFL students. *Language Awareness, 25*(3), 241-255. http://dx.doi.org/10.1080/09658416.2016.1197230

Yilmaz, M. (2016). Improving Turkish EFL learners' writing accuracy: Effects of written languaging and languaging types. *Procedia - Social and Behavior Sciences, 232,* 413-420. http://doi.org/10.1016/j.sbspro.2016.10.057

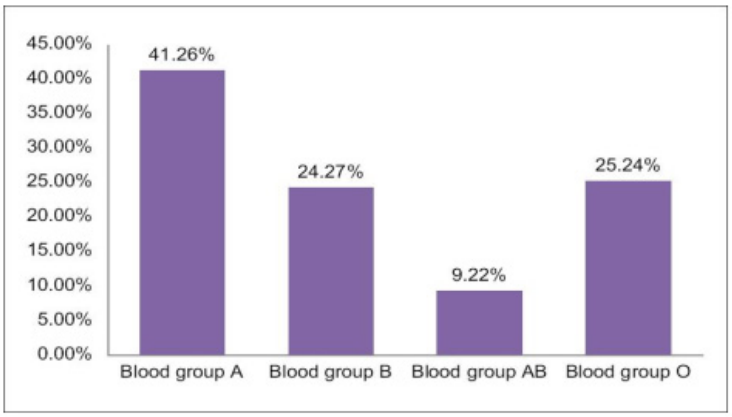
Zeng, G., & Takatsuka, S. (2009). Text-based peer-peer collaborative dialogue in a computer-mediated learning environment in the EFL context. *System, 37*(3), 434–446. https://doi:10.1016/j.system.2009.01.003

Zhang, M. (2021). Understanding L1 and L2 interaction in collaborative writing: A lexico- grammatical analysis. *Language Teaching Research, 25*(3), https://doi.org/10.1177/1362168819859911

APPENDIX A

Error-Identification Task + WL

COVID-19 and Blood Types



(Adopted from Al-Kheikani, 2020)

Read the passage and choose the accurate responses by highlighting them. You can write your languaging episodes on the margins.

این نمودار گزارش می دهد که گروههای خونی ABO در حساسیت به COVID-19 نقش دارد. افراد گروه خونی A در مقایسه با گروه خونی O بیشتر در معرض خطر عفونت قرار دارند. درصد گروه های خونی O ، AB، Bو A در کل بیماران تحت بررسی به ترتیب ۲۵/۲٤ ، ۹/۲۲ ، ۲٤/۲۲ و ٤۱/۲٦ است. گروه خونی O در مقایسه با گروه های خونی غیر O با خطر کمتری همراه است. برعکس ، گروه خونی A در مقایسه با گروه های خونی غیر A با خطر بیشتری همراه است.

3. I don't know this word but I guess it means cruelty.

8. percent or parents?

12. I am not sure about this one!

This picture[1] reports that the ABO blood types plays[2] roles in susceptibility[3] to COVID-19. The[4] people with Blood type A has[5] a higher danger[6] of infection compared to[7] Blood type O. The percent[8] of O, AB, B, and A Blood types in the total investigating[9] patients was 25.24, 9.22, 24.27, and 41.26, respectively[10]. Blood type O is associated in[11] a lower risk compared to the non-O Blood types. To[12] the contrary, Blood type A is associated with a higher risk compared to the non-A Blood types.

[1] Correct/Incorrect
[2] Correct/Incorrect
[3] Correct/Incorrect
[4] Correct/Incorrect
[5] Correct/Incorrect
[6] Correct/Incorrect
[7] Correct/Incorrect
[8] Correct/Incorrect
[9] Correct/Incorrect
[10] Correct/Incorrect
[11] Correct/Incorrect
[12] Correct/Incorrect

1. picture or image?
2. –

4. -
5. -
6. danger is correct?

7. compared against.
9. investigated is correct.
10. relatively is correct. respectively? why?

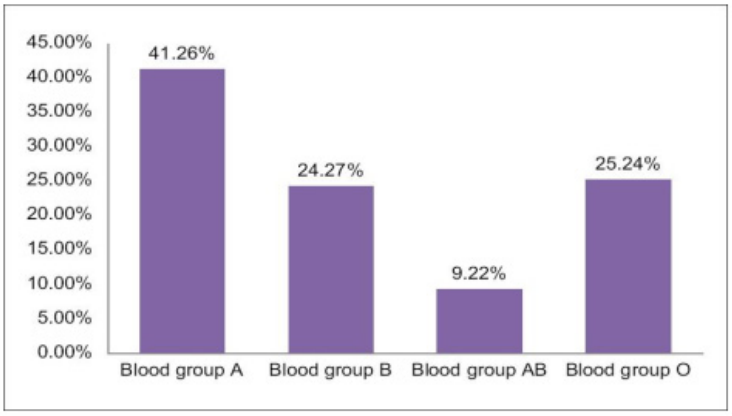
11. -

The Score: 6/12

APPENDIX B

Gap-filling Task + WL

COVID-19 and Blood Types



(Adopted from Al-Kheikani, 2020)

Read the passage and write the best English equivalence in the blanks. You can write your languaging episodes on the margins.

این نمودار گزارش می دهد که گروههای خونی ABO در حساسیت به COVID-19 نقش دارد. افراد گروه خونی A در مقایسه با گروه خونی O بیشتر در معرض خطر عفونت قرار دارند. درصد گروه های خونی O، AB، B و A در کل بیماران مورد بررسی به ترتیب ۲۵/۲٤ ، ۹/۲۲ ، ۲٤/۲۲ و ٤۱/۲٦ است. گروه خونی O در مقایسه با گروه های غیر O با خطر کمتری همراه است. برعکس ، گروه خونی A در مقایسه با گروه های غیر A با خطر بیشتری همراه است.

1. I don't remember the exact word! I think it can't be a picture. Pictures don't show information!

2. Take roles or play roles?

3. I am sure it is not the best word but it is close in meaning.

4. –

5. People are in. I think I should write 2 words here.

This *graph* [1] reports that the ABO blood types *play* [2] roles in *sensitiveness* [3] to COVID-19. - [4] people with Blood type A *are in* [5] a higher *danger* [6] of infection compared *to* [7] Blood type O. The *percentage* [8] of O, AB, B, and A Blood types in the total *observed* [9] patients was 25.24, 9.22, 24.27, and 41.26, *in order* [10]. Blood type O is associated *with* [11] a lower risk compared to the non-O Blood types. *On* [12] the contrary, Blood type A is associated with a higher risk compared to the non-A Blood types.

6. I am not sure danger or condition!

7. –

8. Percentages or percentage?

9. I don't know the exact word for it, but patients can't be studied. Or can they? It must be observed.

10. Why it has come at the end of the sentence? It is very confusing!

11. –

12. To or On? I am not sure.

The Score: 11/12

https://doi.org/10.17323/jle.2023.14717

Academic Procrastination among Indonesian University Learners: Interaction with Cheating, Absenteeism, and L2 Achievement

Adaninggar Septi Subekti

Universitas Kristen Duta Wacana, Yogyakarta, Indonesia

ABSTRACT

Background. Many studies suggested that academic procrastination is particularly prevalent among learners at university level. However, empirical data on the interactions between academic procrastination and, respectively, learners' attitudes towards cheating (AtC), absenteeism, and learning achievement, are either generally inconclusive or non-existent, especially in English as Foreign Language (EFL) literature. Thus, it is worthwhile to conduct a study to examine these issues in the Indonesian EFL context, home to one of the largest communities of EFL learners in the world.

Purpose. The aim of this study was to investigate academic procrastination of Indonesian EFL learners at university level and the interactions of these learners' procrastination with AtC, absenteeism, and second/foreign language (L2) achievement.

Method. The study used an online survey method and 164 learners from non-English departments participated in this study.

Results. On the basis of descriptive statistics, it was found that the participants reported a moderate level of procrastination in English class. Furthermore, this study found that learners' procrastination significantly and positively correlated with their AtC and absenteeism. This indicated that the more learners procrastinated, the higher their approval of cheating behaviour, and the more likely they were to be absent in English classes. The predictive power of learner procrastination was 16.4% on AtC, and at 8.3% on absenteeism. Moreover, the study also found a significant, negative, and moderate relationship between learner procrastination and their L2 achievement with learners' procrastination being able to predict 16.5% of the total variance in L2 achievement.

Conclusion. Teachers are suggested to promote project-based tasks in groups where the step-by-step progress of learners is continually monitored, feedback given, and rewarded. This could discourage procrastination, absenteeism, as well as cheating behaviours, and potentially promote more optimal L2 achievement.

KEYWORDS

English as Foreign Language (EFL), academic procrastination, attitudes toward cheating, absenteeism, second/foreign (L2) achievement

Citation: Subekti A. S. (2023). Academic procrastination among indonesian university learners: Interaction with cheating, absenteeism, and L2 achievement. *Journal of Language and Education, 9*(1), 128-137. https://doi.org/10.17323/jle.2023.14717

Correspondence:
Adaninggar Septi Subekti,
adaninggar@staff.ukdw.ac.id

Received: August 08, 2022
Accepted: March 15, 2023
Published: March 31, 2023

INTRODUCTION

Etymologically, the term "procrastination" comes from the Latin word "procrastinate-" meaning "defer until the morning". Procrastination happens in daily life, including in academic environments where learners postpone working on or completing academic tasks (Alexander & Onwuegbuzie, 2007), hence the term academic procrastination. Academic procrastination is a form of situational procrastination in which learners intentionally delay or defer works that must be completed (Schraw et al., 2007). Learners inclined towards academic procrastination may be aware or not aware that they are engaging in such behaviour (Janssen, 2015). Whilst academic procrastination may happen among learners from all levels of education (Swaraswati et al., 2017), this phenomenon is prev-

alent among learners at the university level (Gonda et al., 2021; Yurtseven & Doğan, 2019).

Interactions between Academic Procrastination and Learning Achievement

There seems to be a debate over whether procrastination merely has detrimental effects or it may have functional effects. Several researchers have stated that procrastination does not always influence learning in a negative way (Cao, 2012; Chu & Choi, 2005). Chu and Choi (2005) argued that procrastinators can be categorised into active and passive procrastinators. Whilst the passive tend to postpone working on tasks unintentionally, the active defer working on tasks because they know they work better under pressure (Chu & Choi, 2005). In other words, despite seemingly delaying, these active procrastinators have control of their tasks and have the intention of performing satisfactorily and meeting the given deadlines (Chu & Choi, 2005). In a similar vein, a study involving learners from undergraduate and graduate levels by Cao (2012) reported that some participants identifying themselves as procrastinators, did so because they felt confident with their abilities to obtain satisfactory results despite procrastinating. This suggested that they did not lose control of their work. A study by Babadoğan (2010) in Turkey reporting a non-significant association between procrastination and achievement may indirectly give support to the claim of these aforementioned researchers.

Despite the aforementioned claim on the possible functional effect of procrastination, many studies have suggested its detrimental effects. Several studies investigated the possible relationship between academic procrastination and L2 achievement in several different contexts such as in Turkey, Iran, and China (Akpur, 2017; Aydoğan & Akbarov, 2018; Kafipour & Jafari, 2021; Korkmaz et al., 2018; Yurtseven & Akpur, 2018; Zhang & Zhang, 2022). They suggest negative associations between procrastination and academic achievement. A study involving 211 Turkish university learners by Akpur (2017) found a significant, negative association between learners' academic procrastination and L2 achievement, $r_{(209)} = -.58$ $p < .05$. In a similar vein, another study in a Turkish context by Yurtseven and Akpur (2018) also reported a significant negative association between the two variables, albeit at a weak level ($r = -.28$, $p < .01$). Slightly similar in findings, a recent study by Zhang and Zhang (2022) in China reported that learners' academic procrastination had a significant negative effect on the readability of their L2 writing. This indicated that the more learners procrastinated, the greater the decrease in the readability of their writing, suggesting a lower quality. Moreover, a recent study involving L2 learners of English from a medical department by Kafipour and Jafari (2021) even reported that learners' procrastination contributed to 90.8% of the total variance in the learners' L2 writing performance.

This highlights the contrasting views and findings on association as well as the causal relationship between academic procrastination and achievement. Therefore, more studies investigating the effects of academic procrastination on achievement are necessary. Moreover, in the Indonesian L2 context, home to the second largest community of English as Foreign Language (EFL) learners after China, such studies are still very limited, if not non-existent.

Possible Factors Affecting Academic Procrastination

There are several contributing factors or antecedents to academic procrastination. One of them is task aversiveness, referring to actions learners find unpleasant (Steel, 2007). Learners try to avoid aversive stimuli and as such, they try to avoid them, for instance by procrastinating. Aversion to a certain task could be attributed to various personal characteristics such as motivation. However, when learners find a task unappealing or even unpleasant, the more likely they are to defer working on it (Steel, 2007). Furthermore, teachers could also be antecedents of academic procrastination (Schraw et al., 2007). Teachers who expect less, who negotiate deadlines and who are not strict with grading could unintentionally promote procrastination. In comparison, learners are less likely to procrastinate knowing their teachers expect good-quality work (Schraw et al., 2007). Furthermore, several studies also reported the role of gender in affecting procrastination (Balkis & Duru, 2017; Roy & Banerjee, 2022; Zhou, 2018). A study in Turkey by Balkis and Duru (2017), for example, reported that male learners had a higher level of academic procrastination than females.

Interestingly, recent studies suggested that the use of the internet contributes to academic procrastination (Herdian & Zamal, 2021; Mohammadi et al., 2015; Wulandari et al., 2021). A study involving 30 Iranian learners of English by Mohammadi et al. (2015) found a medium positive relationship between internet use and academic procrastination, suggesting that the more the participants used the internet, the higher the tendency to procrastinate. Furthermore, two studies in general education in Indonesia also reported that learners procrastinated more during the Covid-19 pandemic-driven online learning (Herdian & Zamal, 2021; Wulandari et al., 2021). Herdian and Zamal (2021) further reported that learners' procrastination was at a moderate level. These findings suggest that the internet often seen as an incredible source of learning can also be a distraction for learners (Satsevich et al., 2021) for example if they access social media and entertainment sites when they are supposed to be participating in an online class or working on certain tasks. For this reason, a study on procrastination in an online learning context will e very useful for further studies in the field of procrastination in an online environment.

Cheating and Absenteeism: Potential Interactions with Academic Procrastination

Several studies, albeit quite limited in number, seem to be interested in the possible association between academic procrastination and academic dishonesty such as cheating. From the perspective of learning, cheating serves as a cognitive shortcut (Anderman & Murdock, 2007). As effective teaching requires learners to use self-regulatory, as well as complex cognitive strategies, cheating is seen by learners as a way to preclude these needs (Anderman & Murdock, 2007). Intuitively, cheating may closely be related to task aversiveness contributing to procrastination. Outside the L2 context, Oktaria et al. (2021) in their study involving learners from a medical department in Indonesia argued that academic procrastination is one of contributing factors to academic dishonesty, including cheating behaviours. However, their study eventually reported that there was no association between learners' academic procrastination with academic dishonesty. In comparison, a study in Iran by Saracaloğlu et al. (2021) reported that academic procrastination among graduate and undergraduate learners correlated positively and moderately with their attitude toward cheating. Thus, the relationship between procrastination and cheating has not been firmly established. Hence, it could be strategic to investigate the association between the two.

Another frequent phenomenon in various L2 learning contexts is absenteeism (Al-Mekhlafi, 2016; Subekti, 2020). A study involving Indonesian English teachers by Subekti (2020) reported the frustration of teachers with the high level of absenteeism among learners. Learners were reported to skip classes despite the efforts of teachers to remind them and despite knowing the consequences on their grades (Subekti, 2020). A study in Yemen by Al-Mekhlafi (2016) also reported discontent among teachers with attendance and motivation of learners in joining the English class. Absences could lead to learners falling behind their peers and having fewer opportunities to obtain complete class content (Al-Mekhlafi, 2016). This may also stimulate task aversion leading to procrastination. Hence, it could be worthwhile to conduct a study investigating the association between procrastination and absenteeism considering such studies are still rare in literature.

Considering these rationales, this study intends to answer the following research questions. First, what is the level of academic procrastination among Indonesian university learners in online English classes? Second, what is the relationship between the level of procrastination and, respectively, their attitudes toward cheating, absenteeism, and L2 achievement?

A study with the aforementioned research objectives in the relatively under-researched Indonesian L2 context could generally contribute to future studies on procrastination concerning other relevant components in L2 learning in the online learning environment.

METHOD

Research Design

The study used a quantitative design by distributing an online Google Form questionnaire. The use of a quantitative design in this study was attributed to several factors. First, a large number of research studies on procrastination have been conducted quantitatively (Akpur, 2017; Aydoğan & Akbarov, 2018; Babadoğan, 2010; Korkmaz et al., 2018; Luján et al., 2021; Zhang & Zhang, 2022), suggesting the popularity in the field of procrastination across various contexts. Secondly, considering the scarcity of such studies in the Indonesian L2 context, such a quantitative study may produce generalisable data for further relevant studies in the Indonesian L2 context.

The online questionnaire used in the study consisted of several parts: explanations about the purpose of the study, consent forms, demographic information, eight questionnaire items on attitude toward cheating (AtC) in English class, and ten questionnaire items on academic procrastination in English class. The eight questionnaire items on AtC were adapted from a study by Carpenter et al. (2006) with the necessary adjustments to fit the L2 context of the present study. Likewise, the ten questionnaire items on academic procrastination, which would be the focus of the present study, were adapted from the Pure Procrastination Scale developed by Lien et al. (2014) in the field of psychiatry. For example, "I delay making decisions until it is too late" in the original questionnaire was modified into "Related to assignments from English class, I delay making a decision until it is too late" to help the participants contextualise their responses to the English class context. Four possible responses were available, "Strongly agree" (converted into 5 points), "Agree" (4 points), "Disagree" (2 points), and "Strongly disagree" (1 point). In the present study, the questionnaires on AtC produced .86 Cronbach's alpha coefficient and .86 McDonald's omega coefficient, indicating high internal reliability. Similarly, the questionnaires on academic procrastination produced .88 Cronbach's alpha coefficient and .89 McDonald's omega coefficient, indicating high internal reliability.

Research Setting and Participants

The research setting was General English (GE) classes Levels 1, 2, and 3 at a private university in Java, Indonesia. GE classes were non-credited matriculation classes taken by learners from various non-English departments. At the time of registration at the university, learners took a placement test to determine their English level at matriculation. They were required to pass GE Level 3 to be able to take a credited English for Academic Purposes (EAP) class in their re-

spective departments. Each level consisted of 16 meetings taken over a semester. At the time of data collection, the meetings were conducted online and most of these meetings were conducted synchronously through the Zoom platform with a 75-minute duration each. Additionally, some 500 learners enrolled in GE classes, with the majority of them taking Level 3.

From these 500 enrolled learners, 164 learners participated in the present study. Of these learners, 164 learners participated in the present study. Of these learners, 79 (48.2%) were males and 85 (51.8%) were females, with a maximum age of 22 and a minimum of 16 (M = 19.27). 118 (72%) were learners at GE Level 3, 40 (24.4%) were at GE level 2, and 6 (3.7%) were at GE Level 1. As the GE classes were conducted online, at the time of data collection these participants resided in various islands or regions. 124 participants (75.6%) resided in Java, 11 (6.7%) in Nusa Tenggara, 10 (6.1%) in Sumatera, 8 (4.9%) in Kalimantan, 5 (3%) in Papua, 3 (1.8%) in Sulawesi, 1 (0.6%) in Bali, and 2 (1.2%) in other islands or regions. They were from eight different departments and the details can be observed in Table 1.

Ethical Consideration

The study did not obtain any ethical clearance as there was no ethics committee at the university at which I worked prior to the data collection. However, this study faithfully employs several principles of research ethics. First, autonomy or voluntary participation was maintained through the distribution of a consent form (Cascio & Racine, 2018) to be completed by the prospective participants in the first part of the online questionnaire. The consent form detailed the purposes of the study, the expectations of the participants and their rights including that of being able to withdraw their participation at any time. This was to ensure that the participants understood the study before participating in it (Farrow, 2016; Weinbaun et al., 2019). Additionally, of 164 participants, 63 (38.4%) expressed their willingness to be invited for interviews, should follow-up studies be required. This indicated the fairly high level of enthusiasm of the participants in the present study. The other 161 (61.6%) were not willing to be involved in any follow-up studies, demonstrating autonomy. Furthermore, the principle of beneficence or maximising the benefits for the participants (Weinbaun et al., 2019) was adhered to by making the questionnaire as simple and easy to complete as possible. Monetary rewards were also given to some randomly selected participants as a token of gratitude. Finally, the study also employed the principle of confidentiality (Ramrathan et al., 2016). Though the participants were required to write their names when filling out the questionnaire, these names were kept confidential and were not publicised in the report.

Data Collection and Analysis

The data collection was conducted in the second semester of the 2021/2022 academic year. After permission to conduct the study was granted from the Head of the Language Centre, the organiser of the GE classes, the link of the online questionnaire was shared to learners taking GE classes by GE class teachers in their respective class *WhatsApp* groups and Learning Management System (LMS). The online questionnaire distribution was conducted from 9 May 2022 up to 27 May 2022. The questionnaire data was then downloaded in an Excel file and moved to SPSS 25 for further analysis. The data on the final grades and the total absences of the participating learners was obtained from the course secretary at end of the semester. This data was recorded to SPSS 25 per the names of the participants.

After all the necessary data had been recorded in SPSS 25, several procedures of data analysis were employed. First, descriptive statistics were used to answer the first research question on the level of academic procrastination among learners in English classes. The data was presented in the form of mean scores and percentages. Before parametric tests were performed, the data was tested for normality and homoscedasticity. After all the variables were found to be normally distributed ($p > .05$). Homoscedasticity tests

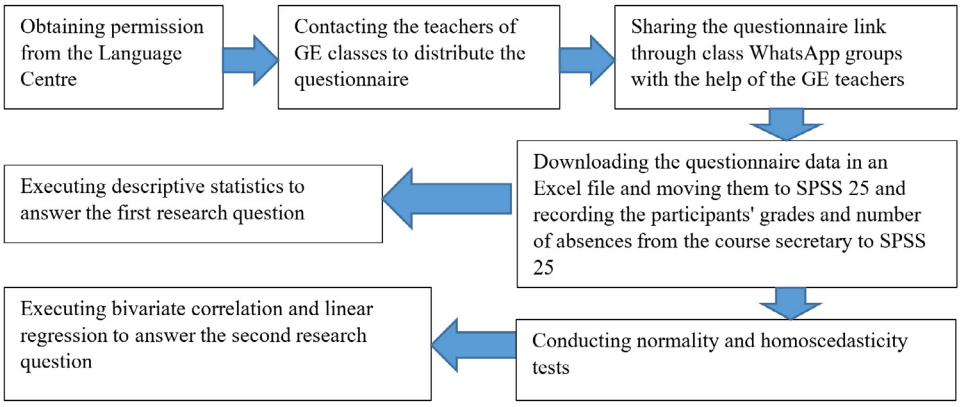
Table 1

The Learner Participants' Departments

No	Departments	Number of participants	Percentages (%)
1.	Accounting	23	14.0
2.	Architecture	24	14.6
3.	Biology	9	5.5
4.	Product design	5	3.0
5.	Informatics	33	20.1
6.	Medical	10	6.1
7.	Management	54	32.9
8.	Information System	6	3.7

Figure 1

The Flow of Data Collection and Analysis



were performed and all the dependent variables, AtC, absenteeism, and L2 achievement had homoscedasticity. Next, bivariate correlation and bivariate (linear) regression formulas were used to answer the second research question on the relationship between procrastination and these components: AtC, absenteeism, and L2 achievement. The sequence of data collection and data analysis can be seen in Figure 1.

RESULTS

Learners' Academic Procrastination in Online English Class

The composite mean score of the ten questionnaire items on procrastination was 25.54, on a scale of 10 up to 50 (*SD = 7.75*). This indicated that in general terms, the learners who participated reported a moderate level of procrastination behaviours in English class. The detailed results in each of the items can be observed in Table 2.

As seen in Table 2, of the ten questionnaire - items on academic procrastination - items 4, 6, and 8 produced the highest mean scores. Item 4, "When working on assignments from English class, I often waste time by doing other things" produced the third highest mean score, at 2.90 on a scale of 1 up to 5, with which 78 participants agreed (47.6%). Then, item 6, "I often find myself performing tasks that I had intended to do days before" produced the highest mean score, at 3.12, with 91 participants (55.5%) agreeing with the statement. Next, item 8, "I generally delay before starting on work I have to do" produced the second highest mean score, at 3.02. 85 participants (51.8%) agreed with the statement.

The Relationship between Learners' Procrastination and Their Attitudes toward Cheating (AtC), Their Absenteeism, and L2 Achievement

The findings on AtC, absences, and L2 achievement among learners are as follows:

First, the mean score of the learner participants' AtC obtained from eight items on AtC was 16.46 on a scale of 1-40, suggesting generally lower AtC or disapproval toward cheating behaviours. Second, the mean score of learners' grades was 78.62 on a scale of 0-100, indicating that they generally had fairly good L2 achievement as measured by their grades. In terms of the total absences in GE classes, the mean score was .79 with the maximum number of absences being six, and the minimum being zero (always present) (*SD = 1.22)*, suggesting that learners generally demonstrated diligence in attending the GE classes.

The results of the bivariate correlation formula on the relationship between learners' procrastination and other components - AtC, absenteeism, and L2 achievement, can be seen in Table 3.

As seen in Table 3, learners' academic procrastination positively correlated with learners' AtC and absenteeism, and negatively correlated with learners' L2 achievement. All three associations were statistically significant. The more learners procrastinated, the more favourable attitude toward cheating they had *(r = .41, p< .01).* The more they procrastinated, the higher the number of absences they had *(r = .29, p < .01).* Lastly, the more they procrastinated, the lower their L2 achievement tended to be *(r = -.41, p< .01).*

Table 2

Learners' Academic Procrastination in English Class

No	Statement	Mean Scores	Standard Deviation *(SD)*	Strongly agree (%)	Agree (%)	Disagree (%)	Strongly disagree (%)
1.	Related to assignments from English class, I delay making a decision until it is too late.	2.20	.97	0.6	18.3	62.8	18.3
2.	Even after I make a decision related to assignments in English class, I delay acting upon it.	2.17	.93	1.2	15.2	66.5	17.1
3.	I waste a lot of time on trivial matters before getting to the final decisions.	2.75	1.22	4.9	36.6	45.7	12.8
4.	When working on assignments from English class, I often waste time by doing other things.	2.90	1.22	5.5	42.1	41.5	10.9
5.	Even assignments that require little else except sitting down and doing them, I find that they seldom get done for days.	2.32	1.10	2.4	22.0	56.1	19.5
6.	I often find myself performing tasks that I had intended to do days before.	3.12	1.19	7.3	48.2	37.8	6.7
7.	Related to assignments from English class, I am continually saying "I will do it tomorrow or another time".	2.77	1.15	4.3	36.0	51.8	7.9
8.	I generally delay before starting on work I have to do.	3.02	1.20	6.7	45.1	39.6	8.6
9.	I do not get things related to English class done on time.	2.17	1.01	1.2	17.7	59.1	22.0
10.	I am not very good at meeting deadlines in English class.	2.13	1.03	1.2	17.7	54.9	26.2

Table 3

Bivariate Correlation between Learners' Procrastination and Other Components – Learners' AtC, Absenteeism, and L2 Achievement

		Learners' AtC	Learners' absenteeism	Learners' L2 achievement
Learners' academic procrastination	Pearson Correlation	.405**	.289**	-.407**
	Sig. (2-tailed)	.000	.000	.000
	N	164	164	164

*Note. **. Correlation is significant at the 0.01 level (2-tailed).*

Furthermore, in order to establish the extent to which learner procrastination impacted their AtC, absenteeism, and L2 achievement, bivariate linear regression formulas were employed with learners' procrastination as the independent variable. Table 4 shows the model summary of the bivariate linear regression results with learners' AtC as the dependent variable.

From Table 4, it can be inferred that academic procrastination impacted 16.4% of the total variance in learner AtC, R^2 = .16, F (1, 162) = 31.88, p < .001. Other variables accounting for the rest 83.6% were outside the formula. Furthermore, academic procrastination significantly predicted learners' AtC, β = .29, t = 5.65, p < .001).

Table 5 shows the model summary of the bivariate linear regression results with learner absenteeism as the dependent variable. As seen in Table 5, the study found that academic procrastination could predict 8.3% of learner absences in English classes, R^2 = .08, F (1, 162) = 14.75, p < .001. Furthermore, academic procrastination significantly predicted learners' absenteeism, β = .05, t = 3.84, p < .001).

Finally, Table 6 shows model summary of the bivariate linear regression results with learners' L2 achievement as

Table 4

Regression results with learners' AtC as the dependent variable

Model	R	R Square	Adjusted R Square	Std. E
1	.405[a]	.164	.159	4.99882

Note. a. Predictors: (Constant), Learners' procrastination

Table 5

Regression results with learners' absenteeism as the dependent variable

Model	R	R Square	Adjusted R Square	Std. E
1	.289[a]	.083	.078	1.174

Note: a. Predictors: (Constant), Learners' procrastination

Table 6

Regression results with learners' L2 achievement as the dependent variable

Model	R	R Square	Adjusted R Square	Std. E
1	.407[a]	.165	.160	16.11405

Note: a. Predictors: (Constant), Learners' procrastination

the dependent variable where it can be seen that learners' academic procrastination could predict 16.5% of their L2 achievement, R^2 = .17, F (1, 162) = 32.09, p < .001. Academic procrastination significantly predicted learner L2 achievement, β = -.92, t = -.58, p < .001).

DISCUSSION

Learners' Academic Procrastination in Online English Class

The present study found a moderate level of academic procrastination among learners from non-English majors. This finding was in line with the findings of recent studies (Herdian & Zamal, 2021; Zhang & Zhang, 2022). A moderate level of procrastination was found in a study by Herdian and Zamal (2021) involving 305 Indonesian pre-service teachers from various departments in a general education context and a study by Zhang and Zhang (2022) involving 55 Chinese learners of English. The similarity could indicate that a certain degree of procrastination was prevalent among university learners regardless of subjects and learning contexts. These findings also confirm the reiteration of several authors stating the prevalence of procrastination among learners at the university level (Gonda et al., 2021; Yurtseven & Doğan, 2019).

Furthermore, learners in the present study also reported that they became distracted by other things whilst working on assignments and delayed work they were supposed to do. With the exception of various intrinsic factors such as level of motivation, this could be attributed to various external factors. In Iran, Mohammadi et al. (2015) found a moderate positive correlation between use of the internet by learners and academic procrastination, suggesting that the internet could be a factor distracting learners from working on their tasks promptly. In the online learning context of the present study where learning was highly dependent on the use of the internet, learners may have been tired of using the internet for educational purposes all the time. As such, the temptation to use it for entertainment, for example accessing social media and entertaining posts, may occur. A study involving school learners by Wulandari et al. (2021) also reported procrastination among learners was quite widespread during Covid-19-driven online learning. Besides this, the limited interaction between teachers and learners during online learning may exacerbate the situation, since low-achieving learners may be falling behind their peers with limited opportunities to ask for help from teachers and peers. In turn, they decided to procrastinate due to task aversiveness (Steel, 2007).

The Relationship between Learners' Procrastination and Their Attitudes toward Cheating (AtC), Their Absenteeism, and L2 Achievement

The study found that academic procrastination among learners could predict 16.4% of their AtC with a statistically significant, moderate and positive association between the two variables. This finding was different from a finding of a study by Oktaria et al. (2021) at a medical faculty where they found no association between the two variables. The finding

was in line with a finding of a study in Iran by Saracaloğlu et al. (2021) involving 357 learners from both graduate and undergraduate levels at various departments. They reported that academic procrastination among learners correlated positively and moderately with their AtC. The similarity between the finding of the present study conducted in an L2 context and that of a study by Saracaloğlu et al. (2021) conducted outside the L2 context may give an early picture of the direction of interaction between academic procrastination and AtC. Intuitively speaking, procrastinating learners may find themselves with limited time left to complete tasks because they delay working on the tasks until much later. In such situations, they may be tempted to take cognitive shortcuts such as cheating, in order not to miss the deadline. Needless to say, however, more empirical studies in L2 learning contexts are needed to establish the interactions between the two variables in the field of L2 learning.

Furthermore, it was found that academic procrastination could predict 8.3% of absences with a statistically significant, weak and positive association between the two variables. When learners procrastinated due to certain reasons, their understanding of materials may not be optimal. This can lead to avoidance behaviours such as skipping classes due to a fear of unpleasant experiences. Procrastination may thus lead to negative feelings toward the class eventually leading to absences. The finding of this study could provide an early empirical 'picture' of the interaction between academic procrastination and learner absences in L2 classes. This finding takes into account the rarity of such studies in the literature, let alone L2 literature despite absenteeism having been a frequently reported issue in studies in various English as L2 learning contexts (Al-Mekhlafi, 2016; Nizar & Flah, 2014; Subekti, 2020).

Moreover, the study also found that academic procrastination could predict 16.5% of learner L2 achievement with a statistically significant, moderate, and negative association between the two variables. This finding conformed with the findings of several previous studies suggesting the negative association between the two variables albeit to different degrees (Akpur, 2017; Aydoğan & Akbarov, 2018; Kafipour & Jafari, 2021; Korkmaz et al., 2018; Yurtseven & Akpur, 2018; Zhang & Zhang, 2022). Kafipour and Jafari (2021), for example, found that learner procrastination could predict a staggering 90.8% of the total variance in L2 writing achievement, whilst two studies in Turkey found a weak to moderate association between procrastination and L2 achievement (Akpur, 2017; Yurtseven & Akpur, 2018). These varieties may be attributed to the nature of the assessment comprising L2 achievement. In the present study and the two studies in Turkey (Akpur, 2017; Yurtseven & Akpur, 2018), the L2 achievement variable was comprised of various types of language assessments. In comparison, the achievement variable in the study by Kafipour and Jafari (2021) was specific in L2 writing. In addition, in the present study, the level of interest among participants in the English class and their level of motivation in joining such non-credited, yet mandatory, GE classes in the present study may explain why their procrastination could to a certain degree negatively affect their L2 achievement.

Despite the findings and possible contributions of the present study, limitations should be acknowledged. The first and foremost limitation is the nature of the self-report questionnaire. In this study it means that the quantitative findings were solely based on the honesty (and possibly dishonesty) of participants in responding to the questionnaire items. Second, this quantitative study did not account for gender differences as a possible contributing factor in the equations. Hence, the results may be seen with gender differences as a possible confounding factor. Furthermore, previous studies on the association between procrastination and AtC in the L2 learning context and procrastination and absenteeism seem to be very limited. Hence, the comparison between the finding of the present study and those of the previous ones were limited. As a result, the findings on both procrastination/AtC and procrastination/absenteeism interactions should be interpreted with caution and may warrant further investigations involving different participants in different L2 learning contexts. This may especially be the case since the sample of this study was generally quite limited, both in number and in breadth and diversity.

CONCLUSION

This study found that academic procrastination among Indonesian L2 learners was generally at a moderate level. It also found statistically significant positive relationships between academic procrastination and two components, learner AtC and absenteeism. In comparison, academic procrastination negatively correlated with learner L2 achievement. Their academic procrastination could predict 16.4%, 8.3%, and 16.5% of the total variance in their AtC, absenteeism, and L2 achievement respectively.

Considering the positive relationships between procrastination and the two components – AtC and absenteeism, teachers could condition instructional design in such a way as to discourage procrastination, cheating, and absenteeism at the same time. Teachers can employ project-based tasks in groups completed during several consecutive meetings. During these meetings, learners consult their progress where submission of progress is also graded. In this way, learners can be more easily assisted, if they have difficulty working on tasks and they are also encouraged to make step-by-step progress. Promoting more group work may also reduce the chance of procrastination, since learners need to coordinate with their group members to work on the task otherwise the quality of the work may not be satisfactory.

Furthermore, this study contributes to establishing the relationship between academic procrastination and three components: AtC, absenteeism, and L2 achievement. In terms of the procrastination/AtC and procrastination/absenteeism relationships, specifically, this study may open the way for further studies in the L2 context considering the infrequency of such studies in L2 literature. However, considering this rarity, the findings should be interpreted with caution, and be treated as an exploration to possibly warrant further investigations.

Future researchers may be interested in conducting a survey involving a bigger sample and employing a stepwise regression, in order to investigate the predictive power of factors such as procrastination, AtC, and absenteeism toward L2 achievement. Conducting studies focusing on the efforts of teachers in reducing procrastination among learners in L2 classes should be encouraged. Whilst it is realised that young adult learners at the university level are responsible for their own learning, there is much that teachers can do to mitigate situations where procrastination is widespread and negatively affects learning.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

Akpur, U. (2017). Predictive and explanatory relationship model between procrastination, motivation, anxiety and academic achievement. *Eurasian Journal of Educational Research*, *67*, 221–240. https://doi.org/http://dx.doi.org/10.14689/ejer.2017.69.12

Al-Mekhlafi, M. A. A. (2016). The impact of class attendance on final examination scores of Yemeni EFL student teachers. *Chetana Education International*, *1*(1), 110–125.

Alexander, E. S., & Onwuegbuzie, A. J. (2007). Academic procrastination and the role of hope as a coping strategy. *Personality and Individual Differences*, *42*(7), 1301–1310. https://doi.org/10.1016/j.paid.2006.10.008

Anderman, E. M., & Murdock, T. B. (2007). *The psychology of academic cheating* (1st ed.). Elsevier Academic Press, Inc.

Aydoğan, H., & Akbarov, A. (2018). A case study on academic procrastination in EFL setting in Turkey. *Acta Didactica Napocensia*, *11*(3), 79–88. https://doi.org/https://doi.org/10.24193/adn.11.3-4.6

Babadoğan, C. (2010). The impact of academic procrastination behaviors of the students in the certificate program in English language teaching on their learning modalities and academic achievements. *Procedia - Social and Behavioral Sciences*, *2*(2), 3263–3269. https://doi.org/10.1016/j.sbspro.2010.03.499

Balkis, M., & Duru, E. (2017). Gender differences in the relationship between academic procrastination, satifaction with academic life and academic performance. *Electronic Journal of Research in Educational Psychology*, *15*(1), 105–125. https://doi.org/10.14204/ejrep.41.16042

Cao, L. (2012). Differences in procrastination and motivation between undergraduate and graduate students. *Journal of the Scholarship of Teaching and Learning, 12*(2), 39–64. https://doi.org/10.14434/josotl.v12i2.2018

Carpenter, D. D., Harding, T. S., Finelli, C. J., Montgomery, S. M., & Passow, H. J. (2006). Engineering students' perceptions of and attitudes towards cheating. *Journal of Engineering Education*, *95*(3), 181–194. https://doi.org/10.1002/j.2168-9830.2006.tb00891.x

Cascio, M. A., & Racine, E. (2018). Person-oriented research ethics: Integrating relational and everyday ethics in research. *Accountability in Research*, *25*(3), 170–197. https://doi.org/10.1080/08989621.2018.1442218

Chu, A. H. C., & Choi, J. N. (2005). Rethinking procrastination: Positive effects of "active" procrastination behavior on attitudes and performance. *The Journal of Social Psychology*, *145*(3), 245–264. https://doi.org/http://dx.doi.org/10.3200/SOCP.145.3.245-264

Farrow, R. (2016). A framework for the ethics of open education. *Open Praxis*, *8*(2), 93–109. http://dx.doi.org/10.5944/openpraxis.8.2.291

Gonda, D., Pavlovičová, G., Tirpáková, A., & Ďuriš, V. (2021). Setting up a flipped classroom design to reduce student academic procrastination. *Sustainability*, *13*. https://doi.org/10.3390/su13158668

Herdian, H. H., & Zamal, N. (2021). The phenomenon of academic procrastination in students during pandemic. *International Journal of Social Science and Human Research*, *04*(08), 2066–2072. https://doi.org/10.47191/ijsshr/v4-i8-15

Janssen, J. (2015). *Academic procrastination: Prevalence among high school and undergraduate students and relationship to academic achievement* [Unpublished doctoral dissertation]. Georgia State University. https://doi.org/https://doi.org/10.57709/7132571

Kafipour, R., & Jafari, S. (2021). Procrastination and its effect on English writing errors and writing performance of medical students. *Indonesian Journal of Applied Linguistics*, *10*(3), 743–750. https://doi.org/10.17509/ijal.v10i3.31763

Korkmaz, O., Ilhan, T., & Bardakçı, S. (2018). An investigation of self-efficacy, locus of control, and academic procrastination as predictors of academic achievement in students diagnosed as gifted and non-gifted. *European Journal of Education Studies*, *4*(7), 173–192. https://doi.org/https://doi.org/10.5281/zenodo.1253354

Lien, R. M. M., Lucien, R., Philipe, G., & Martial, V. D. L. (2014). Validation of a French version of the Pure Procrastination Scale (PPS). *Comprehensive Psychiatry*. https://doi.org/10.1016/j.comppsych.2014.04.024

Luján, R. J. M., Inacio, E. J. H., Calderón, V. E. H., Villamar, R. M., & García, A. K. (2021). Attitudes toward learning english and procrastination in students from a private institution specialized in foreign languages in the city of Lima-Peru. *Revista Colombiana de Psicologia*, *30*(2), 27–39. https://doi.org/10.15446/RCP.V30N2.83678

Mohammadi, M., Tahriri, A., & Hassaskhah, J. (2015). The relationship between internet use and academic procrastination of EFL learners across years of study. *International Journal of Applied Linguistics and English Literature*, *4*(1), 231–241. https://doi.org/10.7575/aiac.ijalel.v.4n.1p.231

Nizar, B. A., & Flah, M. (2014). The impact of absenteeism on undergraduates' academic performance: Evidence from Tunisian English students. *Tayr Quarterly Journal*, *1*(2), 1–17.

Oktaria, D., Azzahra, S. S., & Angraini, D. I. (2021). The relationship of academic procrastination and academic dishonesty in undergraduate medical students. *Jurnal Pendidikan Kedokteran Indonesia - The Indonesian Journal of Medical Education*, *10*(3), 207–214. https://doi.org/10.22146/jpki.63137

Ramrathan, L., Grange, L., & Shawa, L. B. (2016). Ethics in educational research. In *Education studies for initial teacher development* (pp. 432–443). Juta.

Roy, R., & Banerjee, D. (2022). Procrastination amongst undergraduate students in relation to gender and stream. *Journal of Positive School Psychology*, *6*(6), 8808–8817.

Saracaloğlu, A. S., Altin, M., & Eken, M. (2021). Examination of the relationships among prospective teachers' academic procrastination, cheating attitude, academic self-efficacy and academic locus of control. *International Journal of Contemporary Educational Studies, 7*(1), 126–135. https://dergipark.org.tr/en/pub/intjces/issue/64208/942687

Satsevich, K., Bisen, I. E., & Yildirim, Y. (2021). Is online education a setback or step forward? *Eurasian Journal of Higher Education*, *2*(5), 19–35.

Schraw, G., Wadkins, T., & Olafson, L. (2007). Doing the things we do: A grounded theory of academic procrastination. *Journal of Educational Psychology*, *99*(12), 12–25. https://doi.org/10.1037/0022- 0663.99.1.12

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94. https://doi.org/https://psycnet.apa.org/doi/10.1037/0033-2909.133.1.65

Subekti, A. S. (2020). University English teachers' challenges in handling unmotivated learners: A case study. *Langkawi: Journal of The Association for Arabic and English*, *6*(2), 100–113. https://doi.org/10.31332/lkw.v0i0.1926

Swaraswati, Y., Winarno, A. R. D., & Goeritno, G. (2017). Academic procrastination of undergraduate students: The role of academic self-efficacy and the big five personality traits. In G. P. Widanarto, A. Hasjiandito, & H. Dedi (Eds.), *ICSET 2017: Proceedings of The 9th International Conference for Science Educators and Teachers, 13-15 September 2017, Indonesia* (pp. 735–740). Atlantis Press.

Weinbaun, C., Landree, E., Blumenthal, S., Piquado, T., & Gutierrez, C. I. (2019). *Ethics in scientific research: An examination of ethical principles and emerging topics*. RAND Corporation.

Wulandari, I., Fatimah, S., & Suherman, M. M. (2021). Gambaran faktor penyebab prokrastinasi akademik siswa SMA kelas XI SMAN 1 Batujajar di masa pandemi Covid-19. *FOKUS (Kajian Bimbingan & Konseling Dalam Pendidikan)*, *4*(3), 200–212. https://doi.org/http://dx.doi.org/10.22460/fokus.v4i3.7237

Yurtseven, N., & Akpur, U. (2018). Perfectionism, Anxiety and Procrastination as Predictors of EFL Academic Achievement: A Mixed Methods Study. *Novitas-ROYAL (Research on Youth and Language)*, *12*(2), 96–115.

Yurtseven, N., & Doğan, S. (2019). Structural relationships among academic procrastination, academic motivation, and problem solving skill in prep class college students. *Pegem Egitim ve Ogretim Dergisi*, *9*(3), 849–876. https://doi.org/10.14527/pegegog.2019.027

Zhang, C., & Zhang, W. (2022). The impact of academic procrastination on second language writing: The mediating role of L2 writing anxiety. *Frontiers in Psychology*, *13*, 1–11. https://doi.org/10.3389/fpsyg.2022.851120

Zhou, M. (2018). Gender differences in procrastination: The role of personality traits. *Current Psychology*, *39*(4), 1445–1453. https://doi.org/10.1007/s12144-018-9851-5

https://doi.org/10.17323/jle.2023.12262

On the Identifiability of Cognitive Diagnostic Models: Diagnosing Students' Translation Ability

Mona Tabatabaee-Yazdi, Aynaz Samir

Tabaran Institute of Higher Education, Mashhad, Iran

ABSTRACT

Background. In recent years Cognitive Diagnostic Models (CDMs) have attracted a great deal of attention from researchers in a variety of educational fields. However, they have not been taken into consideration in Translation Quality Assessment (TQA), in the aims of presenting fine-grained information about the strengths and weaknesses of translation students.

Purpose. The present study compares the ACDM, DINO, DINA, HO-DINA, and G-DINA models, in order to define the strengths and weaknesses of Iranian translation students and to examine whether the required translation attributes are compensatory, non-compensatory, additive, or hierarchical.

Method. 200 BA translation students translated a two-English-text translation, which was scored by three experienced translation raters using the Translation Quality Assessment Rubric (TQAR). The professional translators, established the relationships between the TQAR items and the nine proposed target translation attributes by constructing a Q-matrix.

Results. Based on the results, HO-DINA can be considered the best-fitting model. Bibliography and technical skills, together with work methodology skills, are shown to be the most difficult attributes for translation students.

Conclusion. HO-DINA is a non-compensatory model, thus the study findings assert that for a correct response to a test item, all measurable attributes need to be mastered.

KEYWORDS

attribute, diagnostic classification models, item response theory, Q-matrix, translation ability, test fairness

INTRODUCTION

According to William (2003), translation is a strategic conscious activity aimed at establishing communication between diverse cultural settings in a controlled way. With the rapid development of globalization and its impact on localization, an increasing volume of products are being imported and exported into other countries. Thus translation plays a significant role in cultural communication (Yan, 2013, p. 36). Indeed, «the need to deliver information quickly and efficiently has put translation at the heart of diverse international cultural, economic, and military enterprises» (Jimenez-Crespo, 2020, p. 375). Due to the importance of globalization and human communication needs, the role of translation has become more valuable. Translation, as an effective way of communication, permits people from different languages and cultures to learn about diverse aspects of the international community, as well as their cultures and ideologies. With the rapid development of translation technology and communication around the world, the ability to translate from or to other languages has become particularly important. According to Nord (1999), translator training programs at universities offer skills and knowledge, which enable students to enhance the required skills, abilities, and translation competences through training, guidance, practice, and getting experience. In this regard, Paradis et al. (1982) believe that the ability to translate foreign languages is a complex integrated cognitive task associated with underlying cognitive components, beyond the

Citation: Tabatabaee-Yazdi M., & Samir A. (2023). On the identifiability of cognitive diagnostic models: Diagnosing students' translation ability. *Journal of Language and Education, 9*(1), 138-157. https://doi.org/10.17323/jle.2023.12262

Correspondence:
Aynaz Samir
aynazsamir@yahoo.com

Received: April 03, 2021
Accepted: Mach 19, 2023
Published: March 31, 2023

ability to speak or understand words and linguistic structure in two languages. In the views of Kelly (2005) and Mackenzie (2004), the translation of different technical texts requires significant interaction among different translation competences and attributes.

In the field of Translation Studies, translation competence [TC] development has been approached from different perspectives. Domínguez Milanés (2015, p. 29) stated that professional competences are «the ensemble of knowledge, skills, personal traits, emotions, motivation, abilities, values, and attitudes that allow an individual to perform successfully not only in a given, situated activity but also in a number of unexpected scenarios in the current, deeply unstable market conditions» (as cited in Martínez-Carrasco, 2017, p. 152). Martínez-Carrasco (2017), studying the nature of translation competence, asserted that «the most common translation sub-competences cover the (inter)linguistic reality of translation, the extra-linguistic input, the role of ICT and other sources of information, [a number of] inter-personal and professional [elements], a psycho-physiological frame, the strategic component, and transferability» (p. 220). For PACTE, translation competence refers to «the underlying system of knowledge required to translate» (2011, p. 321). As «the degree of TC is reflected in both the process and the product of translating» (2014, p. 88), TC encompasses a mix of procedural and declarative knowledge represented in PACTE as a model with five sub-competences (2003): strategic sub-competence, bilingual sub-competence, instrumental sub-competence, extra-linguistic sub-competence, knowledge about translation sub-competence, and psycho-physiological components. Kiraly and Hofmann (2016) add that translation competence is acquired "in a step-by-step, cumulative and essentially linear manner» (p. 72).

Mackenzie (2004) states that translation competence consists of management skills, information technology (IT) skills, marketing ability, linguistic-cultural skills, and interpersonal skills. Kelly (2005, 2008) also suggests her own translation competence model comprising seven sub-competences: (a) interpersonal competence; (b) attitudinal (or psycho-physiological) competence; (c) subject-matter competence; (d) strategic competence; (e) cultural and intercultural competence; (f) professional and instrumental competence; (g) communicative and textual competence in at least two languages and cultures. Perez (2014, as cited in Kabát, 2020, p. 59) also finds that linguistic, intercultural, thematic, technological, info-mining, and translation service provision competences are the most important components of TC. Piecychna (2013) suggests psychological, thematic, textual, and linguistic sub-competences (p. 155). Other translation scholars introduce other important translation skills as part of TC, including grammar skills (Conde, 2013; Dewi, 2015; Hansen, 2010; Lee & Ronowick, 2014), terminological skills (Angelone, 2013; Goff-Kfouri, 2005; Lee & Ronowick, 2014), spelling skills (Beeby, 2000; Dewi, 2015; Doyle, 2003; Waddington, 2001), creativity (Dewi, 2015; Polliastri & Paulina, 2009), time management (Dewi, 2015; Doyle, 2003), problem-solving skills (Dewi, 2015; Nord, 2009), accuracy (Farahzad, 1992; Khanmohammad & Osanloo, 2009; Polliastri & Paulina, 2009), and fluency (Conde, 2011; Dewi, 2015; Farahzad, 1992).

For PACTE (2000) translation competence development can be defined as:

(1) A dynamic, spiral process, which, like all learning processes, evolves from novice knowledge (pre-translation competence) to expert knowledge (translation competence); it requires learning competence learning strategies). During the process, both declarative and procedural types of knowledge are integrated, developed, and restructured.

(2) A process in which the development of procedural knowledge - and consequently of the strategic sub-competence - is essential.

(3) A process in which the TC sub-competences are developed and restructured (as cited in Hurtado Albir, 2015, p. 260).

Thus, assessing student competence in translation programs is essential for the improvement of their professional skills (Beeby et al., 2011; Bonyadi, 2020). Yamaguchi and Okada (2018) mention that achievement tests assess a number of skills and the student's current knowledge. However, translation students with the same overall scores in a translation exam may differ widely in their competences, strengths, and weaknesses. Therefore, a comparison of the sum of the scores of students with the same knowledge differences, cannot really reveal their strengths and weaknesses (Tabatabaee-Yazdi, 2020). Since overall assessment scores cannot satisfy the expectations of the teachers or examiners, the differences in difficulty of the questions should not be ignored (Tabatabaee-Yazdi, 2020). It is crucial to collect detailed information on exam items and more precise data on the specific attributes or skills of students, rather than classifying students based on their scores (DiBello & Stout, 2007; Lee & Sawaki, 2009a).

Only a few empirical studies have been conducted on assessing translation quality with regard to translation process and products (Angelelli & Jacobson, 2009; Stobart & Gipps, 1997). In the words of Conde, "evaluation is still a field in which much remains to be explored" (2012, p. 68). This contrasts with Newmark's (1998) views on assessment: "translation quality assessment is a vital link among translation theory and its practice; it is also a pleasing and informative exercise, specifically, if the assessor evaluates two or more different translations of the same text based on translation standards" (p. 184). Although in some educational settings, statistical assessment techniques have been used to calculate overall skills (Genesee 2002; Genesee & Upshur 1996;

ReaDickins & Germaine, 1993; Tabatabaee-Yazdi, Motallebzadeh, Baghaei, 2021), these techniques have not been used to identify and diagnose the strengths and weaknesses of translation students.

Cognitive Diagnostic Models (CDMs) have emerged to model statistically the examinees' cognitive operations. Their aim is to provide diagnostic feedback, improve teaching and learning processes, and remove the shortcomings of some traditional statistical assessments including classical test theory and factor analysis used to construct tests, and interpret test results by focusing merely on the overall scores. CDMs are given greater attention in some assessment settings since they can increase the opportunity of learning by "pinpointing why students perform as they do" (Leighton & Gierl, 2007, p. 5). The CDM defines the examinees' strengths and weaknesses based on certain specific attributes (Chen & de la Torre, 2013; De la Torre & Lee, 2013; Von Davier, 2005). Rather than placing each examinee on a continuous ability scale based on their scores, CDMs yield categorical diagnostic information about examinees' strengths and weaknesses with different fluency profiles.

CDMs help students answer exam items properly by breaking down the questions/items into several strategies (Birenbaum et al., 1993). Accordingly this permits CDMs to produce "multidimensional diagnostic profiles based on statistically-driven multivariate classifications" of students with regard to performance levels on each of the required skills (Kunina-Habenicht et al., 2012, p. 64).

Ravand & Robitzsch (2018) determine two main purposes for using CDMs: (a) categorizing students into similar skill profiles regarding their answer patterns and (b) identifying compensatory or non-compensatory relations between the attributes of a given skill. There are different arrays of CDMs with various theoretical principles regarding relations between attributes (Ravand & Baghaei, 2020). This is a series of certain mental processes, knowledge, strategies, skills, and competences in which students must answer the items of a test correctly (Leighton & Gierl, 2007).

Types of CDMs

The main factor that distinguishes CDMs from each other is the way in which they model the association between the required attributes when performing a test item or a given task, as well as the probability of a correct response (Table 1, Ravand, 2016, p.3). According to the correlation between attributes, CDMs are classified into two types: (1) conjunctive or non-compensatory models, and (2) disjunctive or compensatory models. If the performance of one or more of the attributes can compensate for the non-performance of other attributes, *compensatory models (Disjunctive)* are used (Ravand & Baghaei, 2020; Tabatabaee-Yazdi, 2020). Conversely, if the performance of one or more of the attributes cannot compensate for the non-performance of other attributes, non-compensatory models (Conjunctive), are used. Thus, in order to achieve a high probability of a correct answer, performance of all the required attributes is needed (Ravand & Baghaei, 2020; Tabatabaee-Yazdi, 2020).

Additive CDMs (de la Torre, 2011) have been proposed as another classification of CDMs. Unlike compensatory models, respondents are credited for the number of attributes performed, signifying that the performance of any one of the attributes can increase the possibility of a correct answer (Ravand & Baghaei, 2020). Finally, the most recent extensions or versions of CDMs, proposed by Templin & Bradshaw (2013), are hierarchical and non-hierarchical CDMs. Hierarchical CDMs (HCDMs) model the structural relations between the required attributes and the impact which the order of teaching materials (where learning a skill is prioritized upon other skills) has on increasing the likelihood of obtaining a correct answer to an item (Tabatabaee-Yazdi, 2020).

More recently, CDMs have been categorized under two major categories: specific and general. In specific CDMs, just one type of association (disjunctive, conjunctive, and additive) can be possible within any assessment. However, general CDMs do not hypothesize any pre-specified relations between underlying sub-skills. Therefore, several kinds of interactions are possible within the same assessment, assuming different relationships between attributes across the items (Ravand & Baghaei, 2020; Tabatabaee-Yazdi, 2020).

Nearly all the models used in the context of the assessment show that they are effective for offering diagnostic feedback in the teaching and learning process (Nichols, 1994). Therefore, a number of CDMs have been used in language assessment studies such as the DINA (Chen & Chen, 2016), the General Diagnostic Model (GDM) (Von Davier, 2005), the G-DINA (Chen & Chen, 2016; Effatpanah et al., 2019; Ravand & Baghaei, 2020; Ravand et al., 2020).), the DINO (Chen & Chen, 2016), the reduced reparameterized unified model (RRUM) (Aryadoust, 2018; Kim, 2015; Lee & Sawaki, 2009a; Li, 2011), and the hierarchical diagnostic classification model (HCDM) (Tabatabaee-Yazdi, 2020).

The objective of this study was to apply and compare the DINO, DINA, ACDM, HO-DINA, and G-DINA models to recognize the strengths and weaknesses of the translation ability of Iranian BA students in a translation test. It also aimed to test whether the required translation attributes are compensatory, non-compensatory, additive, or hierarchical. Accordingly, DINO (as a specific disjunctive model), DINA (as a specific conjunctive model), ACDM (as a specific additive model), HO-DINA (as a specific hierarchical model), and G-DINA (as a general multifunctional model) were applied.

Table 1

CDM Categorization

	DCM Type	Examples	Author(s)
Specific	Disjunctive	• deterministic-input, noisy-or-gate model (DINO)	• Templin & Henson (2006)
		• noisy input, deterministic-or-gate (NIDO) model	• Templin (2006)[1]
	Conjunctive	• deterministic-input, noisy-and-gate model (DINA)	• Junker & Sijtsma (2001)
		• noisy inputs, deterministic "and-gate (NIDA)	• DiBello, Stout, and Roussos (1995); Hartz (2002)
	Additive	• Additive CDM (ACDM)	• de la Torre (2011)
		• compensatory reparameterized unified model (C-RUM)	• DiBello, Stout, and Roussos (1995); Hartz (2002)
		• non-compensatory reparameterized unified model (NC-RUM)	• Hartz (2002)
		• linear logistic model (LLM)	• Maris (1999)
	Hierarchical	• hierarchical DINA (HO-DINA) model	• de la Torre (2008)
General / Saturated	Disjunctive, Conjunctive, & Additive Hierarchical	• general diagnostic model (GDM)	• Von Davier (2005)[2]
		• log-linear CDM (LCDM)	• Henson, Templin & Willse (2009)
		• generalized DINA (G-DINA)	• de la Torre (2011)
		• hierarchical diagnostic classification model (HDCM)	• Templin & Bradshaw (2013)

Note. Adapted from "Application of a Cognitive Diagnostic Model to a High-Stakes Reading Comprehension Test", by H. Ravand, 2016, Journal of Psychoeducational Assessment, 34, p. 3, (https://doi.org/10.1177/0734282915623053). Copyright 2016 by SAGE

DINO

Developed by Templin and Henson (2006), DINO is considered the first in the line of CDMs. In this model, which is very similar in structure and composition to its counterpart DINA, the performance of any one of the attributes increases the chance of a right answer to the test item or the given task. Therefore, in the DINO model, students "are classified into two latent classes: those who have not mastered any of the required subskills and those who have mastered at least one of the subskills" (Aryadoust, 2018, p. 7). "The DINO model is often used in the application of psychiatric assessment, for which the positive response to a diagnostic question (item) could be due to the presence of one disorder (attributes) among several" (Fang, Liu, & Ying, 2019, p. 8).

DINA

DINA, as a non-compensatory model (Junker & Sijtsma, 2001), is known for its parsimony, understandability, and as an easy fit for the data (Fang et al., 2019). Local independence at the attribute level is one of its main features, indicating that the performance of one attribute does not affect that of another. Therefore, "measured attributes in an item are independent of one another" (Galeshi, 2012, p. 18). In other words, the item or task can be answered correctly, if all the required attributes have been mastered thoroughly. However, students who have not mastered any of the attributes may hazard a guess and answer the item accurately (Rupp, Templin & Henson, 2010).

ACDM

Additive CDM (ACDM; de la Torre, 2011), another compensatory model similar to G-DINA (de la Torre, 2011) as a general model, permits both disjunctive and conjunctive relations between attributes within the same test. The ACDM hypothesizes that the probability of producing a correct answer to an item or a task can be increased by mastering each of the required attributes. The absence of one attribute can be compensated for by the existence of other attributes (Effatpanah, 2019). Therefore, ACDM assumes that mastering each of the required attributes advances the likelihood of success, while the absence of one attribute can be compensated by the mastery of other attributes. Moreover, the existence of each sub-skill is independent of the other subskills (Galeshi, 2012).

HO-DINA

HO-DINA is an extension of DINA, which hypothesizes that the required attributes are hierarchically structured (De La

[1] Templin, J. (2006). *CDM User's Guide*. University of Kansas.

[2] Von Davier, M. (2005). *mdltm* [Computer software]. Educational Testing Service.

Torre & Douglas, 2004; De La Torre & Minchen, 2014), and fit the data better than DINA (De La Torre & Douglas, 2004). The first assumption of HO-DINA is that performance of the required attributes is interrelated to "a higher-order and broadly-defined ability parameter similar to the unidimensional θ parameter in the IRT models" (Aryadoust, 2018, p. 6); and secondly, the existence of each attribute is independent of the other attributes (De La Torre & Douglas, 2004). Therefore, HO-DINA can be considered a non-compensatory model which claims that the only essential condition for a correct response to a test item or a given task is performance of all the required attributes measured by that item (Aryadoust, 2018).

G-DINA

G-DINA (De la Torre, 2011) is a general CDM which allows for both compensatory and non-compensatory relationships within the same test, signifying a different model for each item on the same test (Ravand & Robitzsch, 2018). In the G-DINA model, unlike the DINA model, the non-performance of one, some, or all of the required attributes leads to an unequal probability of success for the students. This indicates that irrespective of how many attributes students have mastered, as long as they have mastered at least one of the required subskills (De La Torre & Minchen, 2014), they have the same probability of answering the items correctly. Consequently, knowing one or all the attributes does not necessarily lead to a higher chance of giving a correct response (De La Torre & Minchen, 2014).

METHODS

Participants and Settings

The data set of the present study comprises 200 Iranian junior and senior university students studying *English Translation* at undergraduate level in various national and private universities in Iran. The sample consisted of 51 males (25.5%) and 149 females (74.5%) whose ages ranged between 20 to 44 (M=24.79 years, SD= 3.89). Of the total sample, 157 (78.5%) were studied at national universities and 43 (21.5%) at private universities (Table 2).

Three raters assessed the translations. All are assistant professors and were native speakers of Persian and proficient in English as their foreign language. They had at least 8 years of experiencing in teaching and translating. Three possessed an MA in *Translation Studies* and had been teaching translation courses in different universities of Iran for more than seven years. They were one male and two females between the ages of 27 to 38 years old ($mean_{age}$=36, SD= 1.2). Each rater was provided with a Translation Quality Assessment Rubric, consisting of 23 items to score the translations (Samir & Tabatabaee-Yazdi, 2020).

Instruments

Two instruments were used to collect the data required for the specific objectives. The first instrument was a translation test comprising two texts (a political text of 304 words and a journalistic text of 303 words) extracted from two textbooks on political translation, and news articles taught to BA students at Iranian universities. The political text was about the cabinet government system in different countries. The journalistic text examined terrorism in Asian countries. The two texts were classified as advanced. The text were to be translated into Persian in class within one hour. Students were allowed to use any kind of resources including the internet, online dictionaries, Computer-Aided Translation (CAT) tools, and/or any other software.

The second instrument was Translation Quality Assessment Rubric (TQAR) developed and validated by Samir and Tabatabaee-Yazdi (2020; see Appendix B). The rubric consists of 23 items on a four-point Likert scale, ranging from 4 (*superior*), 3 (*advanced*), 2 (*fair*), to 1 (*poor*). The rubric can be used to identify and score the translation abilities of non-native

Table 2

Demographic Profile of Respondents (n=200)

	Category	Frequency	Percentage
Gender	Male	51	25.5
	Female	149	74.5
Age	20-24	97	48.5
	25-29	57	28.5
	30-34	25	12.5
	35-39	15	7.5
	Above 40	6	3
University	National	157	78.5
	Private	43	21.5

English-speaking translation students in the university context.

For the sake of CDM analysis, and constructing the required Q-matrix, the four-point Likert scale was merged into *mastered* and *non-mastered* . This was achieved by categorizing "poor" and "fair" scales into "non-mastered= 0", and "advance" and "superior" into "mastered= 1". Thus, the TQAR turned into a dichotomous response scale of yes=1 and no=0 to assess the nine proposed translation attributes.

Procedure

Data collection began in October 2019 and lasted for about four months. A consent form was distributed beforehand to the fifty-five English Translation departments currently offering translation programs in Iranian universities, in order to seek their permission and willingness to be included in the study. Only 18 signed consent forms were sent back to the researchers. Thus, only 18 universities, representing National and Azad Universities, participated in this study. In total, 200 translation students were included. Three one-hour training sessions were held for translation raters, in order to increase their awareness about the research objectives, translation attributes, TQAR, Q-matrix construction, and the rating scale.

Analysis

In order to identify accurate correlations between target attributes and test items, a Q-matrix was hypothesized (Table 3, Appendix A). According to Li (2011), Tabatabaee-Yazdy, (2020), and Tatsuoka (1983) a Q-matrix illustrates rows[3] and columns[4] which symbolize test items and required underlying traits to answer each item correctly. In order to identify the required underlying traits or subskills, the present study used Samir and Tabatabaee-Yazdi's Translation Quality Assessment Rubric (2020; see Appendix B) as the Q-matrix items and sub-skills. In a joint session, the researchers and the three raters agreed on the required attributes. Then, the translation raters created a Q-matrix by organizing the 23 items of the TQAR into nine translation attributes, as follows:

(1) linguistic skills (lexical) in the source language (LSSL: L), used to assess students' ability to recognize the meaning of a wide variety of terminology and lexical items in the source text (ST) accurately and appropriately,

(2) linguistic skills (grammatical) in the source language (LSSL: G), aiming at evaluating students' ability in recognizing the ST grammatical rules such as the relative order of subject, verb, modifiers, clauses, and syntactic elements,

(3) linguistic skills (lexical) in the target language (LSTL: L), used to assess students' ability to use specific terms in the translation of the technical text,

(4) linguistic skills (grammatical) in the target language (LSTL: G), used to identify students' ability to follow target language (TL) grammatical rules, such as subject and verb agreement or arrange the words according to the TL rules,

(5) organizational knowledge (ORG), used to assess students' ability to translate and organize ideas cohesively and coherently in order to convey all sections (sentences, titles, headlines…) of the ST to TT,

(6) cultural knowledge (CUL), used to assess students' ability to produce correct and idiomatic use of the target language and to preserve an appropriate register in the translation,

(7) translation skills (TRN,. aiming to study students' ability to produce the target text (TT) at an acceptable level of fluency and avoid words and expressions of ambiguous meanings,

(8) work methodology skills (WM), used to assess students' ability to manage time and respect the deadline,

(9) bibliography and technical skills (BT), used to assess students' ability to use a relevant terminological database or bibliography.

The final Q-matrix presented in Table 3 was proposed after several rounds of revisions and the necessary modifications. Of the total items, 11 of them were associated with attribute 1 (LSSL: L), 11 items, with attribute 2 (LSSL: G); 13 items, with attribute 3 (LSTL: L); 13, with attribute 4 (LSTL: G); 11, with attribute 5 (ORG); 8, with attribute 6 (CUL); 13, with attribute 7 (TRN)*;* 4, with attribute 8 (WM); and 3, with attribute 9 (BT). 1s in Table 3 indicate that the probability of producing a correct answer on each item is conditional on the mastery of the attributes, whereas 0s show that the item does not need the attributes. For example, in order for students to get the correct answer to item 7 (to deal with terminological terms correctly), they should have mastered the first, third, and seventh attributes. Therefore, in this item, students are not required to master attributes 2, 4, 5, 6, 8, and 9.

[3] Rows in this matrix signify the number of items on the test.

[4] Columns in this matrix show the number of test's underlying attributes.

RESULTS

R statistical software and CDM package version 6.1-10 (Robitzsch, Kiefer, George & Uenlue, 2018[5]) were used to analyze and compare the fit of the five selected CDMs, including DINO, DINA, ACDM, HO-DINA, and G-DINA. The CDM package illustrates the various fit indices which can be applied, in order to identify the best model among the competing models (relative fit indices), It can also be used to check the fit of a model to the observed response data (absolute fit indices) (Effatpanah, Baghaei, & Boori, 2019; Rupp, Templin, & Henson, 2010; Tabatabaee-Yazdi, 2020). The models were compared to the relative and absolute fit indices using AIC[6], BIC[7], CAIC[8], Mx2[9], MADcor[10], SRMSR[11], MADQ3 [12], and MADRESCOV[13].

Optimal Model Fit

The fit of the data to CDMs identifies the accuracy of the correlation between attributes and items. Table 4 shows the Absolute and Relative fit indices of the five models. Although there are no agreed-upon cut-off values for the absolute fit indices of CDMs, the smaller the effect size (values), the better a model fits (Robitzsch et al., 2015[14]).

As can be seen in Table 4, the Npars (number of parameters) column shows that the G-DINA model estimated 1094 item parameters: ACDM, 156 parameters; DINA and DINO 92, parameters; and HO-DINA, 64 parameters. This shows that HO-DINA is a parsimonious model and G-DINA is the most complicated model. Moreover, there is a non-significant value (> 0) for MaxX2 for all the models. As to the MADcor, SRMSR, and MADRESIDCOV, the G-DINA had the lowest values. However, with respect to AIC, BIC, and CAIC, the value of HO-DINA was the lowest when compared to the other models. Therefore, since BIC carries a great penalty for more highly parameterized models, it can be anticipated that the G-DINA model has the worst value (Li, Hunter, & Lei, 2016).

As a result, after considering all the indices, the HO-DINA was selected as the best-fitting specific CDMs for further examination, in order to study whether the model can precisely detect the students' translation skills.

HO-DINA Analysis

HO-DINA Parameters

Table 5 describes the study's model (HO-DINA) parameters. Ttwo items (items 7 and 8) are shown in Table 5 as examples. The successful performance of item 7 (Terminology and False Friends) requires students to have mastered attributes 1 (LSSL: L), 3 (LSSL: L), and 7 (TS). Students who have mastered none of the required attributes only have a 30% chance to respond correctly. However, students who have mastered only LSSL: L have a 31% probability of guessing the correct item. Students who master attribute 1 had a 0.30 + 0.31 = 0.61% probability of success on item 7. In the same vein, if students have mastered only attribute 3, there is a 0.30% chance of answering the item correctly. Thus, those who have mastered attribute 3 have a 0.30 + 0.30 = 0.60% probability of success on this item. Finally, those respondents who have mastered the three attributes have a 56% probability of getting the right answer and, consequently, 86% (0.56 + 0.30 = 0.86) probability of success on this item. Another example would be *spelling* (item 8), where students are required to know attribute 3 (LSSL: L). If the students have not mastered the attribute, they only have a 0.74% chance to perform well on this item. However, by mastering the attribute, the probability of performing item eight increases to 99%.

5 Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, C. (2018). CDM: Cognitive diagnosis modeling (Rpackage version 6.1-10). https://cran.r project.org/web/packages/CDM/index.html

6 Akaike Information Criterion (AIC; Akaike, 1974) is one of the relative fit indices used to select between non-nested models. Models with lower AIC are more preferable.

7 Bayesian Information Criteria (BIC; Schwarz, 1978) is another relative fit index used to choose between non-nested models. Models with lower BIC are more preferable.

8 Consistent AIC (Akaike, 1974).

9 Mx2 (Chen & Thissen, 1997) is the test of global model fit. It is the mean difference between the model-predicted and observed response frequencies. If CDM fits the data well, "the x2 test statistic is expected to be 0 within each latent class as the attribute profile of the respondents would perfectly predict the observed response patterns" (Rupp et al., 2010, p. 269).

10 The mean absolute difference for the item-pair correlations (MADcor) statistic (DiBello, Roussos, & Stout, 2006) shows the mean of absolute discrepancy between observed and predicted pairwise item correlations across all item pairs. A MADcor value of 0.049 in Jang (2005) was suggested as a good fit for the DCM to the data.

11 The standardized root mean square residual (SRMSR) is the square root of the difference between the observed correlation and the model covariance matrix. Models with SRMSR values below 0.05 are considered models with "substantively negligible amount of misfit" (Maydeu-Olivares, 2013, p. 84) and models with values below 0.08 as good fit (Hu & Bentler, 1999).

12 The MADQ[3] is "the Pearson product-moment correlation of a set of residuals from the IRT model" (Chen & Thissen, 1997, p. 280; Yen, 1984). This index is to some extend less sensitive than the Mx[2] (Aryadoust, 2018).

13 The average of absolute values of pairwise item covariance residuals (MADRESCOV; McDonald & Mok, 1995) illustrates the average deviations between matrices of observed and reproduced item correlations.

14 Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2015). CDM: Cognitive Diagnosis Modeling. R package version 4.5-0. http:// CRAN.R-project.org/package=CDM

Table 3

The Final Q-Matrix

Attribute	Att. 1	Att. 2	Att. 3	Att. 4	Att. 5	Att. 6	Att. 7	Att. 8	Att. 9
Item	LSSL: L	LSSL: G	LSTL: L	LSTL: G	ORG	CUL	TRN	WM	BT
1	0	1	0	1	0	0	1	0	0
2	1	1	1	1	1	1	0	0	0
3	1	1	1	1	1	1	1	0	0
4	1	1	1	1	1	1	1	0	0
5	0	0	0	0	1	0	1	0	0
6	0	1	0	1	0	0	0	0	0
7	1	0	1	0	0	0	1	0	0
8	0	0	1	0	0	0	0	0	0
9	1	1	1	1	0	0	1	0	0
10	1	1	1	1	1	1	1	0	0
11	1	1	1	1	1	1	1	0	0
12	1	1	1	1	1	0	1	0	0
13	0	0	1	1	1	0	1	0	0
14	1	1	1	1	1	1	1	0	0
15	1	1	1	1	1	1	1	0	0
16	1	0	1	1	1	1	1	0	0
17	0	0	0	0	0	0	0	1	0
18	0	0	0	0	0	0	0	1	0
19	0	0	0	0	0	0	0	1	0
20	0	0	0	0	0	0	0	1	0
21	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	1
23	0	0	0	0	0	0	0	0	1

Table 4

Fit Indices

		Relative Fit Indices	.			Absolute Fit Indices			.
Model	Npars	AIC	BIC	CAIC	MaxX2 (p)	MADcor	SRMSR	MADQ[3]	MADRESIDCOV (MADRCOV)
DINO	92	3936.19	4239.64	4331.64	178.314(0)	0.06378	0.10439	0.07110	0.98298
DINA	92	3935.68	4239.12	4331.12	178.313(0)	0.06347	0.10411	0.06874	0.98105
ACDM	156	3924.42	4438.96	4594.96	**174.389(0)**	0.05037	0.09266	**0.05990**	0.77066
HO-DINA	**64**	**3923.23**	**4134.33**	**4198.33**	178.357(0)	0.06577	0.10502	0.07812	1.00438
G-DINA	1094	5567.13	9175.49	10269.49	175.756(0)	**0.04445**	**0.08931**	0.06571	**0.70469**

Table 5

HO-DINA Parameters

Item No.	Required Attributes	Mastery Patterns	Probability
7	Att.1–3–7	A000	0.30
7	Att.1–3–7	A100	0.31
7	Att.1–3–7	A010	0.30
7	Att.1–3–7	A001	0.31
7	Att.1–3–7	A110	0.30
7	Att.1–3–7	A101	0.31
7	Att.1–3–7	A011	0.31
7	Att.1–3–7	A111	0.56
8	Att.3	A0	0.74
8	Att.3	A1	0.99

Table 6

Attribute Difficulty

Attributes	Attribute probability 1
Linguistic skills (lexical) in the source language	0.814
Linguistic skills (grammatical) in the source language	0.808
Linguistic skills (lexical) in the target language	**0.932**
Linguistic skills (grammatical) in the target language	0.808
Organizational knowledge	0.818
Cultural knowledge	0.815
Translation skills	0.814
Work methodology skills	0.086
Bibliography and technical skills	**0.074**

Attribute Difficulty

Table 6 shows the performance status of the nine translation attributes. As the table shows, bibliography and technical skills, and work methodology skills, with probabilities of 7% and 8%, respectively, are the most difficult attributes. Only 7% of the students have mastered and can mobilize their bibliography and technical skills while translating a text. Conversely, linguistic skills (lexical) in the target language were shown to be the easiest attribute to master (93% probability), indicating that 93% of the students mastered and are able to mobilize linguistic skills (lexical) in the target language satisfactorily.

Attribute Correlations

CDMs are also used to calculate the correlation between the attributes to show the extent of similarities between them. Values larger than 0.70 reflect a strong correlation; 0.50 and 0.70 are considered as moderate, and less than 0.50 as weak (Henson, Templin, & Douglas, 2007; Kunina-Habenicht, Rupp, & Wilhelm, 2012).

As Table 7 shows, all the values show a strong correlation between the attributes, signifying a non-compensatory, and complementary nature of the HO-DINA model (Aryadoust, 2018; Effatpanah, 2019; Lee & Sawaki, 2009b; Ozaki, 2015; Stone & Zhang, 2003; SU, 2013; Wang, Zheng & Chang, 2014), as wll as the existence of a relationship between the translation attributes. This indicates that these attributes require almost the same underlying cognitive processes. For example, if a respondent performed well on attribute one, he/she would probably perform well on attributes 2, 4, 5, 6, and 7 as well. However, "hierarchy relationships of attributes differ from pre-requisite relationships of attributes in which attributes are ordered in difficulty" (Lim, 2015).

Class Probabilities

As Table 8 illustrates, class probabilities show how students are ranked in different latent classes. The number of latent classes in CDMs, is calculated through (2^k), where *k* is the number of attributes. Therefore, there are 2^9=512 latent classes according to the present study's Q-matrix configuration. The first column shows some of the latent classes (2^9 = 512) and the second column signifies the related attribute patterns.

For space considerations, the table shows only some of the latent classes. The data shows that the attribute profile α128= [111111100] was the most populated class with 68% probabilities, This class include approximately 135 students, followed by attribute profiles α1=[000000000] and α512 = [111111111] with 7% probabilities. These classes included approximately 13 students in each, respectively. Therefore, regarding α128, the data indicates that 135 students are expected to have mastered the first seven attributes, while not the eighth and ninth ones. Moreover, 13 students have mastered all the nine attributes, and 13 others have mastered none of the attributes.

Class Probabilities for the Respondents

The class probability of the response pattern (11111111011100001000000) is presented in Table 9. The table illustrates the probabilities that each student belonged to each of the 512 latent classes. The first column shows some of the latent classes (2^9 = 512) and the second column shows the response patterns. Table 9 shows that, for example, there are 23% and 17% probabilities for student number 1, belonging to latent classes 128 and 95, respectively. Thus, this student has a 23% chance of mastering all attributes with the exception of the eighth and ninth, while there is only a 7% probability that he/she has mastered all the attributes or belongs to latent class 512.

Skill Mastery Probabilities

Table 10 shows the skill performance probabilities of respondents on each of the required attributes for each of the test items or tasks. Due to space restrictions, the attribute performance probability of only five randomly chosen students is illustrated in Table 10. The values above 0.50 show

Table 7

Attribute Correlations

Attribute Item	Att. 1	Att. 2	Att. 3	Att. 4	Att. 5	Att. 6	Att. 7	Att. 8	Att. 9
Att. 1 (LSSL: L)	1.00								
Att. 2 (LSSL: G)	0.99	1.00							
Att. 3 (LSTL: L)	0.97	0.97	1.00						
Att. 4 (LSTL: G)	0.99	0.99	0.97	1.00					
Att. 5 (ORG)	0.99	0.99	0.97	0.99	1.00				
Att. 6 (CUL)	0.99	0.99	0.97	0.99	0.99	1.00			
Att. 7 (TRN)	0.99	0.99	0.97	0.99	0.99	0.99	1.00		
Att. 8 (WM)	0.96	0.96	0.94	0.96	0.97	0.96	0.97	1.00	
Att. 9 (BT)	0.95	0.96	0.93	0.96	0.95	0.95	0.95	0.97	1.00

Table 8

Class Probabilities

Latent Class	Attribute Pattern	Class probability	Class Expected Frequency
1	000000000	0.07	13.37
64	111111000	0.01	1.17
126	101111100	0.01	2.00
127	011111100	0.01	1.21
128	111111100	**0.68**	135.02
256	111111110	0.02	4.24
384	111111101	0.01	0.73
512	111111111	0.07	13.11

Table 9

Class Probabilities for the Respondents

Latent Class		Response Patterns
		11111111011100001000000
5	001000000	0.07
14	101100000	0.04
49	000011000	0.02
79	011100100	0.08
95	011110100	**0.17**
96	111110100	0.10
123	010111100	0.09
128	111111100	**0.23**
298	100101001	0.00
511	011111111	0.02
512	111111111	0.07

Table 10

Skill Mastery Probabilities

Test Takers	Response Pattern	Attribute Profile	Probability	LSSL: L	LSSL: G	LSTL: L	LSTL: G	ORG	CUL	TRN	WM	BT
3	00000101101101100000101	111110000	0.60	0.03	0.02	0.69	0.02	0.03	0.03	0.02	0.00	0.00
18	10111101111111110011001	011110100	0.81	0.88	1.00	0.99	1.00	0.99	0.92	1.00	0.00	0.00
83	11111111111110100100001	111111110	0.73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.27	0.02
98	11111101111111110111111	111111111	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99
131	00111101100111000001100	011011100	0.34	0.72	0.38	0.92	0.36	0.77	0.74	0.77	0.00	0.00

that the respondents have mastered the attributes with high confidence (Hu, Miller, Huggins-Manley, & Chen, 2016).

As an illustration, the probabilities that student 18 with the skill profile of [011110100] has mastered the attributes are: 0.88, 1.00, 0.99, 1.00, 0.99, 0.92, 1.00, 0.00, and 0.00, respectively. In other words, there is an 88% chance that he/she has mastered LSSL: L and 0% probability for mastering WM and BT.

DISCUSSION

Nearly all educational tests require students to become involved in different forms of cognitive processing. While trying to validate the inferences drawn from the students (Embretson, 1983, 1998; Messick, 1989; Snow & Lohman, 1989), the students' knowledge and strategies in applying any cognitive processing and problem-solving strategies during exams must be taken into account. Most measurement specialists believe that cognitive theory has a significant role in educational assessment (Frederiksen et al., 1993; Irvine & Kyllonen, 2002; Nichols et al., 1995), They consider that it can help scholars investigate the internal features of tests, assess the rules and develop innovative psychometric models, and describe the psychology which underpins students' test performance (Gierl et al., 1999; Hattie et al., 1999; Nichols, 1994; Nichols & Sugrue, 1999). In this regard, Embretson (1983) believes that cognitive theory advances psychometric practices by postulating the construct representation of a test through students' mental processes, knowledge, and the strategies they use to answer the test items. When

these cognitive needs are appropriately explained, they can be gathered into cognitive models to design items which can elicit particular mental processes and problem-solving strategies. Thus, a test score "anchored to a cognitive model is more explainable, and perhaps more meaningful, to a diverse group of users, since performance is described using a set of cognitive competences in a well-defined content area" (Leighton et al., 2004, p. 205). Cognitive models, as a "simplified description of human problem solving on standardized educational tasks which helps to characterize the knowledge and skills students at different levels of learning have acquired and to facilitate the explanation and prediction of students' performance" (Leighton & Gierl, 2007, p. 6), provide an explicit framework for linking cognitively based inferences with specific, fine-grained test score interpretations" (Gierl & Leighton, 2007; as cited in Gierl et al., 2010, p. 319).

In the field of Translation Studies, the student's ability to translate from a foreign language is a complex, integrated cognitive task related to certain underlying cognitive components (Paradis et al., 1982, as cited in Lorenzen & Murray, 2008). Translation students are expected to perform a set of translation competences, in order to translate appropriately in a test. However, an overall assessment score in such tests cannot satisfy the expectations of the teachers or examiners. Overall scores cannot distinguish between learners in terms of their skills and competences. Thus, evaluation of strengths and weaknesses requires more specific information from tests, as well as more precise data on specific attributes or skills, rather than classifying students based on their scores (Lee & Sawaki, 2009b).

CDMs can be an effective tool in measuring how attributes underlying translation interact to produce an appropriate translation, hence the decision to apply the DINO, DINA, ACDM, HO-DINA, and G-DINA models. The fit of the data to CDMs reveals the level of accuracy in the correlation between attributes and items. The results of fit statistics showed that the HO-DINA was the best-fitting model. This also follows Su's (2013) findings, who asserted that Hierarchal DINA is more applicable in small sample sizes. Moreover, HO-DINA as the best fitting model supports the assumption that "the attribute hierarchy provides an interpretative framework to guide both the development of items and the interpretation of examinees' scores, in such a way that test performance can be linked to specific cognitive inferences about examinees' knowledge and skills (Gierl, Alves, & Majeau, 2010, p. 319). Therefore, the study's proposed translation attributes can form a hierarchy which outlines the psychological ordering of the attributes or cognitive skills needed to answer the test items.

The results specify that there is a strong correlation between the translation attributes. This means that some of the attributes required almost the same underlying cognitive processes. This could serve to claim that there is a non-compensatory relation between the translation attributes, and thus translation skills are complementary and interdependent.

The findings also indicate that the two *flat-mastery profile*s called "master of all attributes" α512= [111111111] and "non-master of all attribute" α1= [000000000] were two of the most prevalent skill profiles. Approximately, 7% of the students were classified as master of none of the skills, and 7% were classified as master of all skills. The latent class 128 with an attribute profile of α128= [111111100] has a class probability of about 0.68, indicating that approximately 68% of the students are expected to have mastered the first seven attributes. According to Lee and Sawaki (2009b) and Rupp et al. (2010), flat skill profiles are caused by a high positive correlation among the attributes or unidimensionality of the evaluated scale. Thus a learner who has mastered a skill could be a master of the other skills as well.

Furthermore, the results revealed that LSTL: L is the easiest. It showed that a large number (93%) of the translation students can use specific terms in the translation of the political and journalistic texts, avoid spelling errors which cause misunderstanding about the intended meaning, translate the message and the structure of the ST expression close to the TT, and demonstrate content and meaning at a good level of accuracy in the TT (Abdi, 2019; Alibabaei, 2020; Davaninezhad, 2016; Kazemian & Vasheghani Farahani, 2020; Khatibzadeh & Sameri, 2013; Samir & Tabatabaee-Yazdi, 2020).

The analysis has also shown that BT, TRN, and WM are the most difficult translation attributes to master. Budianto and Fardhani (2010) declare that "the primary difficulty when translating a text into a second language is to produce a natural-sounding target text" (p. 5). Klimkowski (2015) also states that "42.19 percent of the translation students had insufficient skills of time and work management in successful professional functioning" (p. 79). According to Marczak and Krajka (2016), two important skills in order of importance for the translation students to learn are CAT tools as well as time management skills. The results are inconsistent with the findings by Molanazar and Kamyab (2015) who emphasize revising and editing as necessary skills that all students need to acquire at university. Khoshsaligheh et al. (2019) and Sharif (2016) state that adding more practical workshops and courses to the curriculum, such as translation of technical texts and revising and self-assessment skills, prepare students to work as qualified translators for the real workplace. Mossop (2003), additionally arrived at the same conclusion. He has suggested that, in order to overcome these deficiencies and problems, translation training programs should contain certain subjects which use advanced forms of computer software, including word processing. Today, translation students consider that word processing programs are the standard means of transforming a source text into their target text (Taghizadeh & Azizi, 2017). A sim-

ilar conclusion was reached by Abdi (2020) and Austermuhl (2014). They assert that technology is currently of paramount importance in the translation market, and as such translators must have the relevant knowledge and skills to work with national and international clients and access different data resources. Molanazar and Kamyab (2015) have noted that different computer technologies can be used in the translation production process. In view of that, in order "to survive in the Iranian translation market, it is necessary for translation students to acquire CAT tools, such as translation memories, word processing programs, terminology management systems, multilingual dictionaries, or even at times raw machine translation output" (Khazaeefar & Khoshsaligheh, 2014, p. 147).

The findings of this study can have implications for translation students and teachers. Making translation students aware of their strengths and weaknesses in translation can encourage them to concentrate on their problems and enhance their translation competences. It also enables translation teachers to have a better understanding of the competences where translation students face more difficulties.. Building on the results from the students and the weaknesses encountered in TRN, WM, and BT skills, teachers can aid and support students by offering practical workshops in producing a target text (TT) to an acceptable level of fluency, managing the time, and using relevant bibliography, or electronic terminology instruments. These workshops could help students deal with difficult times in translation performance (Kafi et al., 2018; Taghizadeh & Azizi, 2017; Samir et al. 2018). Besides, the proper identification of translation students' competences leads teachers to apply effective teaching techniques and construct translation tests suitable for students' performance and levels.

CONCLUSION

The present study, aimed at investigating the strengths and weaknesses of Iranian Translation students at BA level in translation and diagnosing their translation competences, has confirmed the usefulness of the TQAR in assessing the quality of translation in different genres. The findings indicate that HO-DINA cannot only classify the translation student's skill mastery/non-mastery reliably and properly but also offer informative and valid information about the learning status of translation students. The findings also advance the improvement of translation assessment by offering diagnostic information regarding cognitive processes and the relation between the translation's attributes. Accordingly, more concern must be taken toward educational assessment in translation programs, which requires the help and support of different specialists in the field of translation studies.

Importantly, this study is unique in that, for the first time, CDMs have been implemented in Translation Studies and more specifically in Translation Quality Assessment. However, the current study has a significant limitation that needs to be discussed. The present study analyzed the results of only 18 English Translation departments in a number of universities in Iran. Thus, future research could evaluate the strengths and generalizability of the findings in other countries and universities. Despite the findings recorded, further research is required to discover, test, and compare the existence of different probable hierarchies among the study's proposed attributes.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Mona Tabatabaee-Yazdi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Aynaz Samir: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

REFERENCES

Abdi, H. (2020). Translation and technology: Investigating the employment of computer-aided translation (CAT) tools among Iranian freelance translators. *Theory & Practice in Language Studies*, *10*(7), 811-818. http://doi.org/10.17507/tpls.1007.13

Abdi, H. (2019). A skopos-theoretical analysis of political texts from English into Persian: The case of Donald Trump's executive orders. *Theory and Practice in Language Studies*, *9*(7), 884-890. http://doi.org/10.17507/tpls.0907.20

Alibabaei, F. (2020). English translation of Persian metaphors in Iranian political news discourse: A relevance theory approach. *Journal of Language and Discourse Practice*, *1*(2), 1-10.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on nAutomatic Control, AC-19*, 716–723. http://doi.org/10.1109/TAC.1974.1100705

Angelelli, C. V., & Jacobson, H. E. (2009). Introduction. Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice. In *Testing and assessment in Translation and Interpreting Studies* (pp. 1-10). John Benjamins. http://doi.org/10.1075/ata.xiv

Angelone, E. (2013). The impact of process protocol self-analysis on errors in the translation product. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, *8*(2), 253-271. http://doi.org/10.1075/tis.8.2.07ang

Aryadoust, V. (2018). A cognitive diagnostic assessment study of the listening test of the Singapore–Cambridge general certificate of education O-level: Application of DINA, DINO, G-DINA, HO-DINA, and RRUM. *International Journal of Listening*, 1-24. https://doi.org/10.1080/10904018.2018.1500915

Austermuhl, F. (2014). *Electronic tools for translators*. Routledge. http://doi.org/10.4324/9781315760353

Beeby, A. (2000). *Teaching translation from Spanish to English.* University of Ottawa Press. https://doi.org/10.2307/j.ctt1cn6sgb.10

Beeby, A., Fernández Rodríguez, M., Fox, O., Kuznik, A., Neunzig, W., Rodríguez-Inés, P., & Wimmer, S. (2011). Results of the validation of the PACTE translation competence model: Translation project and dynamic translation index. In S. O'Brien (Ed.), *Cognitive explorations of translation* (pp. 30-53). Bloomsbury. http://doi.org/10.1075/btl.94.22pac

Birenbaum, M., Kelly, A. E., & Tatsuoka, K. K. (1993). Diagnosing knowledge states in algebra using the rule-space model. *Journal for Research in Mathematics Education*, *24*(5), 442-459. http://doi.org/10.2307/749153

Bonyadi, A. (2020). A skill-based approach for assessing EFL students' translation performance. *Journal of Language and Translation*, *10*(2), 35-46.

Budianto, L., & Fardhani, A. E. (2010). *A practical guide for translation skills*. UIN-Maliki Press.

Chen, B. (2013, December 6-8). *Globalization and film titles translation* [Paper presentation]. Proceedings of the international symposium on globalization: Challenges for translators and interpreters, Zhuhai, China.

Chen, H. L., & Chen, J. S. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly, 13,* 218–230. https://doi.org/10.1080/15434303.2016.1210610

Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement, 37*(6), 419-437. https://doi.org/10.1177/0146621613479818

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265-289. http://doi.org/10.2307/1165285

Conde, T. (2013). Translation versus language errors in translation evaluation. In D. Tsagari, & R. Vam Deemter (Eds.), Assessment issues in language translation and interpretation (pp. 97-112). Peter Lang AG.

Conde, T. (2012). Quality and quantity in translation evaluation: A starting point. *Across Languages and Cultures*, *13*(1), 67-80. http://doi.org/10.1556/Acr.13.2012.1.4

Davaninezhad, F. K. (2016). Supplementary practices and decreasing translation problems in translating political English news texts into Persian. *Theory and Practice in Language Studies*, *6*(1), 184-193. http://doi.org/10.17507/tpls.0601.24

De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76* (2), 179-199. http://doi.org/10.1007/s11336-011-9207-7

De la Torre, J. (2008). An empirically-based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45*(4), 343-362. http://doi.org/10.1111/j.1745-3984.2008.00069.x

De la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333-353. http://doi.org/10.1007/BF02295640

De la Torre, J., & Lee, Y. S. (2013). Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, *50*(4), 355-373. http://doi.org/10.1111/jedm.12022

De la Torre, J., & Lee, Y. S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement, 47*(1), 115-127. https://doi.org/10.1111/j.1745-3984.2009.00102.x

De la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa, 20*, 89–97. https://doi.org/10.1016/j.pse.2014.11.001

Dewi, H. D. (2015). *Comparing two translation assessment models: Correlating student revisions and perspectives* [Unpublished doctoral dissertation]. Kent State University.

DiBello, L. V., & Stout, W. (2007). Guest editor's introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement, 44*(4), 285-291. http://doi.org/10.1111/j.1745-3984.2007.00039.x

DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of Statistics, Psychometrics* (vol. 26, pp. 979-1030). Elsevier.

DiBello, L. V., Roussos, L. A., & Stout, W. (2006). 31a review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of statistics, 26*, 979-1030. http://doi.org/10.1016/S0169-7161(06)26031-0

DiBello, L. V., Stout, W. F., & Roussos, L. (1995). Unified cognitive psychometric assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-390). Erlbaum.

Doyle, M. S. (2003). Translation pedagogy and assessment: Adopting ATA's framework for standard error marking. *The ATA Chronicle*, *32*(11), 21-28.

Effatpanah, F. (2019). Application of cognitive diagnostic models to the listening section of the international English language Testing System (IELTS). *International Journal of Language Testing, 9*, 1-28. http://doi.org/10.1186/s40468-019-0090-y

Effatpanah, F., Baghaei, P., Boori, A. (2019). Diagnosing EFL learners' writing ability: A diagnostic classification modeling analysis. *Language Testing in Asia, 9*(12). https://doi.org/10.1186/s40468-019-0090-y

Embretson, S. E. (1999). Cognitive psychology applied to testing. In F. T. Durso (Ed.), *Handbook of applied cognition* (pp. 629–660). Wiley & Sons.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. [Empirical Study]. *Psychological Methods, 3*(3), 380-396. https://doi.org/10.1037/1082-989x.3.3.380

Embretson, S. E. (1983). Construct validity: Construct representation vs. nomothetic span. *Psychological Bulletin*, *93*(1), 179-197. https://doi.org/10.1037//0033-2909.93.1.179

Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika*, *84*(1), 19-40. http://doi.org/10.1007/s11336-018-09658-x

Farahzad, F. (1992). Testing achievement in translation classes. In C. Dollerup, & A. Loddegaard (Eds.), *Teaching translation and interpreting: training, talent, and experience* (pp. 271-278). John Benjamins. http://doi.org/10.1075/z.56.41far

Fišer, D. (2008). The teaching and learning of ICT skills for translators. In R. Dimitru & K. Freigang (Eds.), *Translation Technology in Translation Classes* (pp. 103-12). Institutul European.

Frederiksen, N., Mislevy, R. J., Bejar, I. I. (1993). *Test theory for a new generation of tests.* Erlbaum. https://doi.org/10.2307/2348147

Galeshi, R. (2012). *Cognitive Diagnostic model, a simulated-based study: Understanding Compensatory Reparameterized Unified Model (CRUM)* [Unpublished doctoral dissertation]. Virginia Tech.

Genesee, F. (2002). Evaluation. In R. Carter, & D. Nunan (Eds.), *The Cambridge guide to teaching english to speakers of other languages* (4th ed., pp. 144-150). Cambridge University Press.

Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education*. Cambridge University Press.

Gierl, M. J., Alves, C., & Majeau, R. T. (2010). Using the attribute hierarchy method to make diagnostic inferences about examinees' knowledge and skills in mathematics: An operational implementation of cognitive diagnostic assessment. *International Journal of Testing*, *10*(4), 318-341. http://doi.org/10.1080/15305058.2010.509554

Goff-Kfouri, C. A. (2005). Testing and evaluation in the translation classroom. *Translation Journal, 9*(2), 75-99.

Hansen, G. (2010). Translation 'errors'. *Handbook of translation studies*, *1*, 385-388. http://doi.org/10.1075/hts.1.tra3

Hartz, S. M. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practicality* [Unpublished doctoral dissertation]. University of Illinois.

Hatim, B. & Mason, I. (1997). *The Translator as communicator*. Routledge.

Hattie, J., Jaeger, R. M., & Bond, L. (1999). Persistent methodological questions in educational testing. *Review of Research in Education, 24,* 393-446. http://doi.org/10.2307/1167275

Henson, R., Templin, J. L., & Douglas, J. (2007). Using efficient model-based sum-scores for conducting skills diagnoses. *Journal of Educational Measurement, 44*(4), 361–376. http://doi.org/10.1111/j.1745-3984.2007.00044.x

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using Log-Linear Models with latent variables. *Psychometrika, 74*(2), 191-210. http://doi.org/10.1007/s11336-008-9089-5

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Hu, J., Miller, M. D., Huggins-Manley, A. C., & Chen, Y.-H. (2016). Evaluation of model fit in cognitive diagnosis models. *International Journal of Testing, 16*(2), 119-141. https://doi.org/10.1080/15305058.2015.1133627

Hurtado Albir, A. (2015). The acquisition of translation competence. Competences, tasks, and assessment in translator training. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, *60*(2), 256-280. http://doi.org/10.7202/1032857ar

Irvine, S. H., & Kyllonen, P. C. (Eds.). (2002). *Item generation for test development.* Erlbaum.

Jang, E. E. (2005). *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL* [Unpublished doctoral dissertation]. University of Illinois.

Jiménez-Crespo, M. A. (2020). Localization. In E. Bielsa & D. Kapsaskis (Eds.), The Routledge handbook of translation and globalization (pp. 375-390). Routledge. http://doi.org/10.4324/9781003121848-31

Jiménez-Crespo, M. A. (2013). *Translation and web localization*. Routledge. http://doi.org/10.4324/9780203520208

Jun-song, W. A. N. G. (2017). Translation quality assessment: Bridge the gap between theory and practice. *Journal of Literature and Art Studies*, *7*(10), 1701-1706. http://doi.org/10.17265/2159-5836/2017.10.023

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272. https://doi.org/10.1177/01466210122032064

Kabát, M. (2020). The competencies of software translators from the point of view of Slovak translation studies. *Bridge: Trends and Traditions in Translation and Interpreting Studies*, *1*(1), 57-67.

Kafi, M., Khoshsaligheh, M., & Hashemi, M. R. (2018). Translation profession in Iran: current challenges and future prospects. *The Translator*, *24*(1), 89-103. http://doi.org/10.1080/13556509.2017.1297693

Kazemian, R., & Vasheghani Farahani, M. (2020). Translating metadiscourse features in TED political talks: A parallel corpus-based investigation of semantic underspecification in English and Persian. *Language and Translation Studies (JLTS)*, *53*(2), 127-154. https://doi.org/10.22067/lts.v53i2.85499

Kelly, D. (2005). *A handbook for translator trainers: A guide to reflective practice*. Sent Jerome Publishing.

Khanmohammad, H. O. M., & Osanloo, M. (2009). Moving toward objective scoring: A rubric for translation assessment. *Journal of English language Studies, 1*, 131-153.

Khatibzadeh, P., & Sameri, M. (2013). Translation of binomials in political speeches and reports; A contrastive study of English and Persian. *Journal of Translation and Interpretation, 6*(1), 18-33.

Khazaeefar, A., & Khoshsaligheh, M. (2014). Insights from EMT for translator training in Iran. *Ferdowsi Review: An Iranian Journal of TESL, Literature and Translation Studies*, *1*(1), 135-152.

Khoshsaligheh, M., Moghaddas, M., & Ameri, S. (2019). English translator training curriculum revisited: Iranian trainees' perspectives. *Teaching English Language*, *13*(2), 181-212. https://doi.org/10.22132/tel.2019.96238

Kim, A. Y. A. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, *32*(2), 227-258. http://doi.org/10.1177/0265532214558457

Kiraly, D., & Hofmann, S. (2016). Towards a postpositivist curriculum development model for translator education. In D. Kiraly, L. Rüth, C. Canfora, & A. Cnyrim (Eds.), *Towards authentic experiential learning in translator education* (pp. 67–87). Mainz University Press at V&R Unipress. https://doi.org/10.14220/9783737004954.67

Klimkowski, K. (2015). Entrepreneurial training in translator and interpreter education. *Roczniki humanistyczne*, *63*(11), 67-83. http://doi.org/10.18290/rh.2015.63.11-5

Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models. *Journal of Educational Measurement, 49*(1), 59–81. https://doi.org/10.1111/j.1745-3984.2011.00160.x.

Lauscher, S. (2000). Translation quality assessment: Where can theory and practice meet? *The translator*, *6*(2), 149-168. http://doi.org/10.1080/13556509.2000.10799063

Lee, Y. W., & Sawaki, Y. (2009a). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6*, 239–263. https://doi.org/10.1080/15434300903079562

Lee, Y. W., & Sawaki, Y. (2009b). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, *6*(3), 172–189. http://doi.org/10.1080/15434300902985108

Leighton, J. P., & Gierl, M. J. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and practices*. Cambridge University Press. http://doi.org/10.1017/CBO9780511611186

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's Rule-Space Approach. *Journal of Educational Measurement, 41*(3), 205-237. http://doi.org/10.1111/j.1745-3984.2004.tb01163.x

Leighton, J., Gierl, M. J., & Hunka, S. (1999, April). *Attributes in Tatsuoka's rule-space model.* Poster presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec, Canada.

Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 9*, 17-46.

Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, *33*(3), 391-409. https://doi.org/10265532215590848.

Lim, Y. (2015). *Cognitive diagnostic model comparisons* [Unpublished doctoral dissertation]. Georgia Institute of Technology.

Lorenzen, B., & Murray, L. L. (2008). Bilingual aphasia: A theoretical and clinical review. *American Journal of Speech-Language Pathology, 17*(3). https://doi.org/10.1044/1058-0360(2008/026)

Mackenzie, R. (2004). The competencies required by the translator's roles as a professional. In Malmkjær, K. (Ed.), *Translation in undergraduate degree programs* (pp. 32-38). John Benjamins Publishing Company. http://doi.org/10.1075/btl.59.04mac

Marczak, M., & Krajka, J. (2016). Translator education in the cloud: Students' perceptions of telecollaborative experiences. In E. Smyrnova-Trybulska (Ed.), *E-learning methodology-implementation and evaluation* (pp. 369-388). Uniwersytet Slaski.

Maris, E. (1999). Estimating multiple classification latent class models, *Psychometrika*, *64*(2), 197-212. http://doi.org/10.1007/BF02294535

Martínez-Carrasco, R. (2017). *Epistemological approaches to legal translation education: A situated account* [Unpublished doctoral dissertation]. Universitat Jaume I.

Maydeu-Olivares, M. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives, 11*, 71-137. https://doi.org/10.1080/15366367.2013.831680

McDonald, R. P., & Mok, M. M. C. (1995). The goodness of fit in item response models. *Multivariate Behavioral Research*, *30*, 23-40. https://doi.org/10.1207/s15327906mbr32-001

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Macmillan.

Molanazar, H., & Kamyab, S. (2015). Political and journalistic translation courses and the market demand in Iran. *European Online Journal of Natural and Social Sciences,4*(2), 285-293.

Mossop, B. (2003). What should be taught at translation school? In Pym, A., C. Fallada, J. Ramon Biau, & J. Orenstein (Eds.). *Innovation and E-learning in translator training* (pp. 20-22). Universitat Rovira i Virgili.

Newmark, P. (1988). *A Textbook of Translation*. Prentice Hall.

Nichols, P. (1994). A framework of developing cognitively diagnostic assessments. *Review of Educational Research, 64*, 575-603. http://dx.doi.org/10.3102/00346543064004575

Nichols, P. D., Chipman, S. F., & Brennan, R. L. (1995). *Cognitively diagnostic assessment*. Erlbaum. https://doi.org/10.4324/9780203052969

Nichols, P., & Sugrue, B. (1999). The lack of fidelity between cognitively complex constructs and conventional test development practice. *Educational Measurement: Issues and Practice, 18*, 18-29. https://doi.org/10.1111/j.1745-3992.1999.tb00011.x

Nord, C. (1999). Translating as a text-production activity. *On-Line Symposium on Innovation in Translator and Interpreter Training, Interpreter Training*, 3(4), 11-19.

Ozaki, K. (2015). DINA models for multiple-choice items with few parameters: Considering incorrect answers. *Applied Psychological Measurement*, *39*(6), 431-447. http://doi.org/10.1177/0146621615574693

PACTE. (2011). Results of the validation of the PACTE translation competence model: Translation problems and translation competence. In C. Alvstad, A. Hild, & E. Tiselius (Eds.). *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 317–343). John Benjamins. https://doi.org/10.1075/btl.94.22pac

PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation: Perspectives in process oriented research* (pp. 43-66). John Benjamins. https://doi.org/10.1075/btl.45.06pac

Paradis, M., Goldblum, M. C., & Abidi, R. (1982). Alternate antagonism with paradoxical translation behavior in two bilingual aphasic patients. *Brain and Language*, *15*(1), 55-69. http://doi.org/10.1016/0093-934X(82)90046-3

Piecychna, B. (2013). Legal translation competence in the light of translational hermeneutics. *Studies in Logic, Grammar, and Rhetoric, 34*(47), 141–159. http://doi.org/10.2478/slgr-2013-0027

Polliastri, P., & Paulina, A. (2009). Evaluation criteria for the improvement of translation quality. In M. Forstner, H. Lee-Jahnke, & P. A. Schmitt (Eds.). *Enhancing translation quality: Ways, means, methods* (pp. 239-261). Peter Lang AG.

Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, *34*, 782-799. https://doi.org/10.1177/0734282915623053.

Ravand, H., & Baghaei, P. (2020). Diagnostic classification models: Recent development, practical issues, and prospects. *International Journal of Testing*, *20*, 24-56. https://doi.org/10.1080/15305058.2019.1588278

Ravand, H., Baghaei, P., & Doebler, P. (2020). Examining parameter invariance in a general diagnostic classification model. *Frontiers in Psychology*, *10*, 2930. https://doi.org/10.3389/fpsyg.2019.02930

Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology, 38*(10), 1255–1277. https://doi.org/10.1080/01443410.2018.1489524.

ReaDickins, P., & Germaine, K. (1993). Evaluation. Oxford University Press.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications.* Guilford Press. https://doi.org/10.1080/15366367.2018.1434349

Samir, A., Motallebzadeh, K., Ashraf, H., & Ghonsooly, B. (2018). Development and validation of an MA English translation curriculum effectiveness questionnaire: Using the Rasch Model. *North American Journal of Psychology*, *20*(3), 601-618.

Samir, A., Tabatabaee-Yazdi, M. (2020). Translation quality assessment rubric: A Rasch model-based validation. *International Journal of Language Testing, 10*(2), 101-128.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464. https://doi.org/10.1214/aos/1176344136

Sharif, D. F. (2016). Evaluating the translation competency areas in the Iranian national translators' training curriculum. *International Journal of Innovation and Research in Educational Sciences, 5*(1), 340-349.

Sidiropoulou, M. (2004). *Linguistic identified through translation*. Rodopi. http://doi.org/10.1163/9789004486652

Snell-Hornby, M., Pochhacker, F., & Kaindl, K., (Eds). (1994). *Translation studies: An interdiscipline (Benjamins Translation Library 2)*. John Benjamins Publishing. http://doi.org/10.1075/btl.2

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263-331). American Council on Education, Macmillan.

Stobart, G., & Gipps, C. (1997). *Assessment: A teacher's guide to the issues* (3rd ed.). Hodder & Stoughton.

Stone, C. A., & Zhang, B. (2003). Assessing goodness of fit of item response theory models: A comparison of traditional and alternative procedures. *Journal of Educational Measurement*, *40*(4), 331-352. http://doi.org/10.1111/j.1745-3984.2003.tb01150.x

Su, Y. L. (2013). *Cognitive diagnostic analysis using hierarchically structured skills* [Unpublished doctoral dissertation]. University of Lowa. https://doi.org/10.17077/etd.5xs1u621

Tabatabaee-Yazdi, M., Motallebzadeh, K., Baghaei, P. (2021). A Mokken scale analysis of an English Reading Comprehension Test. *International Journal of Language Testing, 11*(1), 132-143.

Tabatabaee-Yazdi, M. (2020). Hierarchical Diagnostic Classification Modeling of reading comprehension. *SAGE Open*, *10*(2). http://doi.org/10.1177/2158244020931068

Taghizadeh, M., & Azizi, M. (2017). Exploring computer-aided translation competencies of some Iranian translators. *International Journal of English Language & Translation Studies*, *5*(1), 78-87.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345-354. http://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, *30*, 251-275. https://doi.org/10.1007/s00357-013-9129-4

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287-305. https://doi.org/10.1037/1082-989x.11.3.287

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*, 125–145. https://doi.org/10.1177/014662168400800201

Waddington, C. (2001). Different methods of evaluating student translations: The question of validity. *Meta: Journal Des Traducteurs/Meta: Translators' Journal*, *46*(2), 311-325. https://doi.org/10.7202/004583ar

Wang, C., Zheng, C., & Chang, H. H. (2014). An enhanced approach to combine item response theory with cognitive diagnosis in adaptive testing. *Journal of Educational Measurement*, *51*(4), 358-380. https://doi.org/10.1111/jedm.12057

William, K. (2003). *Understanding media theory*. Arnold.

Yamaguchi, K., & Okada, K. (2018). Comparison among cognitive diagnostic models for the TIMSS 2007 fourth grade mathematics assessment. *PlOS ONE*, *13*(2). http://doi.org/10.1371/journal.pone.0188691

APPENDIX A

Assessment Criteria	Attributes								
	linguistic skills in the SL		Linguistic Skills in TL						
	1. SL Lexical	2. SL Grammatical	3. TL Lexical	4. TL Grammatical	5. Organizational Knowledge	6. Cultural Knowledge	7. Translation skills	8. Work methodology skills	9. Bibliography and technical skills
Grammar (Word Form/Part of Speech, Word Order, Syntax)	0	1	0	1	0	0	1	0	0
Usage	1	1	1	1	1	1	0	0	0
(No) Addition	1	1	1	1	1	1	1	0	0
(No) Omission	1	1	1	1	1	1	1	0	0
Completeness	0	0	0	0	1	0	1	0	0
Punctuation	0	1	0	1	0	0	0	0	0
Terminology/False Friend Terminology	1	0	1	0	0	0	1	0	0
Spelling	0	0	1	0	0	0	0	0	0
Capitalization/ Italicization Rules	1	1	1	1	0	0	1	0	0
Faithfulness/literalness	1	1	1	1	1	1	1	0	0
Register/Tone	1	1	1	1	1	1	1	0	0
Genre (Text Style, Text type)	1	1	1	1	1	0	1	0	0
Cohesion/Coherence, Consistency	0	0	1	1	1	0	1	0	0
Accuracy	1	1	1	1	1	1	1	0	0
Fluency)Naturalness, Readability, No Ambiguity)	1	1	1	1	1	1	1	0	0
Creativity/Problem Solving (No Indecision)	1	0	1	1	1	1	1	0	0
17. Organization/time management	0	0	0	0	0	0	0	1	0
18. Initiative	0	0	0	0	0	0	0	1	0
19. Pace of work	0	0	0	0	0	0	0	1	0
20. Revision file, self-assessment	0	0	0	0	0	0	0	1	0
21. Quality of terminological data base	0	0	0	0	0	0	0	0	1
22. CAT skills	0	0	0	0	0	0	0	0	1
23. Relevance of bibliography	0	0	0	0	0	0	0	0	1

APPENDIX B

Items	Rate by the Raters				Rate by the Researchers	
	0		1		0	1
	1 (poor)	2 (Fair)	3 (Advanced)	4 (Superior)	Mastered	Non-Mastered
Grammar (Word Form/ Part of Speech, Word Order, Syntax …)						
Usage						
(No) Addition						
(No) Omission						
Completeness						
Punctuation						
Terminology/False Friend Terminology						
Spelling						
Capitalization/ Italicization Rules						
Faithfulness/literalness						
Register/Tone						
Genre (Text Style, Text type)						
Cohesion/Coherence, Consistency						
Accuracy						
Fluency)Naturalness, Readability, No Ambiguity)						
Creativity/Problem Solving (No Indecision)						
17. Organization/Time management						
18. Initiative						
19. Pace of work						
20. Revision file/Self-assessment						
21. Quality of terminological data base						
22. CAT skills						
23. Relevance of bibliography						
Total						

https://doi.org/10.17323/jle.2023.12399

Effectiveness of cooperative grouping in developing reading skills of university level EFL learners

Jagoda Topalov

University of Novi Sad, Novi Sad, Serbia

ABSTRACT

Background. Collaborative Strategic Reading (CSR) is an instructional method in which students work together in a cooperative framework, jointly construct a model of text and come to its potential meaning through discussions.

Purpose. A quasi-experimental pretest-posttest non-equivalent control group research design was used to examine the effects of cooperative grouping within the framework of CSR, with the aim of determining whether cooperative grouping is effective in developing EFL students' reading skills.

Method. The response variables included the students' scores on questions testing Vocabulary, Factual Information, Prose Summary, Sentence Simplification, Reference Question and Insert text, whereas the explanatory variable was group membership (+/- cooperative), measured across three testing times (the beginning, middle and the end of the experimental intervention).

Results. The results indicate that the students exposed to CSR within cooperative groups significantly developed those reading skills which focus on the comprehension of global information – prose summary, insert text and reference question. A possible explanation is that, in order to answer these questions, readers must approach the text in a holistic manner and focus on its main ideas, which seems to be facilitated by discussions in heterogeneous teams and negotiations of meaning resulting from those discussions.

Conclusion. The main pedagogical implication of the results concerns the need for introducing cooperative grouping as an alternative to a typical university-level foreign language classroom, allowing teachers to organize an effective, interactive context for reading academic texts in English.

KEYWORDS

EFL reading skills, EFL reading instruction, cooperative learning, Collaborative Strategic Reading, university students, quasi-experimental design

Citation: Topalov J. (2023). Effectiveness of cooperative grouping in developing reading skills of university level EFL learners. *Journal of Language and Education, 9*(1), 158-171. https://doi.org/10.17323/jle.2023.12399

Correspondence:
Jagoda Topalov,
jagoda.topalov@ff.uns.ac.rs

Received: May 11, 2021
Accepted: September 01, 2022
Published: March 31, 2023

INTRODUCTION

In the academic context of tertiary education one of the most useful and most employed skills is reading, whereas reading in English, with its growing significance as a lingua franca, places a set of new, highly demanding challenges on EFL learners. The expectations placed on university students to be able to read academic texts in English are increasing, which is why curriculum designers and practitioners alike are tasked with putting forward most effective ways in which EFL reading instruction can be offered. The urgency of adopting an improved approach to reading instruction is, in the Serbian educational context, strikingly apparent in the results of the Programme for International Student Assessment (PISA)[3] for 2018, according to which Serbian students score below the OECD average on the reading proficiency measure, with only 3% of students able to "comprehend lengthy texts, deal with concepts that are abstract or counterintuitive, and establish distinctions between fact and opinion, based on

[1] PISA (Programme for International Student Assessment). (2018). PISA 2018 results. https://nces.ed.gov/surveys/pisa/pisa2018/index.asp#/

implicit cues pertaining to the content or source of the information".[4] Having recognized the need to bring reading instruction up-to-date, the policy-makers from the Serbian Ministry of Education, Science and Technological Development in their Framework of the National Curriculum of the Republic of Serbia[5] underline that cooperative learning is one of the methods which leads to higher achievement and better learning gains, in contrast with individual and competitive learning. In addition to improved learning gains, cooperative learning also answers to the challenges of the changing labour market in which "employers across the world give increasingly higher priority to behavioural skills than to technical competencies" (Cunningham et al., 2010). Teamwork, collaboration and sociability are ranked high in UNESCO's framework for transversal competencies[6] and a timely integration of these inter-personal skills in education would ensure that the students are adequately prepared for the world of work and are equipped to live meaningful, sustainable, and responsible lives in a world that is rapidly becoming interconnected.

Despite the potential benefits of cooperative learning, the organization and dynamics of the teaching process and the implementation of different teaching approaches and strategies in the Serbian educational context, as evidenced by PISA scores, in most cases remain very static and rigid (Milić, 2016). As students are rarely exposed to cooperative framework and seldom taught how to take part in it, previous research shows that, even though they have moderately positive attitudes towards group work and team activities, they exhibit considerably low willingness to participate in it (Topalov et al., 2015; Topalov et al., 2017). In an educational context in which the individual method is promoted and practiced, this study aims at investigating the extent to which structured, heterogeneous cooperative grouping is an effective method in developing reading skills of young adult EFL learners. For this purpose, drawing on the social interdependence theory (Johnson & Johnson, 2005; 2015), a quasi-experimental pretest-posttest non-equivalent control group research design was implemented to examine whether there was a statistically significant change in the students' scores on the reading comprehension measure between the students who were exposed to reading instruction within cooperative groups and those who were taught reading in a traditional classroom setting.

Student Interaction in the Foreign Language Classroom

Recent decades saw a growing interest of researchers and educators in learning through interaction and cooperation, in response to the weaknesses of the teacher-centred classroom. Although the traditional method allows the teachers to more easily manage all aspects of classroom organization, individual differences among students are marginalized, the teaching materials are aimed at the average student, while both gifted and struggling students are side-lined, and the responsibility for achieving lesson aims is largely assumed by the teacher (Estes et al., 2014).

Cooperative learning appeared as one of the responses to the observed need for a shift in the nature of authority and responsibility in the classroom. Social interdependence theory (Johnson & Johnson, 2005; 2015) provides a theoretical basis for the investigation of the effects of cooperative work and has, thus far, been validated by hundreds of research studies (Johnson and Johnson (2008) offer a highly informative overview of studies investigating cooperative learning procedures). As it accounts for how the outcomes of social interaction are affected both by individual actions of those who participate in interaction, as well as by the actions of other participants, it defines cooperative learning as a "pedagogical strategy in which small groups of students work together to achieve a common goal, with each individual responsible for their own learning as well as the success of their peers" (Davies et al., 2013, p. 564). Its aim is to contribute to the development of cognitive and social skills through selected techniques, wherein the learning process is rigorously structured and the tasks used are clearly defined and regulated; the students are responsible both to themselves and to the group; the group represents the core of all classroom interactions, with the teacher's main task being to provide support (Johnson et al., 2014).

Johnson and Johnson (2009) indicate the following four basic principles of cooperative learning: (a) positive interdependence, according to which the success of an individual is linked to the success of the rest of the group in such a way that an individual can only achieve the goal if the other members of the group reach that goal as well; it is promoted by sharing goals, means for achieving those goals and setting boundaries between group members to determine who is interdependent with whom; (b) individual and group accountability, with individual responsibility being reflected in the results of individual assessment and testing, and group accountability in group assessment; (c) promotive interaction, which takes place when group members support each other's efforts taken towards the accomplishment of the group's goals by providing mutual help and assistance, exchanging needed resources, effectively communicating, exerting mutual influence and trust, and constructively managing conflict; and (d) appropriate use of social skills, which

[4] Schleicher, A. (2018). *PISA 2018. Insights and interpretations*. OECD. http://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf

[5] ONKRS (Okvir nacionalnog kurikuluma Republike Srbije). (2015). https://obrazovanje.org/rs/uploaded/dokumenta/Okvir-nacionalnog-kurikuluma_Osnove-ucenja-i-nastave.pdf

[6] UNESCO. (2015). *Transversal competencies in education: Policy and practice*. UNESCO. http://unesdoc.unesco.org/images/0023/002319/231907E.pdf

include personal and small-group skills needed for effective cooperation; in order to coordinate efforts for the purpose of accomplishing shared goals team members must get to know and trust each other, communicate clearly and accurately, support each other, and constructively resolve any conflicts (Johnson & Johnson, 2016).

In examining foreign language learning contexts, this framework has been shown to be effective in overall foreign language acquisition (Law, 2011; Wei & Tang, 2015), including English language fluency (Alrayah, 2018), productive English skills (Fen, 2011), oral proficiency (Namaziandost et al., 2020) and motivation (Ning & Hornby, 2014). Rather than simply motivating individuals to exert greater effort, positive interdependence as part of cooperative dynamics facilitates the development of new insights and the use of higher-level reasoning through promotive interaction (Johnson & Johnson, 2008).

Cooperative Framework of Reading Instruction

Turning to the pedagogy of reading, Collaborative Strategic Reading – CSR is a combination of cooperative learning and modified reciprocal teaching (Bremer *et al*., 2002) and is based on a socio-cognitive theory of reading that emphasizes the importance of social context in the cognitive development of reading skills (Klingner et al., 2012). In this approach to comprehension, the reader actively decodes the text, uses adequate prior knowledge, applies the cognitive resources at his or her disposal, and develops understanding of the text through structured social interaction. It is an approach to teaching comprehension strategies that places students, their cognition and ability to self-regulate at the centre of the reading process. CSR is an instructional method in which students work together in a cooperative framework, jointly construct a model of text, and come to its potential meaning through discussions (Vaughn et al., 2011). In this learning context, students achieve cognitive development that they would not otherwise be able to achieve individually through the help and support of their micro-community, i.e. the team within which they are reading. By working together with their peers on the construction of meaning, they internalize cognitive strategic knowledge through dialogue within small groups (Fan, 2015).

CSR is conducted over two phases. In the first phase, students are exposed to reading strategies, while in the second phase, cooperative teams are formed in which students continue to apply the strategic framework. The principles of modified reciprocal teaching are reflected in the first phase of the reading approach, which uses a combination of reading strategies. Within CSR, the teacher exposes students to groups of strategies that include: 1) initial review of the text, 2) monitoring comprehension, 3) summarizing parts of the text/getting the main idea, and 4) final evaluation. During the initial review of the text, the students should apply two strategies: first, they make predictions about the text, based on the title, subtitle, pictures, etc., and then activate their previous knowledge of the topic. During reading, the students monitor comprehension and are aware of the moment when there is a breakdown in understanding. Within CSR, students are instructed to use several strategies, including identifying contextual information that would help them understand an unknown word or phrase, extending the context to a sentence that precedes or follows the sentence with the unknown word, and parsing the word and identifying roots, prefixes and suffixes. In applying the strategy of summarizing, the students focus on the most important ideas in a particular section of the text and ignore less important information. After reading, during the final evaluation, students are instructed to formulate questions and provide answers, as well as to summarize the entire text in order to test comprehension. Cooperative elements are evident in the second phase, in the structured roles that students receive when interacting, as well as in the structured materials designed to foster positive interdependence and individual and group accountability. When working with students in cooperative roles, the teacher supervises groups, facilitates in the comprehension of unfamiliar words, demonstrates how strategies and cooperative learning techniques are used, and provides assistance (Klingner et al., 2012).

To date, a number of empirical studies examined the effectiveness of CSR in reading instruction, focusing primarily on reading instruction in the first language, with participants of different ages, as well as with students with learning difficulties (Boardman et al., 2016; Kim et al., 2006; Klingner & Vaughn, 2000; Vaughn et al., 2011). Several studies have examined the impact of collaborative strategic framework on the development of foreign language. For instance, in a study with students at the A2 level of CEFR, CSR positively affected the students' reading comprehension scores, with significant positive effects on the identification of the main idea and supporting details (Fan, 2010). CSR also positively impacted students' ability to deal with vocabulary-related comprehension breakdowns (Karabuga & Kaya, 2013), their learning habits and attitudes (Mendieta et al., 2015) and their willingness to participate in cooperative reading tasks (Zoghi et al., 2010). In a study by Topalov and Radić-Bojanić (2016) CSR had positive effects on university students' affinity towards reading texts in the English language, on the degree in which they read English texts and on their belief that they are able to complete academic reading tasks. However, no study, to the best of our knowledge, has focused on the following two aspects investigated in this research: firstly, the cooperative component of CSR alone has not been investigated in such a way that the strategic input was administered to both conditions, while only the factor of cooperative work/individual work differed between the experimental and the control groups, as is the case in this research; secondly, this study examines the context of reading in a foreign language with young adults learning EFL at university, who are at a higher level of foreign language pro-

ficiency and who, therefore, have functional independence in language use and are able to read complex academic texts that require readers to activate higher cognitive levels of text processing, as well as procedural and conditional knowledge of reading strategies (studies conducted thus far in an EFL context have examined students at lower levels of proficiency, or did not include information on the participants' level of English). In practical terms, the rationale for this study is, thus, twofold. On the one hand, in teasing apart the relative contribution of the cooperative component to the success CSR has in developing EFL reading skills, as attested by previous research (Fan, 2010; Karabuga & Kaya, 2013; Zoghi et al., 2010), in addition to providing its empirical validation, this study will examine the potential benefits of introducing the cooperative component into an educational setting that is still largely oriented towards individual work. On the other hand, by using a sample of students at a higher level of EFL proficiency, it will investigate the effectiveness of CSR beyond the level of simple textual comprehension and will focus on its potential usefulness in achieving a holistic understanding of complex texts.

The goal of this study is, therefore, to determine whether the cooperative component significantly contributes to the development of EFL reading skills of university level students. In view of this goal, this study aimed at answering the following research question: Is there a statistically significant change in the students' scores on the reading comprehension measure over time (8 months) between the students who are exposed to reading instruction within cooperative groups and those who are taught reading in a traditional classroom setting?

METHOD

Design

A quasi-experimental pretest-posttest non-equivalent control group research design was applied in this study to examine the effects of cooperative grouping within CSR on the development of EFL reading skills operationalized as the score on the TOEFL reading comprehension test of university students.

The within-subjects factor in the study was Testing Time, with data collected at three different points: at baseline, after four months and after eight months. Between-subjects factor was Group, which included two levels: the Experimental Group followed the strategic framework of CSR within cooperative groups (+cooperative), while the Control Group followed the same strategic framework through individual work (-cooperative) (see Table 1). This intervention setup effectively eliminated the potential for contamination between experimental and control conditions, since both groups received the same strategic framework.

Participants

This study used single stage sampling. The sampling procedure consisted of non-random, convenience sampling, as intact classes were readily accessible (Gall *et al*., 2007). In terms of the validity of this sampling method, when each control classroom is compared to treatment classrooms within the same school, it eliminates school-based differ-

Table 1

The Design of the Experiment

	Collection of dependent variables	Independent variable manipulation	Collection of dependent variables	Independent variable manipulation	Collection of dependent variables
Experimental Group	Baseline test	+ cooperative	Progress test	+ cooperative continued	Post-test
Control Group	Baseline test	- cooperative	Progress test	- cooperative continued	Post-test
Input:	A combination of reading strategies as part of 1) initial review of the text, 2) monitoring comprehension, 3) summarizing parts of the text/getting the main idea, and 4) final evaluation, in a +/- cooperative setting				

ences in those comparisons and provides an opportunity to replicate program impacts across multiple sites (Hitchcock et al., 2009). This further ensures that reading instruction techniques will likely not vary if both classrooms are from the same school, so that minimizing such differences in non-experimental factors will improve power and quality of the results and reduce the necessary sample size. A total of 50 first- and second-year students from the Faculty of Philosophy, University of Novi Sad participated in this study. They were all students from various humanities and social science departments (including the Departments of Psychology, Pedagogy, Philosophy, History, Media Studies, Serbian Literature, Comparative Literature and Serbian Philological Studies), who were taking a course in the English language as a requirement of their respective study program. On average, the participants were 20.1 years old, with the ages ranging between 18 and 22. They were all at B2 level of the *Common European Framework of Reference – CEFR* attested by the results of *Quick Pen and Paper Test*.[7] In the context of the present study this means that students were expected to have intermediate knowledge of the language in order to be able to understand main ideas of complex texts on both concrete and abstract topics and to adjust their reading to the purpose and the nature of the text (Jones, 2002). The participants in this study attended two courses in English as a foreign language (English B2.1 and English B2.2) during two consecutive semesters, as part of a larger cohort of 81 students who were also placed at B2 level. This, in turn, means that roughly two thirds of all students at B2 level took part in this study.

Of the total number of participants, 24 were in the Experimental Group (11 male and 13 female students) and 26 in the Control Group (12 male and 14 female students). Since group equivalence could not be assumed, the gender structure of the sample was internally controlled for appropriateness with a chi-square test. The results indicate there is no statistically significant difference between the two groups (x^2(1, 50)=.001, p>.01). Furthermore, in order to determine whether groups differed prior to the beginning of the experimental input, all participating students were pretested on reading level including its subcomponents (Vocabulary, Factual Information, Prose Summary, Sentence Simplification, Insert Test, Reference Question). No statistically significant differences in the reading level between the Experimental and the Control Groups were found on questions testing the students' knowledge of Vocabulary (mean difference=-0.11, t=-.427, p>.01), Factual Information (mean difference=0.13, t=.1.431, p>.01), Prose Summary (mean difference=-0.12, t=-1.089, p>.01), Sentence Simplification (x^2(1, 50)=2.889, p>.01), Insert Test (x^2(1, 50)=.774, p>.01), Reference Question (x^2(1, 50)=1.035, p>.01), or on their overall reading comprehension score (t=.031, p>.01).

Classroom Intervention

Both groups had 90-minute classes two times a week for the duration of eleven weeks in the first semester and twelve weeks in the second, with the same teacher. Furthermore, both groups were exposed to the same set of reading strategies within the CSR framework, as was elaborated previously in the Introduction (a combination of reading strategies as part of a) initial review of the text, b) monitoring comprehension, c) summarizing parts of the text/getting the main idea, and d) final evaluation). The students in the Control Group applied the strategies with the guidance of the teacher. In this approach, the teacher taught the strategies by encouraging students to form predictions about the text, by asking questions with which the students' previous knowledge was activated, by asking them to come up with the gist of a paragraph and key words, and by giving the students the task of writing short summaries and formulating questions. On the other hand, the manner in which this strategic input was used and practiced with the Experimental Group was modified as classes progressed, from a general discussion about strategies, over teacher modelling, to students' application in reading, first with scaffolded assistance and then, finally, without it (Reynolds, 2017).

During the second stage, which began two weeks into the experiment, heterogeneous cooperative teams of four were formed in the Experimental Group. The grouping took into account the results of the baseline reading comprehension test, so that, whenever possible, groups consisted of both students with higher and lower levels of reading proficiency, and were mutually balanced. Each team member was given a specific role (leader, gist expert, unknown words expert, reporter, cf. Klingner *et al*., 2012) and these roles alternated in regular three-week intervals so that every student could experience different tasks and responsibilities associated with a specific role. The materials were highly structured, and included learning logs and role cards, all with the purpose of fostering positive interdependence and individual and group accountability.

At this point it is also necessary to acknowledge that the factor of the teacher may potentially be considered a confounding factor, one whose potential influences would be eliminated had the sample included more than one of intervention and control groups each. Although, due to practical reasons, it was impossible for the sample to include more groups in the same school during the same time period, as none were available, it nonetheless does not put at ease doubts arising from this experimental set-up. In order to mitigate this possible threat to the validity of the study, the data will be investigated using statistical tests that are designed to increase the power of the study (a Repeated

[7] Oxford University Press/University of Cambridge/Association of Language Testers in Europe (2006). *Quick Placement Test: Paper and Pen Test*. Oxford University Press.

Measures ANOVA in place of two One-Way ANOVAs) and the discussion of the results will proceed cautiously.

Instruments

In order to determine the students' proficiency in the English language, *the Quick Pen and Paper Placement Test* was administered prior to the commencement of the experiment. The test consists of 60 questions with multiple-choice answers, arranged in order of increased proficiency. The test is able to identify the test-taker's level of proficiency relative to the six proficiency levels outlined by CEFR.

The students' reading skills were tested by means of parallel forms of TOEFL IBT Reading Comprehension Test[8]. The section of TOEFL that assesses EFL reading skills is designed to replicate the types of tasks that university students face when reading in an academic context (Jamieson et al., 1999). The following types of questions were included (Cohen & Upton, 2006):

(1) Vocabulary – measured the ability to understand the meaning of particular words and phrases in context (3 items, 1 point each);

(2) Factual Information – examined the ability to find the answer to a question about an important fact explicitly written in the text (3 items, 1 point each);

(3) Prose Summary – assessed the ability to understand the main ideas in the text, the relative importance of information and to distinguish main ideas from minor information or the ideas that are not in the text (1 item graded on a 3-point scale);

(4) Sentence Simplification – examined the ability to recognize main information in a longer, complex sentence consciously ignoring minor details and elaborations (1 item, 1 point);

(5) Insert Test – assessed the ability to find lexical, grammatical, and logical connections in the sequence of sentences by inserting a new sentence into the most appropriate position in the passage that was already read (1 item, 1 point);

(6) Reference Question – evaluated the ability to recognize the links between anaphoric words and their antecedents or postcedents in the text (1 item, 1 point).

The following TOEFL tests were used in chronological order of testing: *Early Cinema*, *The Expression of Emotions* and *Artisans and Industrialization*.

Procedure

Parallel forms of the reading comprehension test were administered at three different testing times: at baseline, four months into the intervention and upon the completion of the intervention, eight months from the beginning. The baseline test was administered during the first week of classes (beginning of October), following which the initial results testing for group equivalence were computed, allowing the experimental treatment to begin in the second week of semester. The second test was administered during the last week of first semester, after eleven weeks of classes (end of December), whereas the final test was administered after another twelve weeks of classes, during the last week in the second semester (end of May). Students from both groups completed the test individually during a 45-minute session. While there was no missing data, the data from a total of three participants were lost due to attrition (one participant only completed the initial test, while two more completed the initial and the second test, but not the final test). The data from these participants were excluded prior to conducting all relevant statistical tests.

Analysis

The collected data were analysed using descriptive and inferential statistical tests. The data were checked for normality, following which a series of Repeated Measures (RM) ANOVAs and Cochran's Q tests were applied, relative to the nature of data.

The data were analysed using SPSS v.20 statistics software.

RESULTS

Following the presentation of the results of descriptive tests, the results in this section will be given in order of the inferential tests that were applied in the analysis of data. First subsection will show the results of RM ANOVAs performed on the continuous variables of Vocabulary, Factual Information, Prose Summary and Total Reading Score. The second subsection will present the results of Cochran's Q tests, which is considered to be an alternative to the RM ANOVA test,[9] performed on the dichotomous variables of Sentence Simplification, Reference Question and Insert Text.

Table 2 shows the descriptive statistics for the continuous variables of Vocabulary, Factual Information, Prose Summary and Total Reading Score.

In examining the distribution of data, the results indicate that the data is either moderately skewed (values between −1 and −½ or between +½ and +1) or approximately symmet-

[8] ETS – Educational Testing Service. (2009*). The official guide to TOEFL® Test*. McGraw-Hill.
[9] Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Table 2

Descriptive Statistics for Vocabulary, Factual Information, Prose Summary and Overall Reading Score across Three Testing Times

	Variable	Group	Mean	SD	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis
Test 1 - Baseline	Vocabulary	Experimental	2.13	0.69	-.547	.536	-.584	1.038
		Control	2.24	0.70	-.368	.501	-.764	.972
	Factual information	Experimental	2.44	0.63	-.227	.536	-1.516	1.038
		Control	2.31	0.80	-.346	.501	-1.008	.972
	Prose summary	Experimental	1.67	0.30	-.210	.536	-.472	1.038
		Control	1.79	0.37	-1.464	.501	.652	.972
	Total reading	Experimental	8.28	1.74	-.583	.536	-.293	1.038
		Control	8.27	1.53	.169	.501	-.329	.972
Test 2 – 4 months from the beginning	Vocabulary	Experimental	2.84	0.32	-1.461	.536	.137	1.038
		Control	2.40	0.56	-.339	.501	-1.050	.972
	Factual information	Experimental	2.21	0.66	-.482	.536	-.188	1.038
		Control	2.11	0.61	-.235	.501	-.218	.972
	Prose summary	Experimental	2.31	0.87	-1.397	.564	1.816	1.091
		Control	1.65	0.34	-.413	.501	-.870	.972
	Total reading	Experimental	9.23	2.17	-.368	.536	-1.100	1.038
		Control	8.64	1.33	-.511	.501	.025	.972
Test 3 – 8 months from the beginning	Vocabulary	Experimental	2.50	0.51	-.121	.536	-2.267	1.038
		Control	2.33	0.58	-.128	.501	-.537	.972
	Factual information	Experimental	2.06	1.00	-.920	.536	.080	1.038
		Control	2.14	0.79	-.272	.501	-1.312	.972
	Prose summary	Experimental	1.75	0.55	-2.399	.536	5.675	1.038
		Control	0.59	0.87	.984	.550	-.916	1.063
	Total reading	Experimental	8.97	1.88	-1.083	.536	1.272	1.038
		Control	7.05	2.06	-.222	.501	-.385	.972

ric (values between −½ and +½,) and that the excess kurtosis in the tested variables shows both positive results, indicating more outliers than normality, and negative, indicating fewer outliers. The kurtosis values are between -1.513 and 1.816, which is considered acceptable in proving normal univariate distribution, for all except the following two variables: Vocabulary and Prose Summary for which the data were collected during the third testing time. This will be addressed in the following section.

For the binary variables of Sentence Simplification, Reference Question and Insert Test, percentages of correct answers are shown in Table 3.

The following sections outline the results of statistical analyses conducted with the aim of answering if there is a statistically significant change in the students' scores on the reading comprehension measure over time (8 months) between the Experimental and the Control Group.

Vocabulary, Factual Information and Prose Summary

A series of mixed within-between repeated measures ANOVAs (RM ANOVAs) were performed to compare the scores of the students from the Experimental and the Control Group across three different sets of questions (Vocabulary, Factual Information and Prose Summary) used to measure EFL reading level at three different points in time (baseline, after 4 months, after 8 months). As previously stated, a repeated measures ANOVA was chosen in place of two one-way ANOVAs in an attempt to increase the power of the study; furthermore, even though all the dependent variables were drawn from the same TOEFL test, possibly suggesting that

Table 3

Percentages of Correct Answers for Sentence Simplification, Reference Question and Insert Text across Three Testing Times

	Group	
	Experimental	**Control**
Sentence Simplification – Test 1	79.0%	76.2%
Sentence Simplification – Test 2	92.4%	90.2%
Sentence Simplification – Test 3	83.3%	82.4%
Reference Question – Test 1	61.1%	71.4%
Reference Question – Test 2	83.3%	73.5%
Reference Question – Test 3	94.4%	82.4%
Insert Text – Test 1	61.8%	69.5%
Insert Text – Test 2	95.2%	61.1%
Insert Text – Test 3	88.9%	76.2%

MANOVA would be a more appropriate statistical test, no significant correlations were found between the dependent variables, providing further confirmation of the appropriateness of a repeated measures ANOVA. The author is aware of the risk of inflating Type I error by choosing this statistical test and will proceed cautiously with the interpretation of the results. Bonferroni post-hocs were performed for within-subjects simple effects and t-tests, as planned comparisons, for between-group simple effects. A preliminary testing of the model assumptions for RM ANOVA was conducted to check for normality and homogeneity of variances, including sphericity; no serious violations were discovered in all cases except with the variables of Vocabulary and Prose Summary for which the data were collected during the third testing time. The results revealed that, although the sphericity assumption was met, the data for these variables violated the assumptions of homoscedasticity and normality of distribution, which prompted a transformation of data, using a log base 10 transformation.[10] Following this, the transformed data were checked and the results indicated that the data met the necessary assumptions.

The results of the RM ANOVA showed that there was a statistically significant main effect of Testing Time (F=3.587, p<.05, η_p^2=.08) and of Group (F=4.106, p<.05, η_p^2=.10) on the students' results on the Vocabulary questions. Bonferroni post-hoc analysis was implemented to isolate the statistically significant mean differences for Testing Time, as a within-group variable. The results reveal that the Experimental Group made statistically significant progress at second testing time when compared to the first (p<.01). Furthermore, in order to make post-hoc comparisons between conditions and determine simple effects, three independent samples t-tests were conducted for each testing time. The results reveal that the Experimental Group outperformed the Control Group during the second testing time (p<.01). In contrast, the results showed that there was no statistically significant effect of the interaction between Testing Time and Group. The effect sizes for both main effects, reported above as partial eta-squared, were considered medium (Cohen, 1988).

With respect to the tasks testing the variable of Factual Information, both main effects and the interaction effect were not statistically significant, with participants from the Experimental and Control Group showing similar scores during three different testing times.

Additionally, in testing the effect of the experimental treatment on the dependent variable of Prose Summary, a statistically significant main effect was found for Testing Time (F=15.693, p<.001, η_p^2=.284) and Group (F=25.976, p<.001, η_p^2=.456), as well as for the interaction between Testing Time and Group (F=9.001, p<.001, η_p^2=.352). The Bonferroni post-hoc analysis yielded several statistically significant mean differences: firstly, the Experimental Group made progress on the second test compared to the first one (p<.01) and, secondly, the Control Group performed statistically significantly worse on the third test compared to both the first (p<.001) and the second test (p<.001). Pairwise comparisons between groups revealed that the Experimental Group outperformed the Control on both the second (p<.01) and the third test (p<.001). The reported effect sizes for the main effects and the interaction effect (partial eta-squared) are considered large (Cohen, 1988).

Sentence Simplification, Reference Question and Insert Text

In contrast to the continuous dependent variables analysed above, the dependent variables of Sentence Simplification,

[10] Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Insert Text and Reference Question were binary variables, which prompted the use of Cochran's Q, as an alternative to the RM ANOVA test.

When investigating within-group differences for Sentence Simplification, Cochran's Q test indicated that no statistically significant results were obtained for either the Experimental or the Control Group. The Experimental Group's scores on the Reference Question task, on the other hand, yielded a statistically significant result (Q=9.222, p<.01). A pairwise post-hoc Dunn test revealed that the difference was statistical both for the second test and the third test when compared to the first test (p<.01). As SPSS does not return the results for effect sizes when conducting Cochran's Q test, the chance-corrected measure of effect size was separately calculated following Berry, Johnston and Mielke (2018). The analysis returned a value of R=.121, indicating approximately 12% within-group agreement above what is expected by chance. The scores of the Control Group showed no statistically significant variation with respect to Testing Time. Likewise, in testing the variable of Insert Text, statistically significant results were found in the Experimental Group (Q=11.455, p<.01), but not in the Control Group. A pairwise post-hoc Dunn test indicated that the Experimental Group achieved a statistically significantly better result on the second and the third test compared to the first test (p<.01). The value of the chance-corrected measure of effect size was R=.213, indicating approximately 21% within-group agreement above what is expected by chance.

In order to test between-group differences with the dichotomous variables of Sentence Simplification, Insert Text and Reference Question, a series of Mann-Whitney U tests were performed, with the results revealing that prior to the beginning of the intervention there were no statistically significant differences between the groups. Furthermore, no statistically significant differences between the groups were found 4 months into the experiment, as well as upon the completion of the experiment for the tasks of Sentence Simplification and Reference Question. In contrast, the difference between the Control Group's and the Experimental Group's scores for the variable of Insert Text was discovered to be statistically significant on the second test, in favour of the Experimental Group (U=103.500, p<.001).

Overall Reading Comprehension

Finally, the participants' overall scores on the reading comprehension measure were analysed by means of an RM ANOVA test. The data were first checked for normality and homogeneity of variances, including sphericity, with no violations observed. The results of RM ANOVA reveal that there was a statistically significant main effect for Group (F=4.668, p<.05, η_p^2=.11) and a statistically significant interaction between Testing Time and Group (F=3.516, p<.05, η_p^2=.08), suggesting that the effect of time depends on whether the participants performed reading tasks in cooperative groups or individually. Specifically, post-hoc testing revealed that the Experimental Group overall performed statistically significantly better than the Control Group on the second test (p<.01) and the third test (p<.01). The effect sizes for the main effect and the interaction effect, indicated above as partial eta-squared, were considered medium. No statistically significant simple effects were discovered for within-subject differences.

DISCUSSION

The current study investigated the effectiveness of the cooperative component within Collaborative Strategic Reading – CSR framework (Klingner et al., 2012) in improving EFL reading skills of young adult learners who are learning English in a university setting. The study aimed at answering whether there was a statistically significant change in the students' scores on the reading comprehension measure over time (8 months) between the Experimental and the Control Group. The findings indicate that the students exposed to reading instruction in cooperative teams performed better than the students in a teacher-led classroom with respect to a number of reading tasks, including on the overall reading scores at the end of the experimental treatment.

Specifically, in answering Vocabulary questions the students in the Experimental Group scored better in the middle of the experiment when compared to the initial test, but not on the final test, which leads to the conclusion that progress in the second semester cannot be confirmed. The Experimental Group also outperformed the Control Group on the second test, but not on the final test, prompting the need to altogether re-examine the approaches to teaching comprehension of explicit textual information, as neither cooperative teamwork, nor teacher-led instruction seem to be beneficial. In the cooperative approach, members of the team decode the text together and deal with breakdowns in comprehension by applying the prescribed strategies outlined in the Methods section. The drawback is that, without a dictionary confirmation or the involvement of the teacher, the students have no way of knowing if their guess is correct, or if they should even doubt it. This would require both metacognitive and metalinguistic awareness, which is an unrealistic expectation for many learners (Soto et al., 2020). On the other hand, in the teacher-led approach, the teacher can ensure that the students reach the correct meaning of the word; however, the teacher is unable to monitor for comprehension breakdowns in every single student, since here, as well, we are relying on the students' metacognitive awareness that they will be able to recognize when they do not understand what they are reading. At this point it is also necessary to address a limitation of this study that is generally related to vocabulary questions. Namely, even though TOEFL test preparation guidelines state that the topics of the texts are general and, thus, do not favour experts in a particular scientific or professional field, or persons with

specific prior knowledge of the topic of the text, it is possible that the topic of emotions presented from a socio-cultural and psychological perspective on the second test was nevertheless closer to students than the labour union theme presented from a political-historical perspective on the third test. Readers with adequate prior knowledge of the subject area of the text are able to create complex and correct text models (Karimi, 2018), with prior knowledge being significantly associated with vocabulary knowledge and success in understanding the text (Bernhardt, 2011; Martínez, 2022). Furthermore, a number of authors have qualms as to what is actually measured by vocabulary tasks, with the results of certain studies indicating that in about a fifth of cases what is actually measured is the reader's ability to make correct inferences (Kremmel & Schmitt, 2016; Schmitt, 1999). In light of the current study, the primary objective of vocabulary tasks was not to test whether the respondents know the meaning of a particular word, but to verify that they understand the meaning of the word in the context of the sentence or the text in which the word appears and that in the case of polysemous words they can decide which meaning is adequate in a given context. In other words, context plays a very important role in every act of text comprehension, which raises the issue of differentiating between the learner's knowledge of a word in context and the learner's skill to understand the word due to contextual clues. This is a question that is beyond the scope of this paper and that is yet to be answered.

With respect to the tasks testing the variable of Factual Information, which examined the reader's ability to find the answer to a question about an important fact explicitly written in the text, no statistically significant main effects were found, with participants from the Experimental and Control Group showing similar scores during three different testing times. The results were already relatively high for both groups, possibly suggesting positive effects of transfer of their L1 reading skills (Sparks et al., 2009). Furthermore, with the initial mean scores already high, it is rather improbable that the students would be able to make significant progress regardless of the framework of instruction they were exposed to.

In testing the variable of Prose Summary, the results indicate that the students from the Experimental Group achieved statistically significantly higher results than the students from the Control Group both on the second test and on the final test. Since summary is one of the strategies that is explicitly taught within the strategic input to which both groups were exposed, the results in favour of the Experimental Group suggest that the cooperative framework offers a more beneficial context for the acquisition and implementation of the strategy than the teacher-centred approach. Within the latter approach, students are essentially free to decide whether and to what extent they will use their cognitive capacities and apply the strategy, whereas the teacher receives confirmation that the student is participating only when the student is called upon to give the answer in front of the whole class. On the other hand, students working together in smaller groups have to participate more actively in the implementation of this strategy, since they are responsible for their participation not only to the teacher, but also to other members of the group. Furthermore, a comparative analysis of the results of the Control Group at three testing times indicates that the Control Group performed statistically significantly worse on the third test compared to both the first and the second test. A possible explanation for such an unexpected result may be found in the very dynamics of practicing writing summaries within the traditional framework. Specifically, the teacher exposed both groups to the strategy, after which both groups practiced writing a summary each time they read a text in English. In the Experimental Group, each of the student teams would read their abstract, after which the teacher and the other students would comment on whether the summary was correct, whether it adequately conveyed the main idea of the text, its key terms, etc. In the Control Group, on the other hand, students practiced this strategy individually, after which a relatively small number of students read the summary in front of the class and received feedback from the teacher, so that the teacher could not check whether all the summaries in the Control Group were correct, or if all the students were paying attention to comments that would help them solve the task properly. Another explanation for this result may lie in the formulation of the task itself. In the first and second tests, where no statistically significant differences were found in the results, students had to choose three of the six sentences offered that best express the main idea of the text. This task was somewhat more complex on the third test. The students first had to select from among seven sentences five that best represent the ideas in the text and then place them in the appropriate category. Although the change in the task is minimal, it is still sufficient to indicate a significant deficiency in the way students in the Control Group approached the understanding of the text and its key parts, suggesting that the knowledge of solving the summary task remained at the declarative level (cf. Anderson, 2013) as the students failed to generalize that knowledge and come up with ways to apply it in a modified context.

With respect to the task of Sentence Simplification, since the results indicate that students exposed to either of the treatments did not make statistically significant gains in their scores, it is impossible to draw conclusions. Neither the traditional nor the cooperative approach can be connected with improvement in skills required to successfully complete this task, leaving open to speculation which instructional treatment may be beneficial in this respect.

It was further found that the statistically significant differences in scores on the tasks of Insert Text and Reference Question were found with the Experimental Group's scores on the second and third tests compared to the initial test, suggesting that the benefits of working in cooperative

teams in terms of finding lexical, grammatical, and logical connections in the sequence of sentences and recognizing the links between anaphoric words and their antecedents or postcedents are visible after a relatively short period of time (four months after the beginning of the experiment) and that longer exposure under these conditions does not contribute to further development of knowledge and skills. Furthermore, as both Insert Text and Reference Question assess the reader's ability to focus on the macrostructure of the text, it appears that cooperative framework can be associated with higher gains in scores on the so-called *global* tasks, i.e. tasks that require learners to view the text as a whole and make inferences based on linguistic input (Becker, 2016). This may be due to the dynamics of group work in cooperative teams, in which the students, in order to accomplish tasks, must make their thinking process explicitly known to the rest of the group, provide arguments for their vantage points and, as a group, negotiate the meaning of the text. This, in turn, may lead to a more developed awareness of the text as a whole, of its main ideas and key issues, which is in line with Goldman and associates (2016).

Finally, it was determined that the Experimental Group significantly outperformed the Control Group on the overall reading measure, both during second and during third testing session, but also that no statistically significant differences were found when the groups' scores were compared individually across different testing times. These results suggest that the cooperative framework was more beneficial than the teacher-centred instruction in developing the students' reading skills. As no within-group differences were discovered, the results further suggest that the strategic framework itself cannot be associated with any gains in reading skills with young adult EFL learners.

CONCLUSION

The main focus of this study was the investigation of the effectiveness of the cooperative component of the Collaborative Strategic Reading framework in developing EFL learners' reading skills. The results offer a mixed pattern of findings, with the most conclusive differences between students exposed to different experimental treatments established with respect to the reading comprehension questions that mostly focus on global information – Prose Summary and, to an extent, Insert Text and Reference Question. In correctly answering these questions, readers must approach the text in a holistic manner and focus on its main ideas, which seems to be facilitated by discussions in structured heterogeneous teams and negotiations of meaning resulting from those discussions.

The findings in this report are subject to several limitations, among which is certainly the instrument used to assess the students' reading proficiency. The issues concerning questions that test the readers' knowledge of vocabulary have already been discussed in the previous section and it has been noted that in one fifth of these tasks the skill of drawing conclusions based on contextual information is measured instead of the knowledge of vocabulary. Also, given the importance of prior knowledge in developing reading comprehension, it is possible that the topics of the texts favoured certain readers, putting others at a disadvantage. Furthermore, the study used a convenience sample of small size that in all likelihood reduced the statistical power of the results. With a small sample size, caution must be applied as the findings might not be directly transferable to other similar classroom contexts.

Notwithstanding these limitations, the present study hopefully adds to the ever-growing body of research on cooperative learning particularly by emphasizing the practical implications that follow from its results. A cooperative learning environment can provide an alternative to a typical university-level foreign language classroom, which mainly includes large groups of students and in which, due to group size constraints, the teacher uses the traditional, lecture-style method most of the time and tries to establish and maintain control over all aspects of classroom organization. Cooperative teams within CSR allow teachers to organize an effective, interactive context for reading academic texts in English, within which they can more easily identify the weaker and stronger sides of individual students and provide assistance at the right time to those who need it. The research findings also highlight the importance of feedback in the development of metacognitive awareness when reading. In the traditional teaching method, relatively few students receive feedback from teachers about their work, as opposed to cooperative teams, in which students receive feedback on a continuous basis, both from their teacher and, to a large extent, from other members of the group, which helps them form a clearer picture of their current knowledge and skills.

DECLARATION OF COMPETITING INTEREST

None declared.

REFERENCES

Alrayah, H. (2018). The effectiveness of cooperative learning activities in enhancing EFL learners' fluency. *English Language Teaching*, *11*(4), 21-31. https://doi.org/10.5539/elt.v11n4p21

Anderson, J. (2013). *Rules of the mind* (e-book ed.). Psychology Press. https://doi.org/10.4324/9781315806938

Becker, A. (2016). L2 students' performance on listening comprehension items targeting local and global information. *Journal of English for Academic Purposes*, *24*, 1-13. https://doi.org/10.1016/j.jeap.2016.07.004

Bernhardt, E. B. (2011). *Understanding advanced second-language reading*. Routledge.

Berry, K., Johnston, J., & Mielke, P. (2018). *The measurement of association. A permutation statistical approach*. Springer. https://doi.org/10.1007/978-3-319-98926-6_2

Boardman, A. G., Vaughn, S., Buckley, P., Reutebuch, C., Roberts, G., & Klingner, J. (2016). Collaborative strategic reading for students with learning disabilities in upper elementary classrooms. *Exceptional Children*, *82*(4), 409-427. https://doi.org/10.1177/0014402915625067

Bremer, C., Vaughn, S., Clapper, A., & Kim, A. (2002). Collaborative Strategic Reading (CSR): Improving secondary students' reading comprehension skills. In *Research to practice brief* (vol. 1, pp. 2-9). National Center on Secondary Education and Transition.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohen, A. & Upton, T. (2006). *Strategies in responding to the New TOEFL reading tasks* (TOEFL Monograph No. MS-33). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2006.tb02012.x

Cunningham, W., Sanchez-Puerta, M. L., & Wuerml, A. (Eds.), (2010). *Active labor market programs for youth. A framework to guide youth employment interventions*. World Bank. http://hdl.handle.net/10986/11690

Davies, R. S., Dean, D. L., & Ball, N. (2013). Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. *Educational Technology Research and Development*, *61*, 563-580. https://doi.org/10.1007/s11423-013-9305-6

Estes. M. D., Ingram, R., & Liu, J. C. (2014). A review of flipped classroom research, practice, and technologies. International HETL Review, *4*. https://www.hetl.org/feature-articles/a-review-of-flipped-classroom-research-practice-and-technologies

Fan, Y. (2010). The effect of comprehension strategy instruction on EFL learners' reading comprehension. *Asian Social Science*, *6*(8), 19-29. https://doi.org/10.5539/ass.v6n8p19

Fan, Y. (2015). Fostering learner autonomy through a socio-cognitive model of reading comprehension instruction. *British Journal of Education, Society & Behavioural Science*, *9*(2), 105-117. https://doi.org/10.9734/BJESBS/2015/18161

Fen, G. (2011). Theme-based group teaching of college oral English: Endorsed by students in China EFL context. *English Language Teaching*, *4*(1), 33-41. https://doi.org/10.5539/elt.v4n1p33

Gall, M., Gall, J., & Borg, W. (2007). *Educational Research: An Introduction* (8th ed.). Allyn & Bacon.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C. D., & Shanahan, C. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, *51*(2), 219-246. http://dx.doi.org/10.1080/00461520.2016.1168741

Hitchcock, J., Kurki, A., Dimino, J., Wilkins, C., & Gersten, R. (2009). Evaluating the collaborative strategic reading intervention: An overview of randomized controlled trial options. *Practical Assessment, Research & Evaluation*, *14*, Article 2. https://doi.org/10.7275/sakx-4c50

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (1999). *TOEFL 2000 framework: A working paper* (TOEFL Monograph No. MS-16). Educational Testing Service.

Johnson, D., & Johnson, R. (2005). New developments in social interdependence theory. *Genetic, Social, and General Psychology Monographs, 131*(4), 285-358. https://doi.org/10.3200/MONO.131.4.285-358

Johnson, D., & Johnson, R. (2008). Social interdependence theory and cooperative learning: The teacher's role. In R. M. Gillies, A. F. Ashman, & J. Terwel (Eds.), *The teacher's role in implementing cooperative learning in the classroom* (pp. 9-37). Springer. https://doi.org/10.1007/978-0-387-70892-8_1

Johnson, D., & Johnson, R. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365-379. https://doi.org/10.3102/0013189X09339057

Johnson, D. W., & Johnson, R. T. (2015). Theoretical approaches to cooperative learning. In R. M. Gillies (Ed.), *Collaborative learning: Developments in research and practice* (pp. 17-46). Nova Science.

Johnson, D. W., & Johnson, R. (2016). Cooperative learning and teaching citizenship in democracies. *International Journal of Educational Research*, *76*, 162-177. https://doi.org/10.1016/j.ijer.2015.11.009

Johnson, D., Johnson, R., & Smith, K. (2014). Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal on Excellence in University Teaching*, *25*(4), 1-26.

Jones, N. (2002). Relating the ALTE framework to the Common European Framework of Reference. In Council of Europe (Eds.), *Case studies on the use of the Common European Framework of Reference* (pp. 167-183). Cambridge University Press.

Karabuga, F. & Kaya E. S. (2013). Collaborative strategic reading practice with adult EFL learners: A collaborative and reflective approach to reading. *Procedia – Social and Behavioral Sciences*, *106*, 621-630. https://doi.org/10.1016/j.sbspro.2013.12.071

Karimi, M. N. (2018). The mediated/unmediated contributions of language proficiency and prior knowledge to L2 multiple-texts comprehension: A structural equation modelling analysis. *Applied Linguistics*, *39*(6), 912-932. https://doi.org/10.1093/applin/amw059

Kim, A., Vaughn, S., Klingner, J., Woodruff, A., Klein Reutebuch, C., & Kouzekanani, K. (2006). Improving the reading comprehension of middle school students with disabilities through computer-assisted collaborative strategic reading. *Remedial and Special Education*, *27*(4), 235-249. https://doi.org/10.1177/07419325060270040401

Klingner, J., & Vaughn, S. (2000). The helping behaviors of fifth graders while using collaborative strategic reading during ESL content classes. *TESOL Quarterly*, *34*(1), 69-98. https://doi.org/10.2307/3588097

Klingner, J. K., Vaughn, S., Boardman, A., & Swanson, E. (2012). *Now we get it! Boosting comprehension with collaborative strategic reading*. Jossey-Bass.

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, *13*(4), 377-392. https://doi.org/10.1080/15434303.2016.1237516

Law, Y. K. (2011). The effects of cooperative learning on enhancing Hong-Kong fifth graders' achievement goals, autonomous motivation and reading proficiency. *Journal of Research in Reading*, *34*(4), 402-425. https://doi.org/10.1111/j.1467-9817.2010.01445.x

Martínez, A. C. L. (2022). The effect of perceived interest and prior knowledge on L2 reading comprehension via several assessment methods in a higher education context. *European Journal of Language and Literature*, *8*(2), 45-52. https://doi.org/10.26417/ejls-2019.v5i1-194

Mendieta, J., Múnera, L., Olmos, T., Onatra, C., Pérez, P., & Rojas, E. (2015). Fostering reading comprehension and self-directed learning in a collaborative strategic reading (CSR) setting. *Íkala, revista de lenguaje y cultura*, *20*(1), 15-42. https://doi.org/10.17533/udea.ikala.v20n1a02

Milić, S. (2016). Analiza interkulturalnih kompetencija u nacionalnom okviru kurikuluma Republike Srbije i poređenje sa nacionalnim okvirima kurikuluma referentnih zemalja [An analysis of intercultural competences in the Framework of the national curriculum of the Republic of Serbia and its comparison with the national frameworks of referent countries]. In D. Petrović & T. Jokić (Eds.), *Interkulturalno obrazovanje u Srbiji – regulativni okvir, stanje i mogućnosti za razvoj* (pp. 45-67). Centar za obrazovne politike.

Namaziandost, E., Homayouni, M., & Rahmani, P. (2020). The impact of cooperative learning approach on the development of EFL learners' speaking fluency. *Cogent Arts & Humanities*, *7*(1), 1780811. https://doi.org/10.1080/23311983.2020.1780811

Ning, H., & Hornby, G. (2014). The impact of cooperative learning on tertiary EFL learners' motivation. *Educational Review*, *66*(1), 108-124. https://doi.org/10.1080/00131911.2013.853169

Reynolds, D. (2017). Interactional scaffolding for reading comprehension: A systematic review. *Literacy Research: Theory, Method, and Practice*, *66*(1), 135-156. https://doi.org/10.1177/2381336917718820

Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation, and word-class knowledge. *Language Testing*, *16*, 189-216. https://doi.org/10.1177/026553229901600204

Soto, C., Gutierrez de Blume, A. P., Carrasco Bernal, M. A., & Contreras Castro, M. A. (2020). The role of meta-cognitive cues on the comprehension of proficient and poor readers. *Journal of Research in Reading, 43*(3), 272-289. https://doi.org/10.1111/1467-9817.12303

Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term crosslinguistic transfer of skills from L1 to L2. *Language Learning*, *59*(1), 203-243. https://doi.org/10.1111/j.1467-9922.2009.00504.x

Topalov, J, Bjelica, M., & Krombholc, V. (2015). Stavovi studenata prema kolaboraciji u univerzitetskoj nastavi engleskog jezika. [Students' attitudes towards collaboration in the university level EFL instruction]. In B. Radić-Bojanić (Ed.), *Strani jezici na Filozofskom fakultetu: primenjenolingvistička istraživanja* (pp. 107-126). Filozofski fakultet.

Topalov, J., & Radić-Bojanić, B. (2016). Collaborative strategic reading instruction as a determinant of developed affective variables. In S. Gudurić & M. Stefanović (Eds.), *Jezici i kulture u vremenu i prostoru* (pp. 507-517). Filozofski fakultet.

Topalov, J., Radić-Bojanić, B., & Bruner, D. (2017). Thai and Serbian student attitudes regarding collaborative tasks in the EFL classroom. In B. Vujin & M. Radin-Sabadoš (Eds.), *English studies today: Words and visions* (p. 445-461). Filozofski fakultet.

Vaughn, S., Klingner, J., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S., & Still-Spisak, S. J. (2011). Efficacy of collaborative strategic reading with middle school students. *American Educational Research Journal*, *48*, 938–964. https://doi.org/10.3102/0002831211410305

Wei, P., & Tang, Y. (2015). Cooperative learning in English class of Chinese junior high school. *Creative Education, 6*(03), 397. https://doi.org/10.4236/ce.2015.63039

Zoghi, M., Mustapha, R., & Maasum, T. N. R. B. T. M. (2010). Collaborative strategic reading with university EFL learners. *Journal of College Reading and Learning*, *41*(1), 67-94. https://doi.org/10.1080/10790195.2010.10850336

https://doi.org/10.17323/jle.2023.16141

Preferences for Oral Corrective Feedback: Are Language Proficiency, First Language, Foreign Language Classroom Anxiety, and Enjoyment Involved?

Watcharapol Wiboolyasarin[1], Phornrat Tiranant[2], Teavakorn Khumsat[3], Tidarat Ngamnikorn[4], Kanokpan Wiboolyasarin[5], Somkiat Korbuakaew[6], Nattawut Jinowat [6]

[1] Research Institute for Languages and Cultures of Asia, Mahidol University, Nakhon Pathom, Thailand

[2] The University of Hong Kong, Hong Kong

[3] Nanjing Tech University Pujiang Institute, Nanjing, People's Republic of China

[4] Busan University of Foreign Studies, Busan, Republic of Korea

[5] Chandrakasem Rajabhat University, Bangkok, Thailand

[6] Suan Sunandha Rajabhat University, Bangkok, Thailand

ABSTRACT

Background. The effectiveness of oral corrective feedback (OCF) in language learning is influenced by learners' comprehension and response to various OCF techniques. Therefore, it is essential for teachers to consider learners' preferences for OCF strategies.

Purpose. This quantitative study aimed to investigate the preferences of Thai as a foreign language (TFL) learners for ten commonly discussed types of OCF. Specifically, it examined whether these preferences are influenced by four learner variables: proficiency level, first language (L1), foreign language classroom anxiety (FLCA), and foreign language enjoyment (FLE).

Method. The study involved 288 university students from Chinese, Japanese, and Korean TFL settings, and the data from questionnaires were analysed using appropriate statistical methods.

Results. The findings indicate that, regardless of proficiency level, L1, FLCA, or FLE level, learners prefer more explicit OCF techniques, such as metalinguistics feedback and explicit correction. However, Korean undergraduates scored lower in the majority of OCF strategies (i.e., ignoring, elicitation, recast, explanation, and public feedback) compared to the other participants. The MANOVA analysis revealed significant differences in ignore, peer correction, recast, and private feedback based on proficiency level and L1 background. Although the differences between the FLE and FLCA approaches were not statistically significant, high FLE and FLCA groups tended to prefer more OCF strategies than the low groups.

Conclusion. This study has significant implications for instructional practices in TFL settings and for L2 lecturers in the classroom. By understanding learners' preferences for OCF, educators can tailor their instructional approaches to meet the specific needs of their students.

KEYWORDS

oral corrective feedback, preference, proficiency level, first language, anxiety, enjoyment

Citation: Wiboolyasarin W., Tiranant P., Khumsat T., Ngamnikorn T., Wiboolyasarin K., Korbuakaew S., & Jinowat N. (2023). Preferences for oral corrective feedback: Are language proficiency, first language, foreign language classroom anxiety, and enjoyment involved? *Journal of Language and Education, 9*(1), 172-184. https://doi.org/10.17323/jle.2023.16141

Correspondence:
Nattawut Jinowat,
nattawut.ji@ssru.ac.th

Received: October 12, 2022
Accepted: March 15, 2023
Published: March 31, 2023

INTRODUCTION

Oral corrective feedback (OCF) refers to remarks made by teachers or interlocutors on errors made by second language (L2) learners during speech production. The use of OCF in the language classroom is a common technique for increasing students' awareness of the mistakes (Yang, 2016). Empirical and observational studies have proved the effectiveness of OCF in accelerating language learning (Fidan, 2015; Geckin, 2020; Gómez Argüelles et al., 2019; Gooch et al., 2016 Ha & Nguyen, 2021; Lee, 2017, Li, 2018, 2021).

Previous research has highlighted that the impact of OCF on language learning is linked to learners' understanding and reaction to different OCF techniques. It has been suggested that discrepancies between teachers' and learners' interpretations of OCF could have negative impacts on language learning (Kartchava & Ammar, 2014; Lyster et al., 2013). In particular, Kartchava and Ammar (2014) found that learners' preferences for specific OCF strategies were correlated with their comprehension of teachers' error repair intentions. Therefore, it is crucial for teachers to consider their learners' preferences for OCF strategies.

However, while several studies (Deptolla, 2019; Lee, 2013; Muslem et al., 2021; Sakiroglu, 2020; Wiboolyasarin et al., 2020; Zhu & Wang, 2019) have focused on learners' general preferences or preferences for obtaining OCF for specific errors, research exploring the influence individual differences in L2 situations other than EFL on learners' preferences for OCF techniques is limited. Hence, it is necessary to further investigate the preferences of language learners for OCF techniques in various L2 contexts.

OCF Techniques and Their Significance in L2 Classrooms

The role of OCF techniques in language learning has been a subject of interest among scholars who have examined the various ways in which these techniques can facilitate L2 acquisition. Lyster and Ranta (1997) provide a succinct framework for classifying feedback, based on their observations in French immersion classrooms. Their classification includes six main types of feedback: recast, explicit correction, elicitation, repetition, clarification requests, and metalinguistic clues. More recent research has resulted in hierarchical taxonomies of OCF methods, based on theoretical understanding of how CF functions for acquisition. Two essential distinctions of OCF techniques are input-providing vs. output-prompting, and explicit vs. implicit CF, as shown in Table 1 (Ellis, 2009; Li, 2018).

There are differing opinions regarding the classification of OCF techniques. For instance, Li (2021) classifies recast and explicit correction in one category, while the other four are put in another category because the former provides the correct form, and the latter (referred to as 'prompts') motivates students to self-correct. OCF may take place both explicitly and implicitly. Explicit feedback or explicit correction comprises a blatant admission of the student's error. It is offered to provide immediate rectification or clarification of grammar. Conversely, implicit feedback happens when the source (typically the instructor) does not make it clear that the student committed an error (Tasdemir & Yalcin Arslan, 2018). The selective OCF strategies, along with their explanations and examples, are presented in Table 2.

The source of OCF has also been classified into three categories: teacher feedback, peer feedback, and self-feedback (Carless, 2006; Ha & Nguyen, 2021; Rollinson, 2005). Teacher feedback has been discussed in the preceding section, whereas peer feedback is a valuable feedback source that involves students commenting on each other's work. Peer feedback can assist L2 learners in becoming critical readers and acquiring the ability to reflect on their feedback experiences, especially when reading the work of others (Rollinson, 2005).

In the L2 classroom, OCF may be perceived as useful for letting students know when they have performed correctly, but it may also be viewed as potentially harmful because it can impair students' openness to learning (Ellis & Shintani, 2014). The affective and cognitive dimensions of OCF are widely recognised, and practitioners are cautions about offering prescriptive guidance on OCF. While OCF can have cognitive benefits, potential emotional harm also be considered. Lyster and Saito (2010) found that prompts had more significant effects than recast, indicating that prompts are likely more effective than recast in the classroom since they are more prominent. Gooch et al. (2016), in Korean EFL classrooms, found that recast only improved the comprehensibility of the sound in controlled tasks, whereas prompts were helpful for both controlled and free output. However, learners' expectations of OCF may be one of the most important factors that influence L2 acquisition, as highlighted by Kartchava and Ammar (2014). Inconsistencies between students' preferences for OCF strategies and instructors' practices when rectifying mistakes can contribute to inefficient teaching (Lyster et al., 2013). Therefore, it is essential to investigate students' perspectives on OCF, particularly their preferences for OCF categories, based on the material discussed above.

Table 1

A Taxonomy of OCF Techniques

	Implicit	Explicit
Input-Providing	Recast	Explicit Correction
Output-Providing	Repetition	Metalinguistic Explanation
	Clarification Request	Elicitation
		Paralinguistic Signal

Table 2
OCF Techniques

OCF Technique	Definition	Example
Elicitation	The instructor elicits the correct form through the use of a technique in which a part of the student's speech is repeated, omitting the incorrect portion and prompting the student to complete it independently.	S: I went there tomorrow. I: I ..... there tomorrow.
Explicit correction	The instructor accurately identifies the error made by the student and subsequently delivers a correction, providing clear indication that the error has been made.	S: I went there tomorrow. I: No, 'went' is incorrect. You should say 'I will go there tomorrow.'
Metalinguistic feedback	The instructor addresses questions or remarks and explains using grammatical or other linguistic concepts in order to elicit information from the students.	S: I went there in Monday. I: Not in Monday, We use a preposition 'on' for days and dates. We say, 'I went there on Monday.'
Recast	The instructor provides only the correct form of the student's erroneous speech without any additional commentary.	S: I went there tomorrow. I: I will go there tomorrow.
Repetition	The instructor utilises a questioning intonation to highlight the error in the student's statement, thus indicating its incorrectness.	S: I went there tomorrow. I: I *WENT* there tomorrow.
Clarification request	The instructor indicates that the student's statement is unclear or incomprehensible, and requests clarification to ensure clear communication.	S: I went there tomorrow. I: What? Can you please explain it to me more precisely?

Learners' Preferences for OCF, FLCA, and FLE

The preferences for various types of feedback have been the subject of numerous studies in the past, but more recent research has started to take into account fresh factors in an effort to ascertain whether preferences for OCF may be influenced by learner characteristics such as gender, proficiency, motivation, learning style, educational programme, or educational context (Fidan, 2015; Deptolla, 2019; Gómez Argüelles et al., 2019; Papi et al., 2021; Sepehrinia et al., 2020; Tasdemir & Yalcin Arslan, 2018; Wiboolyasarin et al., 2020, 2022; Yakisik, 2021). These investigations have been motivated by the possibility that learner traits could affect feedback efficacy (Nassaji, 2016). Additionally, the concept of emotion has been included in the description of learner characteristics. For instance, foreign language learners have reported anxiety due to various factors such as fear of failing tests or public speaking and receiving negative feedback from classmates and teachers (Horwitz et al., 1986). Foreign language classroom anxiety (FLCA) is characterised by the state anxiety that language learners experience during language study and/or use (Horwitz, 2017). Several studies have investigated FLCA to determine its sources and relationship to student performance and accomplishments (e.g., Dewaele, 2017; Horwitz, 2010).

Studies on directed learning situations have examined the link between FLCA and OCF preferences, as well as other variables, among learners. For example, Geckin (2020) discovered that Turkish pre-intermediate EFL female learners differed from male participants in their higher levels of anxiety, preferences for delayed feedback, and preferences for repetition as the primary error correction approach. Both males and females considered feedback as a critical aspect of language learning and assessed the teacher's input more favourably, especially concerning major and personal errors. Rassaei (2015) demonstrated that low-anxiety students benefited from both metalinguistic feedback and recast, whereas metalinguistic input had a stronger impact on their progress. In contrast, learners with high anxiety benefited substantially more from recast than from metalinguistic feedback. Positive emotions, conversely, enhance a student's ability to notice classroom details and increase their awareness of language input (MacIntyre & Gregersen, 2012). Students in a positive emotional state, or with foreign language enjoyment (FLE), have a better comprehension of the foreign language and can reduce the adverse effects of negative emotions. However, individuals overwhelmed by negative feelings are more likely to have a restricted focus, thereby limiting their language input intake (Dewaele & Dewaele, 2020). Positive emotions increase learners' sense of security while studying a foreign language. Although Dewaele and MacIntyre (2014) conducted a study that simultaneously analysed variation in FLCA and FLE, a few studies have examined the influence of both FLCA and FLE on learner-internal factors (such as preference) and learner-external variables (e.g., educational context or, in this case, TFL). This study combines FLCA and FLE in the same research design based on Dewaele and MacIntyre's (2014) questionnaire items to determine whether they influence OCF approaches. Additionally, research (e.g., Yang, 2016) shows that the proficiency level and first language (L1)

context of learners are the main determinants shaping their preferences for OCF. Consequently, it is vital to thoroughly analyse learners' preferences for OCF techniques in connection with their L1, proficiency level, FLCA, and FLE. This study aims to address the following research questions to enrich the understanding of the existing literature:

(1) What are the preferences of TFL students for OCF techniques?

(2) Do the proficiency level and L1 background of TFL learners influence their preferences for OCF techniques?

(3) Do the FLE and FLCA levels of TFL learners influence their preferences for OCF techniques?

METHOD

Context

This study examines Thai as Foreign Language (TFL) programmes offered at Chinese, Japanese, and Korean universities where Thai is used as the medium of instruction. The language departments of these universities offer Thai elective courses to local undergraduate students, aimed at enhancing their academic Thai language skills and cultural knowledge.

Participants

A non-probabilistic convenience sampling was utilised to select undergraduate students from TFL classrooms for this study. To ensure an adequate sample size, G*Power 3.1.9.7 was used to calculate a medium sample size of 0.25, power (1-β) = 0.95, and α = 0.05. The estimation suggested that 252 people would be sufficient. A total of 288 students participated in the study and completed an online questionnaire administrated through Microsoft Forms (see the following section for more information). The participants were non-native speakers of Thai with diverse educational backgrounds, consisted of 207 (71.90%) females and 81 (28.10%) males. Respondents from countries in the East Asian cultural sphere: People's Republic of China (*n* = 145, 50.30%), Republic of Korea (*n* = 82, 28.50%), and Japan (*n* = 61, 21.20%) reported being taught TFL at their universities and were prompted to differentiate themselves using the ACTFL Proficiency Guidelines 2012[3] presented in their native versions on the web. They self-reported their proficiency level and were categorised into three main categories: 48.30% of participants (*n* = 139) claimed proficiency at the intermediate level, while the rest were assessed at the novice level (*n* = 106, 36.80%) and the advanced level (*n* = 43, 14.30%). The participants' mean TFL learning experience was 2.35 (SD = 1.76) years.

Questionnaire

An online questionnaire consisting of four sections were used to collect data. Part 1 began with a section on demographics, from which the preceding information was extracted. Following this, participants were asked to rate their preferences for ten OCF techniques (OCFT) (see Appendix), modified from previously discussed literature and Wiboolyasarin *et al*.'s (2020, 2022) instruments, using a 5-point Likert scale to assess how learners feel when they make an oral error in TFL lessons, ranging from 'very good' (5 points) to 'very poor' (1 point). In Part 3, the FLE scale is a 21-item instrument proposed by Dewaele and MacIntyre (2014) to which participants replied using a 5-point Likert scale ranging from 'strongly agree' (5 points) to 'strongly disagree' (1 point). For measurement of FLCA in Part 4, the FLCA scale constructed by ibid. (2014) contained a total of 8 items. The alpha reliabilities for the OCFT, FLE, and FLCA in the present investigation were 0.81, 0.94, and 0.93, respectively, showing acceptable internal consistency for the instruments (Cohen *et al*., 2018).

Procedure and Data Analysis

Prior to the commencement of the study, the research design and questionnaire were subject to approval by the Ethics Committee of the first author's institution. Prior consent was sought from all participants before administering the survey. The online questionnaire was made accessible via Microsoft Forms for a period of one month. Initially, 288 students completed the questionnaire online. The participants were then segregated into two groups, based on their scores on the FLE questionnaire: high-enjoyment learners (*n* = 134) and low-enjoyment learners (*n* = 154). High-enjoyment learners were classified as those who scored higher than the mean, while low-enjoyment learners were classified as those who scored lower. Similarly, participants were divided into two groups based on their FLCA questionnaire scores: those with high anxiety (*n* = 154) and those with low anxiety (*n* = 134). The mean score on the FLCA questionnaire was 3.55, with a standard deviation of 1.05. Participants who scored above the mean were classified as high-anxiety learners, while those who scored below were classified as low-anxiety learners.

RESULTS

RQ1: What are the Preferences of TFL Students for OCF Techniques?

On the 10 items of Part 2 in the questionnaire, all respondents were required to rate their feelings when making oral blunders in TFL classes. The frequencies and average rating scores on a 5-point Likert scale are presented in Table 3.

[3] ACTFL. (2012). *ACTFL proficiency guidelines 2012.* Retrieved August 14, 2022, from https://www.actfl.org/resources/actfl-proficiency-guidelines-2012

Table 3

Frequencies of Answers and Descriptive Statistics of 10 Items (n = 288)

	OCF Techniques	Answer	*n*	%	M	SD	Skewness	Kurtosis
1	Disregard	5	38	13.20	2.65	1.31	0.45	-0.91
		4	39	13.50				
		3	57	19.80				
		2	94	32.60				
		1	60	20.80				
2	Elicitation	5	40	13.90	2.80	1.30	0.31	-1.05
		4	51	17.70				
		3	55	19.10				
		2	95	33.00				
		1	47	16.30				
3	Explicit correction	5	150	52.10	4.28	0.90	-1.21	-0.13
		4	82	28.50				
		3	46	16.00				
		2	6	2.40				
		1	4	1.40				
4	Peer correction	5	61	21.20	3.66	0.97	-0.40	0.64
		4	103	35.80				
		3	96	33.30				
		2	21	7.30				
		1	7	2.40				
5	Metalinguistic feedback	5	163	56.60	4.39	0.79	-1.11	-0.27
		4	80	27.80				
		3	41	14.20				
		2	3	1.00				
		1	1	0.30				
6	Recast	5	67	23.30	3.66	1.01	-0.41	-0.27
		4	96	33.30				
		3	93	32.30				
		2	24	8.30				
		1	8	2.80				
7	Repetition	5	39	13.50	2.99	1.18	0.13	-0.80
		4	53	18.40				
		3	93	32.30				
		2	73	25.30				
		1	30	10.40				
8	Clarification request	5	40	13.90	3.11	1.17	-0.07	-0.70
		4	63	21.90				
		3	102	35.40				
		2	54	18.80				
		1	29	10.10				

	OCF Techniques	Answer	*n*	%	M	SD	Skewness	Kurtosis
9	Public feedback	5	47	16.30	3.28	1.12	0.18	-0.55
		4	70	24.30				
		3	108	37.50				
		2	42	14.60				
		1	21	7.30				
10	Private feedback	5	104	36.10	3.96	1.00	0.75	-0.56
		4	98	34.00				
		3	61	21.20				
		2	20	6.90				
		1	5	1.70				

Note. The coding of the answers: very good = 5; good = 4; fair = 3; poor = 2; very poor = 1

The results show that TFL students preferred metalinguistic comment (*M* = 4.39) and explicit correction (*M* = 4.28), respectively. Private feedback (*M* = 3.96), peer correction (*M* = 3.66), and recast (*M* = 3.66) also received positive responses, with overall 'good' ratings. These findings suggest that participants preferred OCF techniques that were more explicit. In contrast, ignoring errors was the least desirable OCF type, with mean scores of 2.65. The average values for elicitation, repetition, explanation request, and public feedback ranged from 2.80 to 3.28, indicating that participants tended to consider them unfavourable TFL classroom strategies. Descriptive statistics of the participants' characteristics classified by L1 background, proficiency level, FLE, and FLCA can be found in Table 4.

The skewness and kurtosis values for each OCF strategy were also analysed, and the results showed that the values between -1.21 and 0.75 were distributed normally, which falls within the range of ±2 for skewness. The kurtosis scores ranged from -1.05 to 0.64, indicating a typical distribution, as they fall within the scale of ±*2* (George & Mallery, 2016).

Table 4

Mean Score of the Participants' L1 Background, Proficiency Level, FLE, and FLCA

OCF Technique	Mean Score									
	L1			Proficiency			FLE		FLCA	
	CN	JP	KR	N	I	A	High	Low	High	Low
Disregard	2.86	2.52	2.40	2.60	2.55	3.12	2.66	2.65	2.82	2.47
Elicitation	2.98	2.77	2.50	2.78	2.63	3.40	2.95	2.66	2.79	2.80
Explicit Correction	4.29	4.30	4.24	4.25	4.26	4.42	4.54	4.05	4.28	4.28
Peer Correction	3.72	3.64	3.56	3.61	3.64	3.84	3.87	3.48	3.70	3.61
Metalinguistic Feedback	4.46	4.39	4.28	4.33	4.47	4.30	4.64	4.18	4.45	4.33
Recast	3.82	3.67	3.37	3.62	3.65	3.77	3.88	3.47	3.63	3.69
Repetition	3.12	3.00	2.77	2.91	2.96	3.30	3.24	2.78	3.00	2.99
Clarification Request	3.34	2.82	2.90	3.02	3.04	3.53	3.30	2.94	3.17	3.03
Public Feedback	3.44	3.16	3.07	3.28	3.14	3.72	3.56	3.03	3.16	3.41
Private Feedback	4.00	4.03	3.83	3.96	3.90	4.14	4.26	3.69	3.90	4.02

Note. The abbreviation of the L1: CN = Chinese; JP = Japanese; KR = Korean. The abbreviation of the proficiency level: N = Novice; I = Intermediate; A = Advanced

RQ2: Do the Proficiency Level and L1 Background of TFL Learners Influence Their Preferences for OCF Techniques?

The normality assumption was assessed using the Shapiro-Wilk test, which tested the univariate normality of ten OCF approaches. This test was chosen due to the sample size being less than 2,000 (Adigun, 2021). A multivariate analysis of variance (MANOVA) was then conducted, with OCF techniques serving as the between-subjects variable, and proficiency level and L1 background as dependent variables. The MANOVA results indicated significant differences in proficiency level and L1 background across OCF strategies (Pillai's trace = 0.18, F = 1.30, p = 0.031), suggesting that the interaction of these two factors can explain the variance in the preferences of L2 learners for OCF. To ensure the validity of the MANOVA, it was confirmed that the covariance matrices among groups were equal, as determined by a Box's M test (p < 0.001) (Huberty & Petoskey, 2000).

Based on the aforementioned results, a between-subject analysis or a univariate test was conducted for each dependent variable, as presented in Table 5.

Table 5 displays the outcomes of ten OCF techniques. The values for disregard (F (4, 283) = 3.43, p = 0.01), peer correction (F (4, 283) = 2.38, p = 0.04), recast (F (4, 283) = 2.62, p = 0.03), and private feedback (F (4, 283) = 2.45, p = 0.05) suggested that there were significant differences in the proficiency level and L1 background on OCF. On the other hand, the scores for elicitation (F (4, 283) = 2.26, p = 0.23), explicit correction (F (4, 283) = 1.15, p = 0.33), metalinguistic feedback (F (4, 283) = 0.79, p = 0.29), repetition (F (4, 283) = 1.50, p = 0.20), clarification request (F (4, 283) = 1.51, p = 0.20), and public feedback (F (4, 283) = 0.95, p = 0.44) indicated a small variation between the average of the groups. As is evident, there was a strong interaction of L1 by proficiency on disregard, peer correction, recast, and private feedback. A simple slope was plotted to gain a clearer understanding of the overall pattern of interactions and to identify the factor that predicted whether the participants' L1 background (Chinese, Japanese, or Korean) and their proficiencies were significantly related to these four dependent variables. The slopes of the graphs in Figures 1, 2, 3, and 4 are represented graphically, showing that four OCF techniques for learners had significant interactions with respect to their L1 background and proficiency level.

RQ3: Do the FLE and FLCA Levels of TFL Learners Influence Their Preferences for OCF Techniques?

A MANOVA was conducted with OCF techniques as the between-subjects variable, and the FLE and FLCA levels as the dependent variables. The overall MANOVA, which examined the differences in FLE and FLCA levels across OCF strategies, was found to be non-significant. To further explore the relationship between FLE, FLCA, and L2 learners' OCF preferences, Pillai's trace was employed. The analysis revealed no significant interaction between FLE and FLCA (Pillai's trace = 0.38, F = 1.10, p = 0.58), indicating that only 38% of the variability in L2 learners' OCF preferences can be attributed to the interaction between FLE and FLCA. These results suggest that other factors may have a more significant impact on L2 learners' preferences for OCF, and further investigation is warranted.

In light of these results, the between-subject effects or univariate tests for each dependent variable was conducted, which are depicted in Table 6.

The results of the ten OCF techniques are presented in Table 6. The analyses for disregard (F (4, 283) = 3.58, p = 0.06), elicitation (F (4, 283) = 1.12, p = 0.29), explicit correction (F (4, 283) = 2.65, p = 0.10), peer correction (F (4, 283) = 0.10, p =

Table 5

Test of Between L1 and Proficiency Effects on OCF Techniques

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig	Partial ^{2}
L1 x Proficiency	Disregard	21.99	4	5.50	3.43	0.01	0.05
	Elicitation	9.02	4	2.26	1.42	0.23	0.02
	Explicit Correction	3.78	4	0.94	1.15	0.33	0.02
	Peer correction	9.51	4	2.38	2.57	0.04	0.04
	Metalinguistic Feedback	3.15	4	0.79	1.26	0.29	0.02
	Recast	10.49	4	2.62	2.68	0.03	0.04
	Repetition	8.22	4	2.05	1.50	0.20	0.02
	Clarification Request	7.76	4	1.94	1.51	0.20	0.02
	Public Feedback	4.60	4	1.15	0.95	0.44	0.01
	Private Feedback	9.68	4	2.42	2.45	0.05	0.03

0.76), metalinguistic feedback (F (4, 283) = 3.14, p = 0.08), recast (F (4, 283) = 1.22, p = 0.27), repetition (F (4, 283) = 1.35, p = 0.25), clarification request (F (4, 283) = 1.64, p = 0.20), public feedback (F (4, 283) = 0.003, p = 0.96) and private feedback (F (4, 283) = 1.52, p = 0.22) indicated that there were no statistically significant differences between FLE and FLCA with respect to OCF approaches.

DISCUSSION

It is widely observed that L2 learners tend to prefer receiving OCF as it acknowledges errors as a natural occurrence in the process of acquiring an L2. However, traditional TFL instruction places emphasis on language form, leading to Chinese, Japanese, and Korean TFL students feeling less confident in their ability to speak the target language when teachers disregard their errors. Our study shows that the majority of students preferred the use of explicit OCF techniques, specifically metalinguistic remarks and explicit correction. Therefore, it is recommended that foreign language courses adopt strategies that teach the correct version of errors and linguistics terminology, allowing students to enhance their language skills without the burden of self-correction (Lee, 2017; Ur, 2012).

In L2 programmes, OCF from instructors is regarded as crucial, as students often lack sufficient exposure to the target language in their environment, as Muslem et al. (2021) have emphasised. It is acknowledged that Thai was not employed as a lingua franca in the previous classes of other students, depriving them of enough opportunities to practise their oral Thai. Our findings indicate that learners appreciate being informed directly of their errors and the correct forms, and they do not feel any shame when corrected by their teachers or classmates. This finding reinforces prior studies, suggesting that OCF may not lead to antagonising or demoralising East Asian students (Wiboolyasarin et al., 2020) or other EFL learners (Wiboolyasarin et al., 2022; Yang, 2016; Yu, 2019).

Although it is widely recognised that providing OCF in a public setting can be beneficial for the entire class, this study's results indicate that the instructor should provide

Figure 1

Simple Plot of Interaction Effects Between Proficiency and L1 Towards Disregard

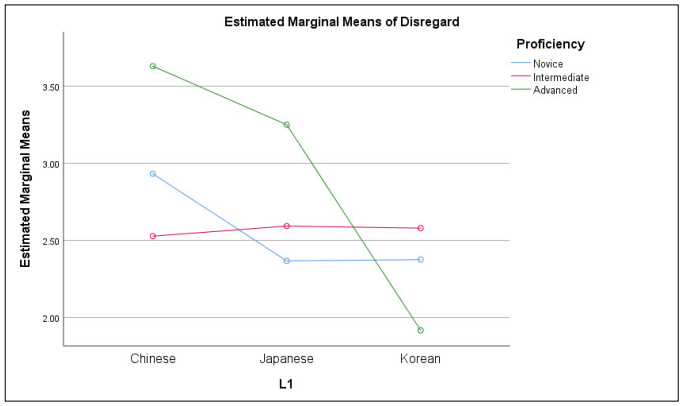


Figure 2

Simple Plot of Interaction Effects Between Proficiency and L1 Towards Peer Correction

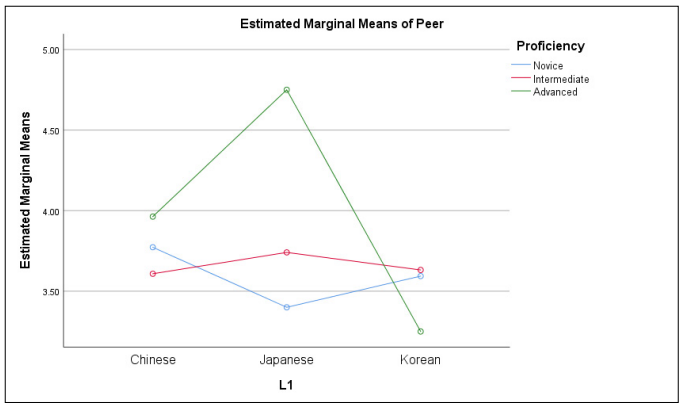


Figure 3

Simple Plot of Interaction Effects Between Proficiency and L1 Towards Recast

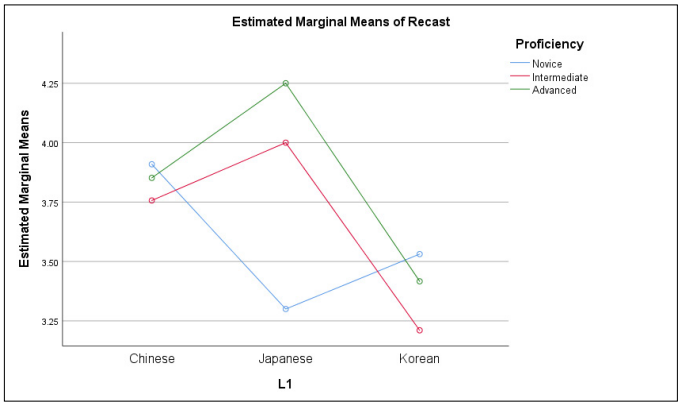


Figure 4

Simple Plot of Interaction Effects Between Proficiency and L1 Towards Private Feedback

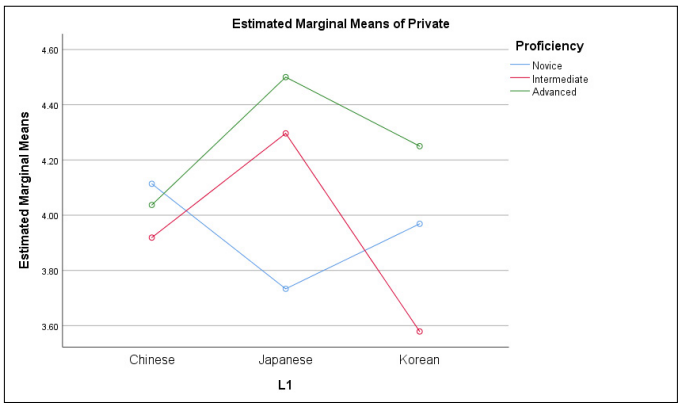


Table 6
Test of Between FLE and FLCA Effects on OCF Techniques

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig	Partial ^{2}
FLE x FLCA	Disregard	6.00	1	6.00	3.58	0.06	0.01
	Elicitation	1.86	1	1.86	1.12	0.29	0.00
	Explicit Correction	1.99	1	1.99	2.65	0.10	0.01
	Peer Correction	0.09	1	0.09	0.10	0.76	0.00
	Metalinguistic Feedback	1.77	1	1.77	3.14	0.08	0.01
	Recast	1.21	1	1.21	1.22	0.27	0.00
	Repetition	1.82	1	1.82	1.35	0.25	0.00
	Clarification Request	2.16	1	2.16	1.64	0.20	0.01
	Public Feedback	0.00	1	0.00	0.003	0.96	0.00
	Private Feedback	1.42	1	1.42	1.52	0.22	0.01

such feedback in private. Some students consider OCF to be restrictive and humiliating when it is frequently used in the classroom (Martínez Agudo, 2013). In particular, students who receive excessive corrective feedback in front of their peers may become bashful and experience an increase in anxiety. This finding is consistent with previous research (Oflaz, 2019), which found a positive correlation between shyness and FLCA, demonstrating conclusively that shyness increases as speaking anxiety grows.

Based on the findings, participants tended to perceive elicitation, repetition, and explanation requests as unfavourable TFL classroom tactics. Several reasons can explain this phenomenon. For example, when the teacher echoes the incorrect speech with a rising intonation and a doubtful expression, trying to imply that there was an error, students may not recognise it as a correction or may not realise that the instructor's pronunciation was different from their own, interpreting it merely as an echo or confirmation. Similarly, active OCF strategies that encourage students to rethink what they have said and construct the correct form independently did not yield positive results for East Asian students who are accustomed to being teacher-fed information.

Li's (2021) research revealed that various factors related to the learner and the context can constrain the effectiveness of feedback. Among these factors, a student's emotional engagement with an instructor's OCF can shape the student's perception and response to the feedback received. It would be intriguing to investigate if two learner variables, namely proficiency level and L1 background, could influence learners' preferences for OCF approaches. The literature has shown that learners' preferences for OCF can differ based on their proficiency and L1 background. Our study confirmed that Korean TFL learners preferred fewer OCF techniques, such as disregard, peer correction, recast, and private feedback, than their Chinese counterparts. In addition, advanced-level students favoured more OCF approaches than intermediate- and beginner-level students. These results may be due to the fact that less proficient students may feel embarrassed when receiving OCF on their erroneous speech, especially when the feedback includes wrong information. These findings suggest that learners from diverse linguistic backgrounds and proficiency levels are aware of their errors and expect instructors to provide OCF feedback appropriately.

Although instructors use various OCF techniques to provide feedback to students, it is crucial to adapt their strategies to suit students' preferences (Tasdemir & Yalcin Arslan, 2018). As Bada and Okan (2000) stated, instructors should consider their students' expectations and provide each student with the opportunity to express their opinions. To our knowledge, previous studies have not explored the relationship between FLCA/FLE levels and OCF preferences extensively. Therefore, our study provides a basis for further research. Although there were no statistically significant differences between FLE and FLCA for OCF techniques, we observed that low and high FLE/FLCA groups had distinct preferences. Specifically, high FLE groups favoured more OCF techniques than low FLE groups, while high FLCA groups selected more OCF techniques than low FLCA groups, except for elicitation, recasting, and public feedback. This finding has significant pedagogical implications, as it suggests that instructors need to use OCF techniques that enable L2 students with different psychological profiles to feel confident that their errors are being detected and corrected appropriately. Furthermore, instructors should focus on frequent and major errors and refrain from providing OCF strategies for some errors. Nonetheless, if instructors believe that feedback on specific errors is essential, they can use metalinguistic feedback and explicit corrective approaches, which are more effective than recast (Lyster, 2004; Lyster & Saito, 2010), and

were the most preferred techniques by both FLE and FLCA groups in our study.

CONCLUSION

The current study offers a novel contribution to the field by examining the characteristics of L2 undergraduate TFL learners in East Asian universities, and exploring the potential impact of proficiency level, L1 background, FLE, and FLCA on preferences for OCF approaches. Such information could prove useful for educators working with this cohort of students across different regions. Despite the complexity of the findings, the results suggest that TFL learners generally prefer metalinguistic explanation and explicit correction, although there is a significant relationship between L1 and proficiency with regards to disregard, peer correction, recast, and private feedback. Furthermore, FLCA and FLE were found to have a notable impact on individuals' perceptions of the value of more prompts and explicit correction. Thus, TFL teachers may wish to provide learners with more explicit OCF on oral errors, particularly those with high FLE and FLCA, as this could promote greater engagement with feedback. This study offers insights into the factors that may shape OCF preferences and highlights the potential significance of these features for OCF preference formation.

It is important to note several caveats in interpreting the results of this research. Firstly, the participants were drawn from a population of students at universities in three countries where Thai is the language of instruction, and therefore, the findings may have limited generalisability to other institutional contexts. Additionally, as this study only examined four characteristics of student preferences for OCF, its findings may not fully capture their impact on learning outcomes. Moreover, the sample size was limited to university-level students, which may not be representative of other age groups. Further research is therefore needed to investigate the preferences for oral feedback across a wider range of students.

ETHICAL STATEMENT

The investigation involving human participants was approved by the Ethics Committee of the first author's university. Informed consent was obtained from all participants, who were informed that participation in the online questionnaire survey was voluntary, and they could withdraw at any time. The online questionnaire was designed to automatically obtain the participants' informed consent upon submission.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Watcharapol Wiboolyasarin: Conceptualization; Project administration; Validation; Writing - review and editing.

Phornrat Tiranant: Conceptualization; Data curation; Investigation.

Teavakorn Khumsat: Data curation; Investigation.

Tidarat Ngamnikorn: Data curation; Investigation.

Kanokpan Wiboolyasarin: Formal analysis; Methodology; Writing - original draft.

Somkiat Korbuakaew: Supervision; Validation.

Nattawut Jinowat: Other contribution; Resources; Visualization; Writing - original draft.

REFERENCES

Adigun, O. T. (2021). Inclusive education among pre-service teachers from Nigeria and South Africa: A comparative cross-sectional study. *Cogent Education, 8*(1), 1–19. https://doi.org/10.1080/2331186X.2021.1930491

Bada, E., & Okan, Z. (2000). Students' language learning preferences. *The Electronic Journal for English as a Second Language, 4*(3), 1–15.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233. https://doi.org/10.1080/03075070600572132

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Deptolla, A. (2019). *Language-learner preferences for corrective feedback* (Publication No. 13812139) [Master's thesis, University of Colorado at Boulder]. ProQuest Dissertations and These Global.

Dewaele, J.-M. (2017). Psychological dimensions and foreign language anxiety. In S. Loewen, & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 433–450). Routledge.

Dewaele, J.-M., & Dewaele, L. (2020). Are foreign language learners' enjoyment and anxiety specific to the teacher? An investigation into the dynamics of learners' classroom emotions. *Studies in Second Language Learning and Teaching, 10*(1), 45–65. https://doi.org/10.14746/ssllt.2020.10.1.3

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal, 1*(1), 3–18. https://doi.org/10.5070/l2.v1i1.9054

Ellis, R., & Shintani, N. (2014*). Exploring language pedagogy through second language acquisition research.* Routledge.

Fidan, D. (2015). Learners' preferences of oral corrective feedback: An example of Turkish as a foreign language learners. *Educational Research and Reviews, 10*(9), 1311–1317. https://doi.org/10.5897/ERR2015.2162

Geçkin, V. (2020). Do gender differences affect foreign language anxiety and preferences for oral corrective feedback? *Journal of Theoretical Educational Science, 13*(3), 591–608. https://doi.org/10.30831/akukeg.64744

George, D., & Mallery, P. (2016). *IBM SPSS statistics 23 step by step: A simple guide and reference* (14th ed.). Routledge.

Gómez Argüelles, L., Hernández Méndez, E., & Perales Escudero, M. D. (2019). EFL teachers' attitudes towards oral corrective feedback: A case study. *PROFILE: Issues in Teachers' Professional Development, 21*(1), 107–120. https://doi.org/10.15446/profile.v21n1.69508

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation development: Teaching English /ɹ/ to Korean adult EFL learners. *System, 60*, 117–127. https://doi.org/10.1016/j.system.2016.06.007

Ha, X. V., & Nguyen, L. T. (2021). Targets and sources of oral corrective feedback in English as a foreign language classrooms: Are students' and teachers' Beliefs aligned? *Frontiers in Psychology, 12*, 1–10. https://doi.org/10.3389/fpsyg.2021.697160

Horwitz, E. K. (2010). Foreign and second language anxiety. *Language Teaching, 43*, 154–167. https://doi.org/10.1017/S026144480999036X

Horwitz, E. K. (2017). On the misreading of Horwitz, Horwitz, and Cope (1986) and the need to balance anxiety research and the experiences of anxious language learners. In C. Gkonou, M. Daubney, & J.-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 31–47). Multilingual Matters.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70*, 125–132. http://doi.org/10.1111/j.1540-4781

Huberty, C. J, & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In H. E. A. Tinsley, & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 183–208). Academic Press. https://doi.org/10.1016/B978-012691360-6/50008-2

Kartchava, E., & Ammar, A. (2014). Learners' beliefs as mediators of what is noticed and learned in the language classroom *TESOL Quarterly, 48*(1), 86–109. https://doi.org/10.1002/tesq.101

Lee, E. J. (2013). Corrective feedback preferences and learner repair among advanced ESL students. *System, 41*(2), 217–230. https://doi.org/10.1016/j.system.2013.01.022

Lee, I. (2017). *Classroom assessment and feedback in L2 school contexts.* Springer. https://doi.org/10.1017/S0261444819000247

Li, S. (2018). Corrective feedback in L2 speech production. In J. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–9). John Wiley & Sons.

Li, S. (2021). Oral corrective feedback. In H. Mohebbi, & C. Coombe (Eds.), *Research questions in language education and applied linguistics: A reference guide* (pp. 353–357). Springer. https://doi.org/10.1007/978-3-030-79143-8_63

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*(3), 399–432. https://doi.org/10.1017/S0272263104263021

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition, 32*(2), 265–302. https://doi.org/10.1017/S0272263109990520

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms *Language Teaching, 46*(1), 1–40. https://doi.org/10.1017/S0261444812000365

MacIntyre, P. D., & Gregersen, T. (2012). Emotions that facilitate language learning: The positive-broadening power of the imagination. *Studies in Second Language Learning and Teaching, 2*, 193-213. https://doi.org/10.14746/ssllt.2012.2.2.4

Martínez Agudo, J. D. (2013). An investigation into how EFL learners emotionally respond to teachers' oral corrective feedback. *Colombian Applied Linguistics Journal, 15*(2), 265–278.

Muslem, A., Zulfikar, T., Astilla, I., Heriansyah, H., & Marhaban, S. (2021). Students' perception toward oral corrective feedback in speaking classes: A case at English education department students. *International Journal of Language Education, 5*(4), 244–259. https://doi.org/10.26858/ijole.v5i4.19010

Nassaji, H. (2016). Anniversary article interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research, 20*(4), 535–562. https://doi.org/10.1177/1362168816644940

Oflaz, A. (2019). The effects of anxiety, shyness and language learning strategies on speaking skills and academic achievement. *European Journal of Educational Research, 8*(4), 999–1011. https://doi.org/10.12973/eu-jer.8.4.999

Papi, M., Wolff, D., Nakatsukasa, K., & Bellwoar, E. (2021). Motivational factors underlying learner preferences for corrective feedback: Language mindsets and achievement goals. *Language Teaching Research, 25*(6), 858–877. https://doi.org/10.1177/13621688211018808

Rassaei, E. (2015). Oral corrective feedback, foreign language anxiety and L2 development. *System, 49*, 98–109. https://doi.org/10.1016/j.system.2015.01.002

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal, 59*(1), 23–30. https://doi.org/10.1093/elt/cci003

Sakiroglu, H. Ü. (2020). Oral corrective feedback preferences of university students in English communication classes. *International Journal of Research in Education and Science, 6*(1), 172–178.

Sepehrinia, S., Fallah, N., & Torfi, S. (2020). English language teachers' oral corrective preferences and practices across proficiency groups. *PROFILE: Issues in Teachers' Professional Development, 22*(2), 163–177. https://doi.org/10.15446/profile.v22n2.82369

Tasdemir, M. S., & Yalcin Arslan, F. (2018). Feedback preferences of EFL learners with respect to their learning styles. *Cogent Education, 5*(1), 1–17. https://doi.org/10.1080/2331186X.2018.1481560

Ur, P. (2012*). A course in English language teaching* (2nd ed.). Cambridge University Press.

Wiboolyasarin, K., Kamonsawad, R., Jinowat, N., & Wiboolyasarin, W. (2022). EFL learners' preference for corrective feedback strategies in relation to their self-perceived levels of proficiency. *English Language Teaching Educational Journal, 5*(1), 32–47. https://doi.org/10.12928/eltej.v5i1.4403

Wiboolyasarin, W., Wiboolyasarin, K., & Jinowat, N. (2020). Learners' oral corrective feedback perceptions and preferences in Thai as a foreign language tertiary setting. *Journal of Language and Linguistic Studies, 16*(2), 912–929. https://doi.org/10.17263/jlls.759344

Yakisik, B. Y. (2021). EFL learners' preferences and emotions about oral corrective feedback at secondary education in Turkey: Are there gender and grade-level differences? *Indonesian Journal of English Language Teaching and Applied Linguistics, 6*(1), 103–122. https://doi.org/10.21093/ijeltal.v6i1.889

Yang, J. (2016). Learners' oral corrective feedback preferences in relation to their cultural background, proficiency level and types of error. *System, 61*, 75–86. httpS://doi.org/10.1016/j.system.2016.08.004

Yu, S. (2019). Learning from giving peer feedback on postgraduate theses: Voices from master's students in the Macau EFL context. *Assessing Writing, 40*, 42–52. https://doi.org/10.1016/j.asw.2019.03.004

Zhu, Y., & Wang, B. (2019). Investigating English language learners' beliefs about oral corrective feedback at Chinese universities: A large-scale survey. *Language Awareness, 28*(2), 139–161. https://doi.org/10.1080/09658416.2019.1620755

APPENDIX

Questionnaire

Preferences for oral corrective feedback techniques

Please rate yourself honestly using the following scales based on your true feelings about the given statements.

"When you make an oral mistake in TFL classes, you think it's ...very good, good, fair, poor, or very poor.... if the teacher..."

1. ...disregards it, doesn't correct at all.
2. ...indicates there's a mistake but doesn't actually tell you what's wrong, so you have to try to work it out for yourself.
3. ...says what was wrong and tells you what the right version.
4. ...says what was wrong and gets someone else to say the correct version.
5. ...explains why it was wrong, what the rule is.
6. ...modifies the entirety of your utterance in an error-free manner.
7. ...repeats your mistakes in a high intonation to emphasise them.
8. ...indicates that your utterance contained some mistakes by using phrases like 'Excuse me?' or 'I don't understand.'
9. ...provides feedback to the student in front of others in public (e.g., in the classroom).
10. ...provides feedback to the student one-on-one or in private (e.g., in the teacher's office).

CONTENT

EDITORIAL

RESEARCH ARTICLES